JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

JSCDM

Vol. 6 No. 1 (2025) 86-94
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Intelligent Pathological Diagnosis of Melanocytic Lesions Using Deep Learning

Qian Bian[1,2], Jiayi Zhang[3], ELcid A.Serrano[1]*

[1] *School of Graduate Studies and School of IT, Mapúa University,*
658 Muralla St, Manila, 1002, PHILIPPINES

[2] *Xi'an Siyuan University,*
No 28 Shui'an Road, Xi'an, 710038, CHINA

[3] *Suzhou Institute of Biomedical Engineering and Technology,*
No 88 Keling Road, Suzhou, 215163, CHINA

*Corresponding Author: easerrano@mapua.edu.ph

DOI: https://doi.org/10.30880/jscdm.2025.06.01.006

Article Info

Received: 28 June 2024
Accepted: 25 February 2025
Available online: 30 June 2025

Keywords

Melanocytic lesions, pathological diagnosis, deep learning, stain normalization

Abstract

Melanocytic lesions occur on the surface of the skin, and melanoma is a malignant type of melanocytic lesion associated with a high mortality rate, posing a serious threat to human health. Histopathological analysis remains the gold standard for diagnosing melanocytic lesions. In this study, a fully automated intelligent diagnosis method based on deep learning is proposed to classify pathological whole slide images (WSIs) of melanocytic lesions. First, color normalization using a CycleGAN neural network was performed on multi-center pathological WSIs. Second, a ResNet-152 neural network-based deep convolutional network prediction model was built using 745 WSIs. Third, a decision fusion model was cascaded to calculate the average prediction probability for each WSI. Finally, the diagnostic performance of the proposed method was verified using internal and external test sets containing 182 and 54 WSIs, respectively. Experimental results showed that the overall diagnostic accuracy of the proposed method reached 94.12% on the internal test set and exceeded 90% on the external test set. Furthermore, the adopted color normalization method was superior to the traditional color statistics-based and staining separation-based methods in terms of structure preservation and artifact suppression. These results demonstrate that the proposed method can achieve high precision and strong robustness in classifying pathological WSIs of melanocytic lesions, highlighting its potential in promoting the clinical application of computer-aided pathological diagnosis.

1. Introduction

Skin cancer is one of the most common cancers in the world. Melanoma is a type of skin cancer that is highly invasive and malignant. It easily causes lymph node and blood metastasis in the early stage of the disease, and the fatality rate is very high (about 80%). It is also classified as a malignant melanocytic lesion, and it can be of atypical and benign types [1]. In clinical practice, pathological diagnosis stands as the gold standard for confirming melanocytic lesions [2]. However, this process is subjective, with inconsistent diagnosis rates between benign and malignant lesions, with the variability being as high as 45.5% [3] and usually higher for atypical lesions. In recent years, the advancement of whole-slide scanning techniques has propelled the development of digital pathology. Artificial intelligence and, notably, deep learning (DL) have shown remarkable performance in computer-aided pathological diagnosis [4]. As the demand for computer-aided

This is an open access article under the CC BY-NC-SA 4.0 license.

pathological diagnosis grows, variations in stain styles among different medical centers pose new challenges to the reliability and generalizability of diagnostic models [5]. This study used whole-slide images (WSIs) and DL to construct an intelligent pathological diagnosis model for melanocytic lesions (malignant, atypical, and benign).

2. Related Work

With respect to intelligent pathological diagnosis of melanocytic lesions, Hekler et al. [6], [7] proposed an intelligent pathological diagnosis model for melanocytic lesions based on ResNet-50 and achieved higher benign and malignant classification accuracy than 11 pathologists (68% vs. 59.2%). On this basis, Brinker et al. [8] constructed an intelligent pathological diagnosis model for benign and malignant melanocytic lesions based on ResNeXt-50 by improving ResNet-50, and the accuracy was comparable to that of 18 senior pathologists (88.0% vs. 90.3%). Li et al. [9] proposed an intelligent pathological diagnosis model with superior classification performance by designing a novel convolutional neural network, which achieved the highest accuracy (92.0%) reported so far among all intelligent pathological diagnosis models of benign and malignant melanocytic lesions.

The results of the above studies show that accurate pathological diagnosis of benign and malignant melanocytic lesions can be achieved using DL. However, the ability to discriminate atypical melanocytic lesions, which is precisely a relative shortcoming of pathological diagnosis, remains to be investigated. In addition, the pathological WSI data used in the above studies were all from patients with melanocytic lesions diagnosed within the same year at the same medical center, and the stain consistency of these pathological slides was good [10]. However, the preparation process of pathological slides is complex, and stain variability among pathological WSIs from different medical centers is high. Therefore, it is a substantial challenge to construct a generalized intelligent pathological diagnosis model for melanocytic lesions based on DL [11], [12]. Ianni et al. [13] proposed a stain normalization method for suppressing multi-center pathological WSI stain variability based on a self-coded convolutional neural network, which effectively equalizes the stain styles of pathological images from multiple centers to the same level, but the method requires paired pathological WSI data, that is, pathological WSI data of the same tissue slide captured by two scanners, which is not practical for actual clinical diagnosis.

