Research Article

NHL Pathological Image Classification Based on Hierarchical Local Information and GoogLeNet-Based Representations

Jie Bai [iD],[1] **Huiyan Jiang** [iD],[2] **Siqi Li,**[2] **and Xiaoqi Ma**[3]

[1]*Northeastern University, Shenyang 110819, China*
[2]*Software College, Northeastern University, Shenyang 110819, China*
[3]*School of Science and Technology, Nottingham Trent University, Nottingham, UK*

Correspondence should be addressed to Huiyan Jiang; hyjiang@mail.neu.edu.cn

Background. Accurate classification for different non-Hodgkin lymphomas (NHL) is one of the main challenges in clinical pathological diagnosis due to its intrinsic complexity. Therefore, this paper proposes an effective classification model for three types of NHL pathological images, including mantle cell lymphoma (MCL), follicular lymphoma (FL), and chronic lymphocytic leukemia (CLL). *Methods.* There are three main parts with respect to our model. First, NHL pathological images stained by hematoxylin and eosin (H&E) are transferred into blue ratio (BR) and Lab spaces, respectively. Then specific patch-level textural and statistical features are extracted from BR images and color features are obtained from Lab images both using a hierarchical way, yielding a set of hand-crafted representations corresponding to different image spaces. A random forest classifier is subsequently trained for patch-level classification. Second, H&E images are cropped and fed into a pretrained google inception net (GoogLeNet) for learning high-level representations and a softmax classifier is used for patch-level classification. Finally, three image-level classification strategies based on patch-level results are discussed including a novel method for calculating the weighted sum of patch results. Different classification results are fused at both feature 1 and image levels to obtain a more satisfactory result. *Results.* The proposed model is evaluated on a public IICBU Malignant Lymphoma Dataset and achieves an improved overall accuracy of 0.991 and area under the receiver operating characteristic curve of 0.998. *Conclusion.* The experimentations demonstrate the significantly increased classification performance of the proposed model, indicating that it is a suitable classification approach for NHL pathological images.

1. Introduction

Lymphoma is a type of malignant tumors that originate from the lymphoid hematopoietic tissues [1], which are divided into the Hodgkin lymphomas (HL) and the Non-Hodgkin's lymphomas (NHL). NHL is one of the common malignant tumors with the fastest growth rate, and chronic lymphocytic leukemia (CLL), follicular lymphoma (FL), and mantle cell lymphoma (MCL) are the major types of NHL. Pathological characteristics vary complicatedly among different lymphoma subtypes. And the vision disparities of different subtypes are indistinguishable. Therefore, accurate classification of various lymphoma histopathology has become one of the difficult challenges for hematologists and pathologists, as well as the development of computer vision methods.

Histopathological image analysis is an important part of computer aided diagnosis (CAD). Complex structure information in pathological images provides great value for diagnosis of many diseases including most tumors. There is a mass of complicated pathological information including abundant spatial structure, various cell types, and variation in cell morphology in histopathological images. Traditionally, these characteristics are found and analyzed by pathologists with clinical experiences, which is time-consuming and inefficient. With the development of CAD, approaches have been designed to automatically extract expressive features from histopathological images for subsequent differentiation tasks and then to facilitate disease diagnosis.

Traditionally, there are a large number of feature extraction and classification approaches based on traditional

machine learning or handcrafted features which reflect low level characteristics involving morphological descriptors, spatial arrangement, color and texture features. Sertel *et al.* [2] obtained semantic description by model-based intermediate representation (MBIR) and combined MBIR with low level texture features for the grading of the follicular lymphoma (FL). However, those three grades of FL have significant morphological differences which make it easier to identify them. Baker *et al.* [3] utilized a coarse-to-fine method to extract the local features in a two-stage way for brain tumor whole slide classification, but they also pointed out that the method might not be efficient for small pathological image classification with identified lesions. There are some approaches with good classification performances but just for binary classification tasks. Fukuma *et al.* [4] extracted morphological features and spatial arrangement features based on graph theory as feature descriptions. Cao *et al.* [5] combined texture features, spatial distribution, and semantic features of the nuclear structure to describe the various morphology of the nucleus and the structural and interpretation information of the image. These binary classification methods are hard to be extended to multiclass classification. Zhou *et al.* [6] developed an image classification platform to give a standard solution for different problems, which sacrificed pertinence for specific disease and resulted in a low accuracy of 70.9% for lymphoma classification. Meng *et al.* [7] developed an approach to lymphoma classification that divided images into 25 subblocks from which a set of 505 color and texture features were extracted. However, its weighting scheme was performed in a class level and majority voting was still used for image label acquisition, which might lead to losing a part of local information. Song *et al.* [8] proposed a method to extract high-dimensional multimodal descriptor, and subcategory discriminant transform (SDT) was used to enhance discriminative power of descriptors. This approach extracted sufficient image information and achieved decent results, but it used deep learning method just for comparison purpose instead of benefitting from making use of the full power of deep learning. Those above approaches extract customized, manually designed features which require domain-specific knowledge, and the extracted features just represent a portion of the image characteristics.

With the rapid development of the theory of deep learning, automatic learning approaches are developed for feature extraction and medical image classification. Zhou *et al.* [9] proposed a multispectral feature learning model based on convolutional sparse coding (CSC) and spatial pyramid matching (SPM), which could automatically learn a set of convolution filter banks from separate spectra. Vu *et al.* [10] proposed an automatic feature extraction framework by learning class-specific dictionaries, and a low-complexity method was used for histopathology classification and disease grading. They tried to explore the histopathological image classification issue from the perspective of feature discovery and dictionary learning and obtained promising results in different datasets. Approaches based on convolutional neural network (CNN) have been widely applied in various medical visual recognition tasks and demonstrated promising performances [11]. Gao *et al.* [12] proposed an automatic framework for HEp-2 cell image classification, where a CNN with three convolutional layers is employed. The recently prevailed various CNN pretrained on ImageNet can be explored to extend this work. Sirinukunwattana *et al.* [13] proposed spatially constrained convolutional neural network (SC-CNN) for nucleus detection and a novel neighboring ensemble predictor (NEP) was developed to more accurately predict the cell nuclei class. Note that those methods based on deep learning are awfully time consuming since they need the training process.

Recently, there are some approaches that combine traditional handcrafted features and high-level representations of deep learning methods. Codella *et al.* [14] extracted a combination of local binary pattern (LBP) and CNN features based on several enhanced and segmented results, so it required an early precise segmentation process. Song *et al.* [15] proposed Fisher vector (FV) encoding combined with multiple types of local features including high-level features from CNN and applied a novel separation-guided dimension reduction method, which achieved state-of-the-art results. However, it had low feature extraction efficiency due to complicated computation.

To sum up, while a number of approaches have been proposed, accurate histopathological image classification, especially for malignant lymphoma images, still remains a challenging task due to some issues. First, the key to classification task is feature extraction and the above methods all explored different features to represent image information, but more efficient and complete representations still need extracting in an appropriate way. Second, some methods performed classification by employing patches-division and showed their ability for local knowledge extraction, but they have not considered different classification contributions for those cropped patches with diverse locations in one image. Third, most of those methods were just based on either handcrafted features or high-level features. Some approaches combined them together, but still needed improvement to achieve better performance with lower computational complexity.

In this study, to address these problems, we propose a hierarchical local information and transfer learning based approach specific to transformed image spaces for the classification of MCL, FL, and CLL histopathological images. Original RGB images are transformed to blue ratio and CIE Lab image space [16] to amplify the texture and color information, respectively. Then image patches are cropped at two hierarchies; parent patches are cropped from transformed images and subpatches are obtained from parent patches by utilizing sliding window. Handcrafted features are extracted from subpatches specific to different image spaces and parent-patch features are statistically computed from corresponding subpatches representations. Random forest classifier is trained for parent patch classification. Different strategies for image level classification are discussed, including a novel distance matrix weighting (*DMW*) method for calculating weighted sum of parent patch results. Furthermore, a pretrained Google Inception Net v3 model is fine-tuned to perform patch-level classification based on transfer learning. Prediction fusion at both feature and image level is employed for more adaptable performance.

(a)   (b)   (c)
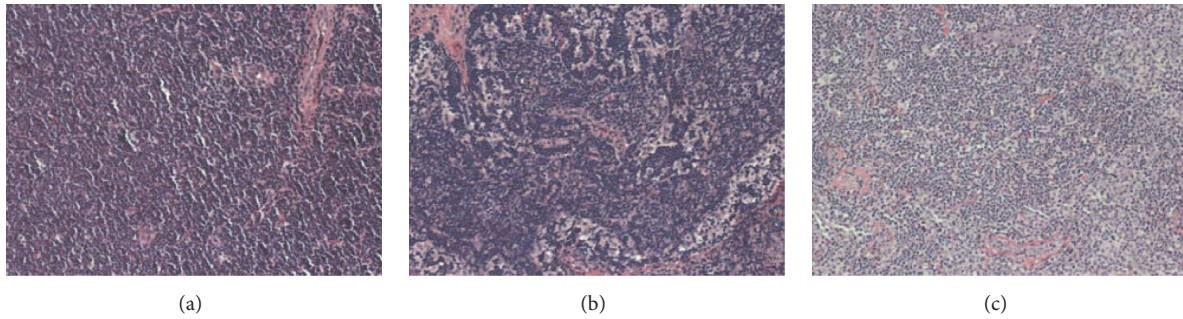
FIGURE 1: The lymphoma histological image samples for different NHL subtypes: (a) CLL, (b) FL, and (c) MCL.

The contributions of this paper can be summarized as follows.

(i) It proposes a novel lymphoma subtypes classification approach utilizing a multiple hierarchies classification method from local to global, which improves the completeness of features and reduces the spatial complexity of computation simultaneously by combining small patch level feature extraction and efficient image level classification strategies.

(ii) It shows the benefits of appropriate image level classification strategy based on patch level results by comparing three strategies, namely, majority voting, mean score of patches, and weighted sum of patch scores.

(iii) It proposes a novel distance matrix weighting (*DMW*) method to calculate appropriate weights for patch level results and presents improved performance.

(iv) It presents that more effective features can be extracted from transformed image space. Different image space transformations are employed to enhance particular image characteristics. Space-specific feature set can be extracted which is better than features extracted from original images.

(v) It shows that transfer learning technique based on deep CNN network can provide better efficiency for lymphoma classification. Pretrained Google Inception Net is fine-tuned to improve the performance and leave out time-consuming training process simultaneously.

This paper is organized as follows. Section 2 introduces the dataset used in this study and illustrates the detailed methodology. Section 3 shows the experimental setup and results. Section 4 presents the comparison with other existing works and discussions for experiment performances. Finally, the conclusions are provided in Section 5.

2. Materials and Methodology

2.1. Dataset. In this study, a public dataset named IICBU malignant lymphoma dataset [18] with respect to studies in National Cancer Institute and National Institute on Aging, both in the United States [19, 20], is used. There are 10

cases for each of these three lymphoma types, namely, CLL, FL, and MCL, and in total 30 H&E stained lymphoma histological slides used for this study. A Zeiss Axioscope white light microscope with a 20× objective and a color digital camera AxioCam MR5 are used to obtain these images. Lesion regions in these slides are digitally photographed and reserved without compression in the *tif* format, using RGB color model, with 1388 × 1040 resolution and 24 bits quantization. The final dataset contains 374 images with 113, 139, and 122 regions of CLL, FL, and MCL. In our study, the whole data set is randomly divided into training and testing sets by a ratio of 7:3.

Figure 1 shows the samples of histological images employed in our experiments to evaluate the proposed approach. A CLL sample is depicted in Figure 1(a), an FL sample is shown in Figure 1(b), and an MCL sample is presented in Figure 1(c). These images are from biopsies of different patients in different hospitals, so the inner class difference could be prominent because of the staining variance [21], which makes the classification task quite challenging.

2.2. Method. The proposed classification pipeline takes an H&E stained malignant lymphoma image section as input and finally yields as output the image level classification result with likelihood estimations for each class. Figure 2 illustrates the overall flow of our proposed methodology which has four basic steps.

(a) Image Space Transformation. Both single channel and multiple channel image transformation methods are considered. The original images are mapped into the blue ratio gray space to emphasize cellular morphology. The CIE Lab color space transformation is applied to original RGB images to amplify the color characteristic. These two space transformation methods are used to enhance different features of original images by which targeted discriminative features can be extracted subsequently. This process can also be considered as a preprocessing step.

(b) Handcrafted Features Extraction and Classification. A set of texture and statistical features are obtained from blue ratio gray space and color features are extracted from Lab space images. A hierarchical local-to-global feature extraction method is used to obtain complete patch-wise features in
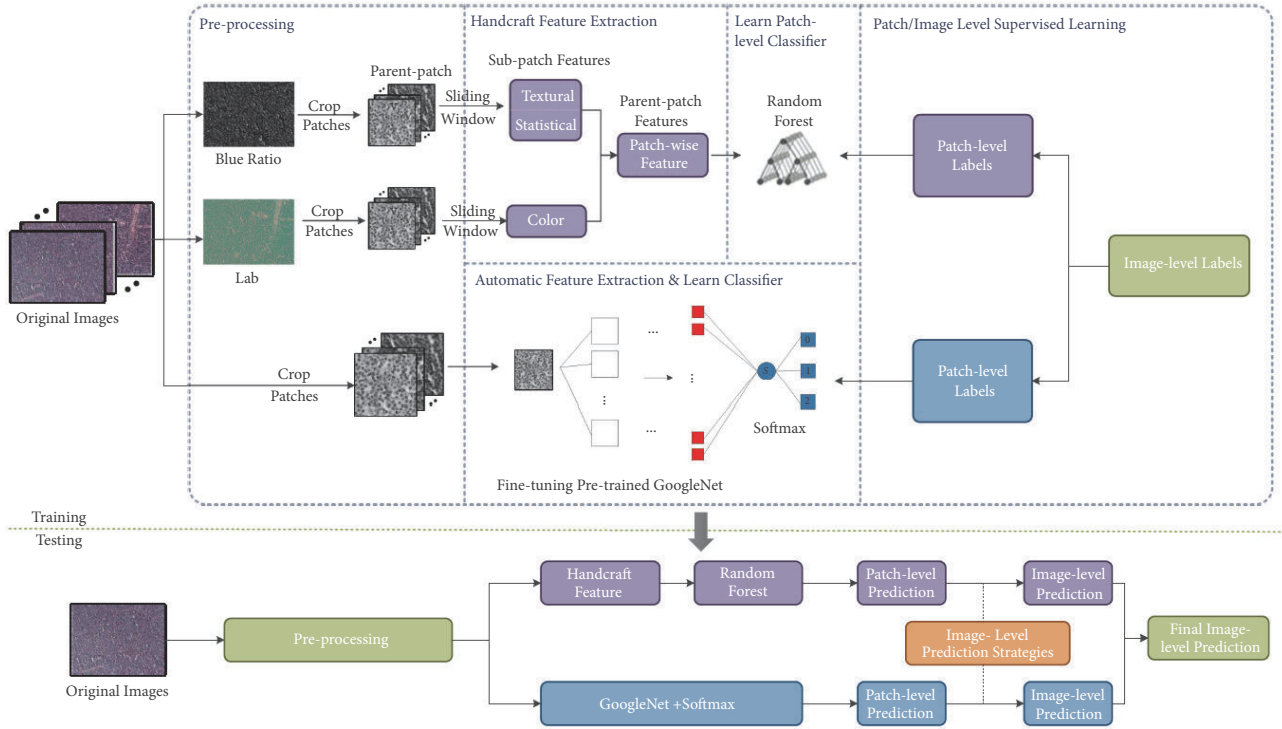
FIGURE 2: The overall flow of the proposed method.

different scales. Then different feature subsets are combined to train a patch level random forest classifier.