In summary, the current studies on computer-aided pathological diagnosis of melanocytic lesions based on WSIs have the following problems. First, there is insufficient research on the differentiation of atypical melanocytic lesions. Clinically, patients with atypical melanocytic lesions have different surgical treatment plans than those with benign or malignant melanocytic lesions. However, the histological patterns and biological characteristics of atypical melanocytic lesions partially overlap with those of benign and malignant melanocytic lesions, leading to frequent confusion. Therefore, precise differentiation of atypical melanocytic lesions is of substantial clinical significance. Second, there is a lack of studies on suppressing stain variations in multi-center pathological WSIs. Many studies have used pathological WSI data from patients with melanocytic lesions within the same year and from the same medical center, in which stain consistency is relatively good. The diagnosis models built on such data lack generalizability and cannot be used for classifying pathological WSIs with substantial stain variations. Although some scholars have proposed DL-based stain normalization methods, they rely on paired pathological WSI data, which cannot meet the actual clinical diagnostic needs.

3. Methods

3.1 Method Architecture

This study proposes an intelligent pathological diagnosis method for melanocytic lesions based on DL, comprising four modules (Fig. 1): Patch generation, stain normalization, DL prediction, and aggregation of prediction results.

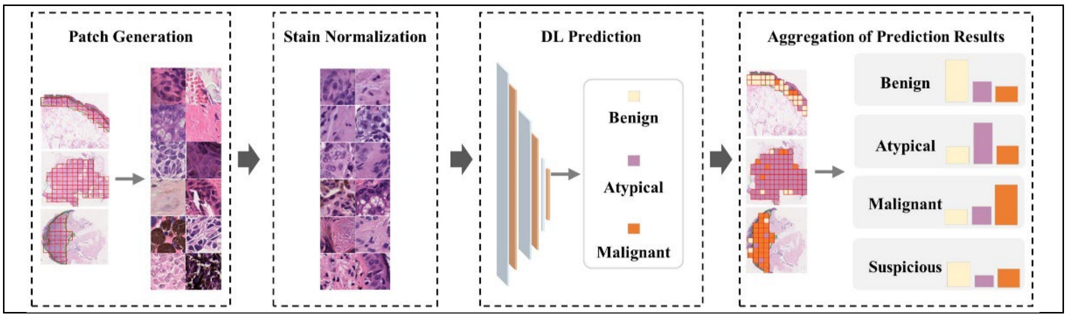


Fig. 1 *Our framework for an intelligent pathological diagnosis method for melanocytic lesions*

3.2 Patch Generation

Constrained by hardware computing resources, DL methods cannot directly handle WSIs containing hundreds of millions of pixels. Therefore, in this study, the annotated information of WSIs was used to extract image patches (224 × 224 pixels) from the annotated regions in a non-overlapping manner, and the image patches with less than 60% effective tissues were discarded. In order to ensure the balance of the image patch dataset for training the intelligent pathological diagnosis model, this study randomly extracted image patches from the malignant, atypical, and benign melanocytic lesion WSIs in the training set at the ratio of 1:3:1. Finally, 20,000 samples were randomly sampled for each category to balance the number of image patches for the three categories of melanocytic lesions.

3.3 Stain Normalization

Inspired by the study of Shaban et al. [14], this study framed stain normalization of pathological images as a color style transfer problem and designed a stain normalization method, CycleGAN-Stain, based on CycleGAN. The method uses a K-means color clustering strategy to generate two sets of pathological images: Domain A (with diverse stain styles) and Domain B (with relatively uniform stain styles). Then, CycleGAN was used to convert the image styles of Domain A to Domain B.

Due to the long sampling time span (from 2001 to 2018) of the slides from Center 1, there were large differences in the stain styles between the WSIs. A proportion of the image patches from Center 1 was randomly selected to form Domain A, which was characterized by different stain styles. K-means clustering was then applied to analyze these differences in stain styles further. K-means stain style clustering was used to establish Domain B with relatively consistent stain styles. The specific steps were as follows.

First, the stain concentration matrix $A = [a]_{ij}$ of each image patch in Domain A was calculated, where $i = 3$ denotes the three channels R, G, and B; and $j = 2$ denotes the two stains of H and E. That is, a_{ij} is the concentration of stain j in channel i. This study calculated the mean value of the three channels of each stain for visualization and mapped an image patch into a two-dimensional feature (a_H, a_E) (Fig. 2). Then, these features were divided into five different stain styles by the K-means clustering method (k = 5). We set the image patches corresponding to the class of features with the most concentrated styles as the Domain B dataset.