(c) Transfer Learning Based High-Level Feature Extraction and Classification. A pretrained deep convolutional neural network is fine-tuned to extract patch level automatic features from cropped original images. A softmax classifier is connected to the output layer to implement patch level classification.

(d) Multi-Path Integration and Image Level Classification. Image level prediction is calculated based on patch level prediction of this image using several strategies. Weighted fusion in both patch level and image level prediction is used to integrate results from different paths.

Detailed description of the proposed pipeline's steps is shown in Figure 2.

2.2.1. Image Space Transformation

(A) Blue Ratio Image. Chang *et al.* [22] proposed that RGB images could be transformed into a one channel gray space named blue ratio image to enhance the nuclear dye and attenuate background. And it is time-consuming or even impractical to extract some kinds of handcrafted features in three channel images. Therefore, in order to calculate texture and statistical features of original images in a more efficient way, the original RGB images can be mapped into blue ratio image space. The conversion policy is defined as

$$BR = \frac{100 \times B}{1 + R + G} \times \frac{256}{1 + B + R + G} \quad (1)$$

where B, R, and G are intensities of blue, red, and green channels and BR represents the intensity at the corresponding position in transformed blue ratio gray space. The first term enhances the nuclear signal and the second term attenuates the background. Figure 3(b) illustrates the blue ratio transformation result for the original image as shown in Figure 3(a). Subsequently, texture and statistical features are extracted in the transformed gray level images.

(B) CIE Lab Color Space. The CIE Lab color space is a color mode enacted by International Commission on Illumination (CIE) in 1976 [16]. It is a device-independent color system based on physiological features. It has three basic coordinates $L * a * b*$, abbreviated as Lab [23]. L component represents the luminosity, and a and b represent the green-red and blue-yellow color value components, respectively. The range of L channel's value is [0, 100], which corresponds to different lightness from black to white. The range of a channel's value is [-128, 127] and b channel has values in range [-128, 127]. Figure 3(c) is a sample of Lab color space transformation result for the original image in Figure 3(a).

RGB color space is transformed into XYZ color space first and then the XYZ space is converted into Lab color space [24], as given by (2) and (3), respectively. Subsequently, color features are extracted from the transformed Lab images.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$
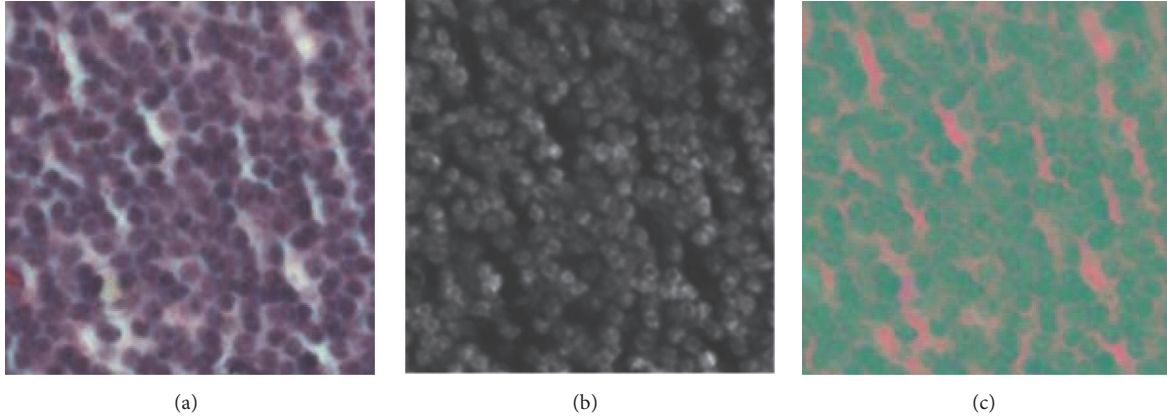
FIGURE 3: Sample patches of different image spaces. (a) One sample patch of original RGB image for CLL. (b) Blue ratio image. (c) Lab color space image.

$$L = 116 \times f\left(\frac{Y}{Y_n}\right) - 16$$

$$a = 500 \times \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \qquad (3)$$

$$b = 200 \times \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right]$$

where $X_n = 0.950456$, $Y_n = 1.0$, and $Z_n = 1.088754$. $f(t)$ is defined as

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases} \qquad (4)$$

2.2.2. Handcrafted Feature Extraction

(A) Overall Process. Based on the image preprocessing results discussed in Section 2.2.1, discriminative patch-wise manual features are extracted in a local to global way. The entire process for manual feature extraction can be expounded in four major steps as follows.

(a) Cropping Parent Patches. Big parent patches are cropped in preprocessed 1388×1040 images with window size s_1. Parent patches are cropped in a nonoverlapping way to reduce subsequent computation complexity. An appropriate value of parent window size s_1 is necessary to capture local information completely and precisely.

(b) Cropping Subpatches. Relatively smaller subpatches are cropped from parent patches using sliding window with window size s_2. In order to fully extract local information, subpatches are cropped in an overlapped way with overlapping ratio r_o.

(c) Extracting Subpatch Features. The preprocessed single or multiple channel images with original size are first sliced into $s_1 \times s_1$ parent patches. Then $s_2 \times s_2$ subpatches are obtained from each parent patch using sliding window. Specific n-dimensional feature sets are extracted from subpatch. For example, texture features are extracted from single channel patch, and three channel Lab images contribute color features.

(d) Calculating Parent Patch Features. Since each parent patch has its corresponding subpatches set, its features can be generated from subpatches' features. As described in Section 2.2.2, n-dimensional features are extracted from each subpatch. First, we can calculate the mean and standard deviation as well as the 10th and 90th percentile of each dimension of the n features among all subpatches cropped from this parent patch, which yields a total of $4n$-dimensional features as the parent patch's representation.

Finally, $4n$ features at parent patch level are extracted. Since it is more likely that parent patches contain complete and representative information, computation of features from subpatch level to parent patch level is done to obtain more meaningful and discriminative features. The detailed kinds of original features extracted from subpatch are discussed below.

(B) Features Extraction in Blue Ratio Images. As described in Section 2.2.1, blue ratio image emphasizes the foreground in H&E stained histopathology images, namely, the nuclear structure, and cellular morphology is discriminative for tumor classification. In addition, blue ratio image is a single channel gray space image which reduces computation cost of some texture and statistic features.

Local binary pattern (LBP) texture representations and a set of statistical features, expressed as r_{LBP} and r_{sta}, respectively, as explained below, are extracted from processed blue ratio image in view of its characteristics.

LBP is a local texture descriptor which has the characteristics of simple calculation, insensitivity to illumination changes, and good texture expression ability [25]. The specific operation steps are as follows:

(a) Set the center pixel of the window with a fixed radius r and sampling points' number p as the threshold, and

compare the pixel values of the p sample points with it.

(b) If the pixel value of the sample point is greater than the center pixel value, the corresponding position of this pixel point is marked as 1, otherwise it is 0.

(c) The above procedure produces a p-bit binary number, which is the LBP value of the center pixel of the window.

Therefore, an LBP operator with p sample points in a circular region of radius r will produce 2^p patterns. As the neighborhood sampling points' number p increases, the type of binary pattern augments exponentially. The statistical histogram of the LBP patterns is usually used to express image information when the LBP operator is used for texture classification, and excessive pattern types will result in too large data quantity and too sparse histogram. Therefore, it is necessary to perform dimensionality reduction on the original LBP patterns, that is to use fewer feature dimensions to represent image information effectively.

Ojala *et al.* proposed using uniform local binary patterns (ULBP) to reduce the dimension of LBP operator since most LBP patterns only contain two 0-1 or 1-0 transitions in real images [26]. The "uniform pattern" is defined as follows: if the cyclic binary value corresponding to an LBP has at most two 0-1 or 1-0 transitions, this binary pattern is one kind of uniform pattern. Specifically, let p be the sampling points' number, there are 2 patterns for 0 transition, $2(p-1)$ patterns for 1 transition, and $(p-1)(p-2)$ patterns for 2 transitions. Moreover, nonuniform patterns with more than 2 transitions are regarded as one pattern; thus a total of $2 + 2(p-1) + (p-1)(p-2) + 1 = p(p-1) + 3$ patterns are generated. As a consequence, the statistical histogram dimension of LBP is reduced from 2^p to $p(p-1) + 3$. For instance, for a fixed radius r, when the sampling points number $p = 8$, the feature dimension is reduced from $2^8 = 256$ to $8 \times (8-1) + 3 = 59$.

The detailed LBP and statistical features extraction settings are as follows:

r_{LBP}: ULBP for radii 1 and 2 with 8 sampling points is extracted, which is 118-dimensional for each subpatch. So r_{LBP} contains 472 features for each parent patch.

r_{sta}: 9 statistical features of the gray level histogram are computed for each subpatch including maximum, minimum, sum, mean, standard deviation, median, first and third quartiles, and interquartile range. So r_{sta} contains 36 features for each parent patch.

(C) Features Extraction in Lab Images. RGB color channels tend to be affected by color variance in images; hence Lab color space is used to extract color features. As illuminated in Section 2.2.1, Lab color space contains three channels of which a and b channels represent color component. Therefore, the color representation r_{Lab} is extracted as follows:

r_{Lab}: Extracting statistics from color histogram with 85 bins of a and b channels, respectively, and cascading the two results. As a result, 170-dimentional features can be extracted from each subpatch and r_{Lab} contains 680 features for one parent patch.

2.2.3. High-Level Feature Extraction.

CNN is a multilayer neural network including convolutional, pooling, and fully connected layers, which can learn features in an automatic and supervised way. It has shown success in biomedical imaging applications [13, 27, 28]. Some customized deep CNN models trained on ImageNet have displayed improved performance, and their effectiveness on biomedical images implies natural and biomedical images which virtually share similar low level features [29].

Since a deep network usually contains a large number of parameters that need to be fully trained and optimized, suitable parameter initialization is extremely helpful for training process, which is of great importance for model's final performance. Poor parameter initialization may lead to slow convergence rate and insufficient parameter training, especially for small biomedical data sets. The deep network trained on ImageNet performs well on the complex classification tasks of other natural images, and some existing studies have shown that classification network learning from natural images has certain transfer effects on biomedical image data sets [30]. This effectiveness, which may originate from the similarity of natural images and biomedical images in some low level texture and shape features, has also been verified by experiments in this paper. Therefore, the pretrained deep network on ImageNet can provide excellent parameter initialization. This paper makes some customized modifications to the network output structure based on the initialized parameters and performs fine-tuning process on the whole network using the lymphoma data set to make it more suitable for our classification task.

Specifically, the Google Inception Net (GoogLeNet) v3 model pretrained on ImageNet [31] is applied to classification task in this study. It is a deep network with 42 layers and provides state-of-the-art classification result for ImageNet dataset. The original RGB image is cropped into $s_3 \times s_3$ patches and 2048-dimensional local features r_G are densely extracted from the last fully connected layer for each patch. The patches are used to perform a basic fine-tuning process, that is, to add a fully connected linear layer with the true number of classes after the last layer in GoogLeNet and perform backpropagation just for last fully connected layers. In this way, the pretrained GoogLeNet is customized to fit characteristics of this specific data set. Finally, a softmax classifier is connected after the network to perform classification. Parameters in other layers keep invariant as in the pretrained model.

By utilizing transferred GoogLeNet, time used for the training process can be saved significantly; meanwhile, the classification performance is improved due to powerful representation ability and high universality of pretrained GoogLeNet.

Note that training a new customized CNN model for specific biomedical imaging application is a more usual choice than using pretrained models. A non-pretrained Google Inception Net v3 implemented by ourselves is experimented for comparison. The model is trained totally using datasets in this study. Experimental result shows the improved efficiency of transfer learning in this dataset with small number of biomedical images.

2.2.4. Integration and Classification

(A) Patch Level Integration and Classification. Different feature sets are extracted for each parent patch in Sections 2.2.2 and 2.2.3, including LBP features r_{LBP}, statistical features r_{sta}, Lab color features r_{Lab}, and high-level features r_G. For handcrafted features r_{LBP}, r_{sta}, and r_{Lab}, an auxiliary random forest classifier is trained on specific feature set F to perform patch level classification. Feature set F can be a single type of manual features, namely, r_{LBP}, r_{sta}, or r_{Lab}, and it can also be a combination of multiple kinds of features, which involves a fusion process at feature level. There are mainly two kinds of feature combination modes considered in this study, which are $F_1 : \{r_{LBP}, r_{sta}\}$ and $F_2 : \{r_{LBP}, r_{sta}, r_{Lab}\}$, since feature set F_1 is extracted from blue ratio image and F_2 is an expansion of F_1 with additional color feature from Lab image. For high-level feature r_G, the softmax classifier cascaded after GoogLeNet is used for patch level classification.

(B) Image Level Classification. One original sized image can be cropped into a set of parent patches. There are different methods for parent patch classification as discussed in Sections 2.2.2, 2.2.3, and 2.2.4. For each parent patch, the probability scores for three classes can be calculated. Image level classification can be performed based on parent patch prediction result considering that patches of the corresponding original image reflect local knowledge of the whole image. For image level classification, three strategies are chosen to make sufficient use of patch level result and to obtain convictive image level prediction. Two of them are commonly used approaches illustrated as follows.

(a) Majority Voting (MV). Assume an original image I can be cropped into m parent patches. Let m_1, m_2, and m_3 denote the numbers of patches classified as the first, second, and third classes. $K(\max([m_1, m_2, m_3]), [m_1, m_2, m_3])$ calculates the final prediction class of the whole image, where $K(x, y)$ is the function that calculates value x's index in array y.

(b) Mean Score (MS) of All Patches. Since each parent patch has probability scores for three classes, scores for the whole image can be calculated taking average over all parent patches' scores. The image level classification result can be obtained based on image scores.

In addition, a novel strategy to calculate weighted sum of patch scores for image level classification is proposed. Different patches are corresponding to different locations of one image and contain various local information, so they own different discrimination and contribution for classification task. Therefore, appropriate weights for different patches are favorable for acquisition of conclusive image classification result. The most crucial part is the determination of weights. A distance matrix weighting (*DMW*) method is proposed to determine weights in this study, which is described below.

Suppose the patch with closer association to other patches is inclined to provide more contribution to the whole image. This hypothesis is made based on the fact that the lymphoma samples in this study are all sections from lesions, so class-specific characteristics are reflected on most cropped patches.

One original image I can be cropped into m parent patches $\{I_1, I_2, \ldots, I_m\}$. Specific feature set can be extracted for each parent patch, which generates m patch-wise feature sets $\{r_1, r_2, \ldots, r_m\}$. And patch level prediction scores set $\{s_1, s_2, \ldots, s_m\}$ can be obtained from the trained classifier based on feature set for each parent patch, where s_i is the classification score for the ith patch and contains three score values corresponding to three classes. For one patch I_k, a weight w_k should be calculated.

First, a distance matrix M_d is generated for all patches based on their feature values. Let $d_{ij} = D(r_i, r_j)$ denote the element of row i and column j in M_d, where $D(x, y)$ is a function to calculate distance between vectors x and y, and Euclidean distance is used in our study. Then the average distance da_k between parent patch I_k and all others is calculated based on M_d as follows:

$$da_k = \frac{1}{m-1} \sum_{i=1}^{m} d_{ki} \tag{5}$$

In essence, da_k is the mean value of the kth row in M_d, and the patch with smaller average distance should be given greater weights. Thus a distance score Sd_k for patch I_k is defined by taking the reciprocal of da_k to represent this inversely proportional relationship, making it convenient to calculate patch weight.

$$Sd_k = \frac{1}{da_k} = (m-1) \frac{1}{\sum_{i=1}^{m} d_{ki}} \tag{6}$$

Then the weight w_k for patch I_k is calculated as follows:

$$w_k = Sd_k \frac{1}{\sum_{i=1}^{m} Sd_i} \tag{7}$$

Finally the image-level classification score S_I is obtained based on patch level prediction scores set $S\{s_1, s_2, \ldots, s_m\}$ and weights set $W\{w_1, w_2, \ldots, w_m\}$ as shown in the following equation:

$$S_I = \sum_{i=1}^{m} w_i \cdot s_i \tag{8}$$

As a result, we consider totally three image-level classification strategies denoted as *MV*, *MS*, and *DMW*.

(C) Image Level Integration. Different feature set selection methods at patch level can be combined with one of the image level classification strategies. The probability scores of multiple methods can be fused at both patch level and image level. There are mainly five method combination modes experimentalized in this study:

(a) $C_1(F_1 + DMW)$: Feature set F_1 at patch level combined with *DMW* at image level.

(b) $C_2(F_2 + MS)$: Feature set F_2 at patch level combined with *MS* method at image level.

(c) $C_3(F_1 + F_2 + DMW)$: For one original image, two patch level prediction scores sets $S_1 = \{s_1^{(1)}, s_2^{(1)}, \ldots, s_m^{(1)}\}$,

$S_2 = \{s_1^{(2)}, s_2^{(2)}, \ldots, s_m^{(2)}\}$ of cropped parent patches can be obtained based on F_1 and F_2, respectively, where $s_i^{(j)}$ is the classification score of the image's i^{th} parent patch based on feature set F_j. Score fusion at patch level can be performed by calculating weighted sum of S_1 and S_2 to integrate characteristics of two methods. Then the final patch level probability score set $S = w_P \cdot S_1 + (1 - w_P) \cdot S_2$ can be obtained. *DMW* strategy is used for image level classification.

(d) $C_4(r_G + DMW)$: High-level feature set r_G combined with *DMW* at image level.

(e) $C_5(C_3 + C_4)$: Weighted summation with weight w_I of image level results S_{I3} and S_{I4} from C_3 and C_4. The final image level probability score $S_{I5} = w_I \cdot S_{I3} + (1 - w_I) \cdot S_{I4}$.

Method combination modes C_1, C_2, and C_3 are based on handcrafted features among which C_3 integrates features in C_1 and C_2, whereas C_4 is totally based on deep automatic features. In order to bind handcrafted and automatic features, the image level probability scores of C_3 and C_4 are fused by weighted summation in C_5.

Note that there are more alternative combination modes at features, patches, and images levels. The core principle is to choose different typical feature sets in different combination modes, and three image-level strategies are all experimentally evaluated with specific features and the best strategy is finally combined with this feature set (except image level fusion method C_5). Finally, only several typical choices are experimentally compared to discuss each part's classification contribution.

3. Experiments and Results

3.1. Evaluation Metrics.
Classification performances at both patch level and image level can be evaluated using several commonly used metrics calculated as follows:

$$ACC = \frac{1}{n}\sum_{i=1}^{n} I\left(f\left(x_i\right) = y_i\right) \tag{9}$$

$$SEN\left(c_i\right) = \frac{Num\left(PC\left(c_i\right)\right)}{Num\left(GT\left(c_i\right)\right)} \tag{10}$$

$$SPE\left(c_i\right) = \frac{Num\left(PC\left(-c_i\right)\right)}{Num\left(GT\left(-c_i\right)\right)} \tag{11}$$

$$P\left(c_i\right) = \frac{Num\left(PC\left(c_i\right)\right)}{Num\left(PT\left(c_i\right)\right)} \tag{12}$$

$$R\left(c_i\right) = \frac{Num\left(PC\left(c_i\right)\right)}{Num\left(GT\left(c_i\right)\right)} \tag{13}$$

$$F_1\left(c_i\right) = \frac{2 \times P\left(c_i\right) \times R\left(c_i\right)}{P\left(c_i\right) + R\left(c_i\right)} \tag{14}$$

where ACC represents the overall classification accuracy. $SEN(c_i)$ and $SPE(c_i)$ mean the sensitivity and specificity for class c_i. $P(c_i)$, $R(c_i)$, and $F_1(c_i)$ are precision, recall, and F_1 score for class c_i, respectively. $I(f(x_i) = y_i)$ defines that if $f(x_i) = y_i$, $I(f(x_i) = y_i) = 1$, otherwise $I(f(x_i) = y_i) = 0$. $-c_i$ denotes all the other classes except c_i. $Num(PC(c_i))$, $Num(PT(c_i))$, and $Num(GT(c_i))$ represent correctly predicted number of class c_i, the total number of all predicted c_i, and the total number of c_i in the ground truth, respectively. Note that recall is equal to sensitivity. Furthermore, area under the receiver operating characteristic curve (AUC) is also calculated for evaluation.

3.2. Parameter Sensitivity.
Some parameters in the study have high sensitivity for final performances, such as sizes of parent and subpatches, overlapping rate of subpatches, number of decision trees of random forest classifier, and fine-tuning iteration times of pretrained GoogLeNet. Sensitivity of these parameters is experimentally analyzed by observing both patch and corresponding image levels classification performances when just one primary parameter is adjusted at a time, as shown in Figure 4. Sensitivity for three parameters: parent patch size s_1, subpatch size s_2, and subpatch overlap rate r_o are depicted in Figure 4. (a), (b), (c) are patch level accuracy and (d), (e), (f) are the corresponding image level accuracy.

Figures 4(a) and 4(d) show the patch and image levels accuracy with different parent patch size s_1, respectively. Let the subpatch size $s_2 = 50$ and its overlapping rate $r_o = 0.5$. The selected feature set is $r_{LBP} + r_{sta}$ extracted from blue ratio image and a random forest classifier with 350 decision trees is trained for classification. Majority voting (*MV*) is applied for image level prediction. Note that the above is the default experimental setup for all the subsequent experiments. Performances for three parent patch scales (100×100, 200×200, and 300×300) are compared at both patch level and image level.

Subpatch size impacts completeness of local knowledge extraction and then may make a difference for the final results. The accuracy for different subpatch size at patch and image levels is shown in Figures 4(b) and 4(e), respectively. Parent patch is cropped in a nonoverlapping way considering calculation complexity, so only the subpatch overlapping rate r_o is experimented as shown in Figures 4(c) and 4(f). Parent patch size is fixed to $s_1 = 300$ in above experiments. Default setup is used for other parameters.

Transfer learning is used by applying the pretrained GoogLeNet model and performing fine-tuning process for the last fully connected layer using dataset in this study. Patch-level accuracy on testing set for different iteration times t in fine-tuning process is shown in Figure 5. Size s_3 of the input original RGB patch is set to 300. Ten percent of training data is randomly divided into validation set to monitor training situation. The model is implemented based on tensorflow. Learning rate $e = 0.001$ and cross entropy loss function are used. Stochastic gradient descent (SGD) is selected as the optimizer and the training batch size is set to 100. As shown in Figure 5, the test accuracy increases as t increases and reaches maximum 83.55% when $t=9000$, but after that, the accuracy begins to decline because of overfitting.

(a) Sensitivity of parent patch size

(b) Sensitivity of subpatch size

(c) Sensitivity of overlapping rate

(d) Sensitivity of parent patch size

(e) Sensitivity of subpatch size
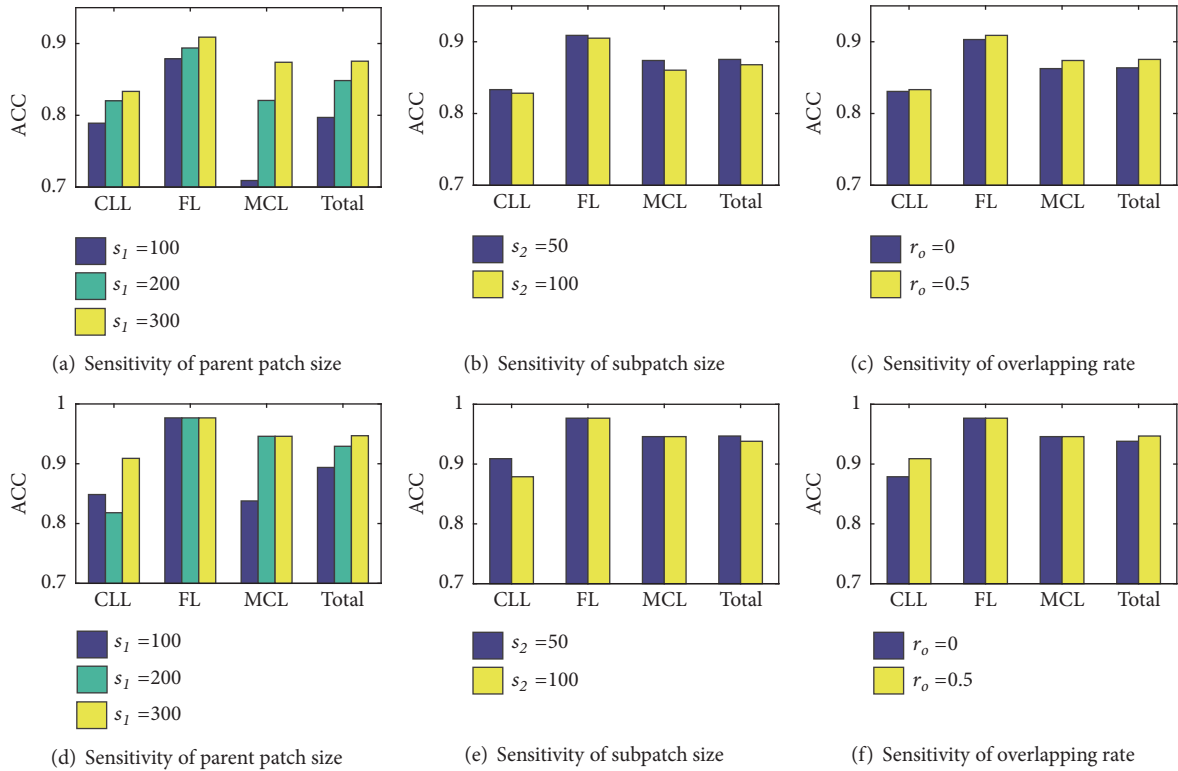
(f) Sensitivity of overlapping rate

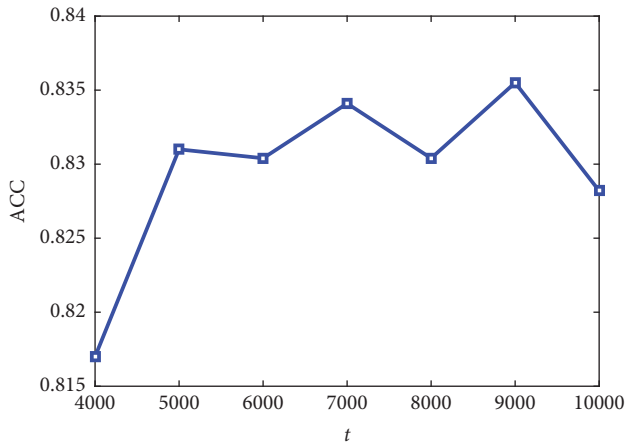FIGURE 4: Sensitivity for three parameters at both patch and image levels.

FIGURE 5: Patch level classification accuracy of transferred GoogLeNet with different fine-tuning iteration times t.
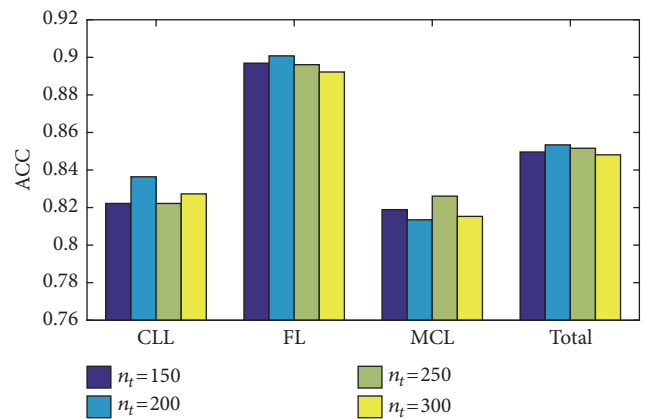

FIGURE 6: Patch level classification accuracy for different decision tree number.

A random forest (RF) classifier is used for patch-wise classification based on handcrafted features as discussed in Section 2.2.4. The influence of RF's decision tree number n_t is experimentally evaluated. Figure 6 exhibits patch level classification accuracy with different decision tree numbers. Parent patch size s_1 is set to 200. Other related parameters are determined using default setup. The effect of decision tree number varies among different classes; however, as a whole, RF classifier obtains best patch level classification result when $n_t = 200$. Note that Figure 6 only shows patch level

performance, and rational choice for image level classification strategy may make additional adjustment for final prediction.

Weighted fusion is used at both patch and image levels as described in Section 2.2.4 with weights w_P and w_I, which are set to 0.7 and 0.5, respectively, using parameter scanning.

3.3. Experimental Results. To evaluate the contributions of different feature sets, several kinds of features involved in this study are tested and compared. Specific RF classifiers are trained for different handcrafted feature sets r_{LBP}, r_{sta}, r_{sta} as well as their combination. Fine-tuned GoogLeNet combined

TABLE 1: Patch level classification accuracy for different feature set.

Feature Set	CLL	FL	MCL	Total
Gist	0.5424	0.5829	0.6252	0.5850
LPQ	0.6556	0.8070	0.6405	0.7083
r_{LBP}	0.7343	0.8171	0.7000	0.7546
r_{sta}	0.7828	0.8713	0.7414	0.8029
r_{Lab}	0.7394	0.8837	0.8631	0.8348
r_G	0.7394	0.8837	0.8631	0.8355
$\{r_{LBP}, r_{sta}\}$	0.8253	0.8969	0.8324	0.8549
$\{r_{LBP}, r_{sta}, r_{Lab}\}$	0.9283	0.9217	0.9324	0.9271
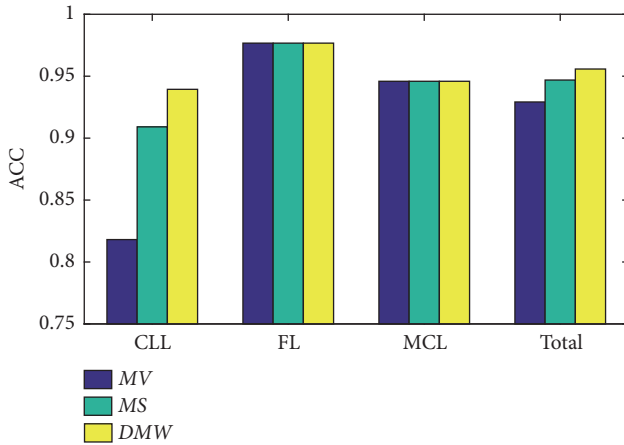
FIGURE 7: Classification accuracy of different image level classification strategies.

FIGURE 8: Image level classification accuracy for different method combination modes.

with softmax is used to test high-level feature r_G. To confirm the efficiency of the concerned features, two additional kinds of handcrafted features Gist [32] and local phase quantization (LPQ) [33] are considered. The accuracy of different feature sets is displayed in Table 1, where parent patch size $s_1 = 200$. Fine-tuning iteration time of GoogLeNet is set to 9000.

Several strategies in image level classification are used to take full advantages of patch level results as illustrated in Section 2.2.4. Figure 7 shows the image level accuracy of the three strategies. Parent patch size $s_1 = 200$, and default setup is used for other parameters. The impact of different image level strategies is mainly reflected in classification performance for CLL in this setup. By utilizing the proposed *DMW* image level strategy, the accuracy for CLL is significantly improved compared with commonly used *MV*. In fact, our experimental result shows the proposed method *DMW* can provide better or comparative performance for different parameters setups, which indicates its efficiency is superior to those of other commonly used methods to calculate image level results based on patch level predictions.

There are mainly several method combination modes (a)~(e) in Section 2.2.4. Fusion at patch level, namely, combination mode C_3, is based on C_1 and C_2. Fusion at image level is implemented in C_5 based on C_3 and C_4. Image level classification accuracy of the five combination schemes $C_1 \sim C_5$ is shown in Figure 8. Performances for C_1, C_2, and C_4 are
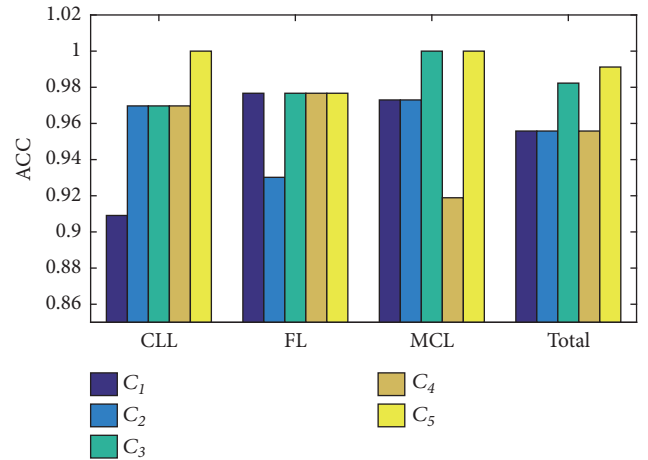
relatively low because these combination modes are based on specific feature set, but they show different discrimination power for three classes. Therefore, combination modes that fuse their characteristics obtain improved performances. C_3 combines C_1 and C_2 by score fusion at patch level which provides prominent performance improvement. C_5 integrates C_3 and C_4 through weighted summation at image level results and obtains further amelioration for classification of CLL, which is considered as the final scheme for this classification task. Table 2 shows the classification indices including overall accuracy (ACC), overall area under the curve (AUC) and AUC, sensitivity (SEN), recall (R), specificity (SPE), precision (P), and F_1 score for each class of different method combination modes $C_1 \sim C_5$.

4. Comparison and Discussions

To further evaluate our model, several existing methods for this lymphoma dataset are used for comparison, as shown in Table 3. Method combination mode C_5 is used as a representative of our method since it has the best performance and shows better classification results than those of the state-of-the-art methods can be obtained. The current state-of-the-art method [15] proposed Fisher vector (FV) encoding combined with multiple types of local features. It applied a novel separation-guided dimension reduction

Table 2: Image level classification indices for different method combination modes.

Index		C_1	C_2	C_3	C_4	C_5
ACC	Overall	0.9558	0.9558	0.9823	0.9558	0.9912
AUC	Overall	0.9846	0.9971	0.9943	0.9894	0.9982
	CLL	0.9709	0.9903	0.9820	0.9875	0.9972
	FL	0.9987	1	1	0.9907	0.9960
	MCL	0.9971	0.9985	1	0.9912	1
SEN/R	CLL	0.9091	0.9697	0.9697	0.9697	1
	FL	0.9767	0.9302	0.9767	0.9429	0.9767
	MCL	0.9730	0.9730	1	0.9189	1
SPE	CLL	0.9750	0.9500	0.9875	0.9500	0.9875
	FL	0.9429	0.9714	0.9857	0.9429	1
	MCL	0.9474	0.9474	0.9737	0.9737	0.9868
P	CLL	0.9677	0.8889	0.9697	0.9412	0.9706
	FL	0.9767	1	1	0.9545	1
	MCL	0.9231	0.9730	0.9737	0.9414	1
F_1	CLL	0.9375	0.9275	0.9697	0.9552	0.9851
	FL	0.9767	0.9639	0.9882	0.9655	0.9882
	MCL	0.9474	0.9730	0.9867	0.9444	1

Table 3: Image level classification performance of our method and existing studies.

Index	Ours	[15]	[17]	[7]	[14]	[8]
ACC(%)	99.1	97.9	85	92.7	95.5	96.8
AUC	0.998	0.993	-	-	-	-

method and provides 97.9% accuracy and 0.993 AUC. Codella *et al.* [14] used descriptors including LBP and CNN based on segmented results. SVM was used as classifier and it obtained 95.5% accuracy. Song *et al.* [8] proposed a method to extract high-dimensional descriptor and subcategory discriminant transform (SDT) was used to enhance discriminative power of descriptors, which achieved the accuracy of 96.8%. Meng *et al.* [7] proposed a framework based on the Collateral Representative Subspace Projection Modeling (CRSPM) supervised classification model for histology image classification. Shamir *et al.* [17] presented an open source utility Wndchrm for standardized classification which provided accuracy of 85% for lymphoma classification. In our study, we achieve 99.1% accuracy and 0.998 AUC with method combination C_5.

Table 3 provides a quantitative comparison in terms of AUC and ACC values, which can be seen as complementary. The results demonstrate the advantages of our proposed model for lymphoma classification over the state-of-the-art approaches. Song *et al.* [15] calculated both ACC and AUC and achieved the best performance in existing approaches with 97.9% ACC, while we achieve 99.1% accuracy employing the C_5 combination mode with a similar AUC value. This suggests that the extracted feature sets and RF classifier in our study is more discriminative for malignant lymphoma pathological images.

According to Table 3, our method achieves better accuracy than all other existing methods. Shamir *et al.* [17] proposed solutions based on an open source platform for standardized support of different problems, including tissue age

differentiation, subcellular protein localization, lymphoma subtyping, and pollen grain distinction. It is not targeted so the performance for lymphoma is relatively low. Meng *et al.* [7] and Song *et al.* [8] extracted high-dimensional features and applied popular machine learning method; however, they have not considered deep learning techniques. According to our research, high-level features of CNN have promising contribution for final classification. Codella *et al.* [14] and Song *et al.* [15] both combined handcrafted and high-level features to perform classification. Codella *et al.* [14] applied a segmentation process which is time-consuming, and the segmentation-based features can easily be nondiscriminative due to imprecise segment results. Song *et al.* [15] utilized Fisher vector to encode the whole image, which resulted in low efficiency in feature extraction stage. In addition, Codella *et al.* [14] and Song *et al.* [15] performed classification with SVM, while our study experimentally confirmed that random forest classifier can achieve better performance for this classification task.

To sum up, accurate lymphoma classification is an important prerequisite for malignant lymphoma CADs. Our proposed classification method performs well in different image space transformations of original images, noting that blue ratio and Lab space amplify the texture and color characteristics of original images, respectively. Therefore, the subsequent feature extraction with customized feature sets specific to image spaces is meaningful. Classification is performed utilizing multiple hierarchies of local knowledge, which ensures the completeness of features and efficiency

of the whole approach. Patch based classification reduces the spatial complexity but sacrifices some local information, so the strategy for image level classification is investigated in detail for a remedy. Transfer learning is used to reduce consumption of time and it shows improved performance. Multiple paths are combined at different stages to fuse their characteristics sufficiently. Random forest classifier is employed, and it shows better performances than commonly used SVM.

5. Conclusions

This paper proposes a new CAD model for lymphoma classification on H&E stained histopathological images. This model includes three major steps, namely, image space transformation, feature extraction, and classification design. First, original images are transformed into different image spaces including blue ratio and Lab. Then two hierarchies of patches are cropped to perform a local-to-global classification. After that, specific feature set is extracted from corresponding image space. RF classifier is then trained for patch level classification based on handcrafted features. Meanwhile, transfer learning is introduced by utilization of a pretrained GoogLeNet model which also performs patch level classification. Finally, image level classification strategies are discussed and fusion at different levels is performed. The identification task is performed on IICBU Malignant Lymphoma Dataset. The experiment results and comparisons with the related work present that out proposed model can achieve better identification performances than others. We demonstrate the benefits of our multihierarchy classification scheme and the novel *DMW* method for weights computation of patch-level results, as well as the advantages of image space transformation.

In the future, we intend to apply the proposed model on more kinds of datasets to evaluate its generality when extended to different problems. The analysis of alternative for pretrained CNN can be performed to explore the potential of deep learning. Moreover, since the lymphoma sections in this study are all from lesions, a preceding lesion identification procedure can be designed for those datasets containing extra normal tissues. Then the proposed hierarchical strategy can be applied to the classification for precise identified lesions.

Data Availability

(1) The dataset includes three types of malignant lymphoma. (2) The data can be accessed from https://ome.grc.nia.nih .gov/iicbu2008/lymphoma/index.html. (3) In addition, there are no restrictions on data access.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] "A. C. Society, Cancer Statistics Center," https://cancerstatisticscenter.cancer.org.

[2] O. Sertel, J. Kong, U. V. Catalyurek, G. Lozanski, J. H. Saltz, and M. N. Gurcan, "Histopathological image analysis using model-based intermediate representations and color texture: follicular lymphoma grading," *Journal of Signal Processing Systems*, vol. 55, no. 1-3, pp. 169–183, 2008.

[3] J. Barker, A. Hoogi, A. Depeursinge, and D. L. Rubin, "Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles," *Medical Image Analysis*, vol. 30, pp. 60–71, 2016.

[4] K. Fukuma, V. B. Prasath, H. Kawanaka, B. J. Aronow, and H. Takase, "A study on feature extraction and disease stage classification for Glioma pathology images," in *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '16)*, pp. 2150–2156, Vancouver, BC, Canada, July 2016.

[5] J. Cao, Z. Qin, J. Jing, J. Chen, and T. Wan, "An automatic breast cancer grading method in histopathological images based on pixel-, object-, and semantic-level features," in *Proceedings of the IEEE 13th International Symposium on Biomedical Imaging (ISBI '16)*, pp. 1151–1154, Prague, Czech Republic, April 2016.

[6] J. Zhou, S. Lamichhane, G. Sterne, B. Ye, and H. Peng, "BIOCAT: a pattern recognition platform for customizable biological image classification and annotation," *BMC Bioinformatics*, vol. 14, p. 291, 2013.

[7] T. Meng, L. Lin, M.-L. Shyu, and S.-C. Chen, "Histology image classification using supervised classification and multimodal fusion," in *Proceedings of the IEEE International Symposium on Multimedia (ISM '10)*, 152, 145 pages, 2010.

[8] Y. Song, W. Cai, H. Huang, D. Feng, Y. Wang, and M. Chen, "Bioimage classification with subcategory discriminant transform of high dimensional visual descriptors," *BMC Bioinformatics*, vol. 17, no. 465, 2016.

[9] Y. Zhou, H. Chang, K. Barner, P. Spellman, and B. Parvin, "Classification of histology sections via multispectral convolutional sparse coding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 3081–3088, Columbus, OH, USA, June 2014.

[10] T. H. Vu, H. S. Mousavi, V. Monga, G. Rao, and U. K. A. Rao, "Histopathological image classification using discriminative feature-oriented dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 3, Article ID 2493530, pp. 738–751, 2016.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, Article ID 3065386, pp. 84–90, 2017.

[12] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "HEp-2 cell image classification with deep convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, Article ID 2526603, pp. 416–428, 2017.

[13] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[14] M. N. Gurcan, A. Madabhushi, N. Codella et al., "Lymphoma diagnosis in histopathology using a multi-stage visual learning approach," in *Proceedings of the SPIE Medical Imaging*, M. N.

Gurcan and A. Madabhushi, Eds., San Diego, Calif, United States.

[15] Y. Song, Q. Li, H. Huang, D. Feng, M. Chen, and W. Cai, "Low dimensional representation of fisher vectors for microscopy image classification," *IEEE Transactions on Medical Imaging*, vol. 36, no. 8, pp. 1636–1649, 2017.

[16] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, vol. 8 of *Color Research & Application*, John Wiley and Sons, New York, NY, USA, 2nd edition, 1983.

[17] L. Shamir, N. Orlov, D. M. Eckley, T. Macura, J. Johnston, and I. G. Goldberg, "Wndchrm—an open source utility for biological image analysis," *Source Code for Biology and Medicine*, vol. 3, no. 1, article 13, 2008.

[18] N. I. on Aging, "NIA Intramural Research Program Laboratory of Genetics," https://ome.grc.nia.nih.gov/iicbu2008/lymphoma/index.html.

[19] N. C. Institute, http://www.cancer.gov/.

[20] N. I. on Aging, https://www.nia.nih.gov/.

[21] L. Shamir, N. Orlov, D. M. Eckley, T. J. Macura, and I. G. Goldberg, "IICBU 2008: a proposed benchmark suite for biological image analysis," *Medical & Biological Engineering & Computing*, vol. 46, no. 9, pp. 943–947, 2008.

[22] H. Chang, J. Han, A. Borowsky et al., "Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association," *IEEE Transactions on Medical Imaging*, vol. 32, no. 4, pp. 670–682, 2013.

[23] O. Sertel, J. Kong, G. Lozanski et al., "Texture classification using nonlinear color quantization: application to histopathological image analysis," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 597–600, Las Vegas, NV, USA, 2008.

[24] T. Acharya and A. K. Ray, *Image Processing: Principles and Applications*, Wiley-Interscience, Hoboken, NJ, USA, 2005.

[25] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[26] S. Kaplan Berkaya, H. Gunduz, O. Ozsen, C. Akinlar, and S. Gunal, "On circular traffic sign detection and recognition," *Expert Systems with Applications*, vol. 48, pp. 67–75, 2016.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '15)*, Lecture Notes in Computer Science, pp. 234–241, Springer International Publishing, 2015.

[28] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 550–566, 2016.

[29] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *Proceedings of the IEEE 12th International Symposium on Biomedical Imaging (ISBI '15)*, pp. 294–297, Brooklyn, NY, USA, April 2015.

[30] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, pp. 2818–2826, Las Vegas, NV, USA, June 2016.

[32] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[33] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Proceedings of the International Conference on Image and Signal Processing (ICISP '08)*, vol. 5099 of *Lecture Notes in Computer Science*, 2008, pp. 236–243.