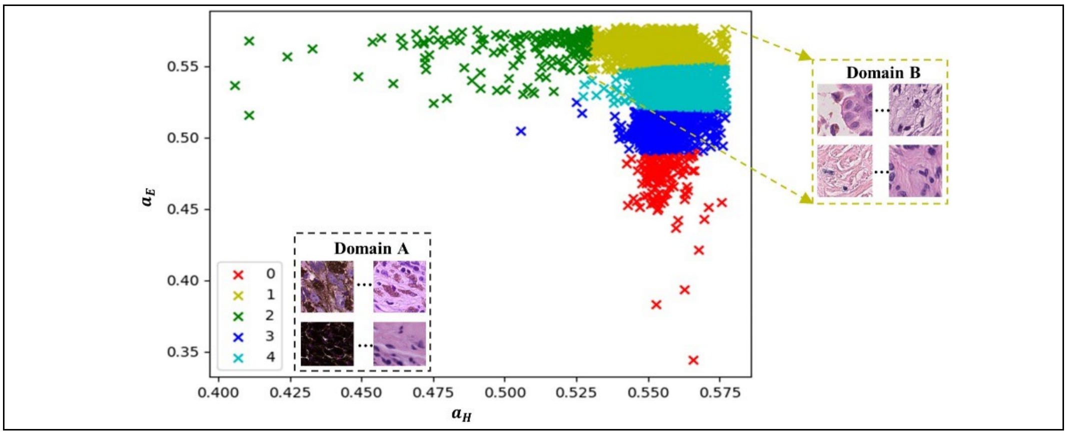


Fig. 2 *K-means stain style clustering visualization*

The architecture of CycleGAN-Stain is shown in Fig. 3. It comprises two cyclic networks: Domain A → Domain B (forward cycle, green solid arrows) and Domain B → Domain A (reverse cycle, red solid arrows). In the forward loop, the cycle consistency loss between a and $\hat{a}$ is computed as follows:

$$L_{cycle}(G_A, G_B, A) = \mathbb{E}[\left\| a - G_B\left(G_A(a)\right)\right\|_1] \tag{1}$$

The discriminator D_B is used to determine the truth of the real image b and generated image $\hat{b}$ in Domain B, and the adversarial loss in the forward cycle is defined as follows:

$$L_{adv}(G_A, G_B, A, B) = \mathbb{E}[log D_B(b)] + \mathbb{E}\left[log\left(1 - D_B\left(G_A(a)\right)\right)\right] \tag{2}$$

The cycle consistency loss function and the adversarial loss function in the reverse cycle are calculated in a similar way to the forward cycle, and the overall loss function is the sum of the four losses. The generator and discriminator use ResNet and a 70 × 70 PatchGAN network, respectively.

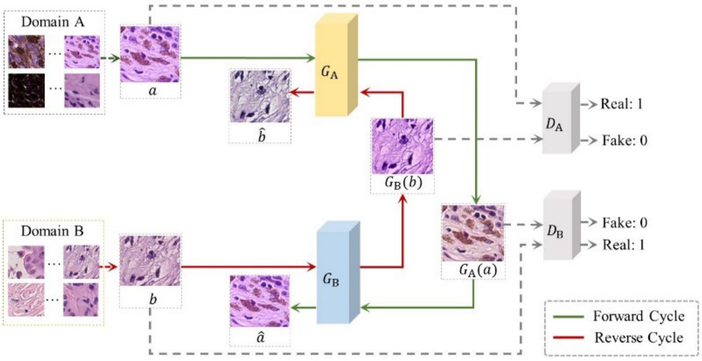


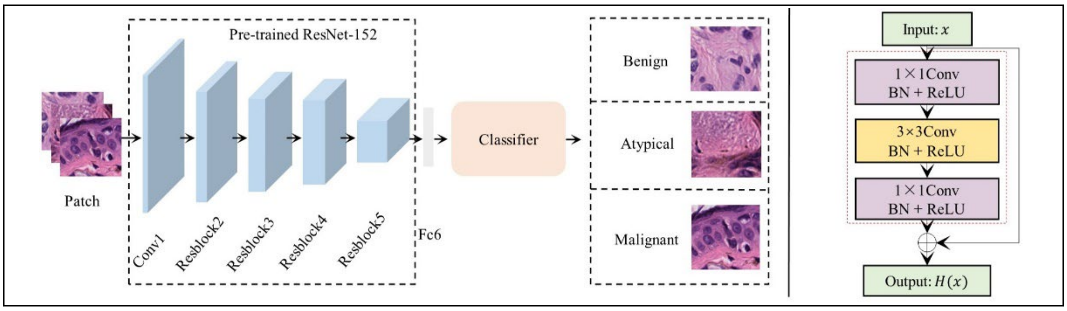
Fig. 3 *Architecture of CycleGAN-Stain*

3.4 Deep Learning Prediction

Given the high similarity of the histopathologic images in terms of the morphological structure and texture features, shallow convolutional neural networks have limited ability to extract complex image features, and it is difficult to uncover deep information about the pathological images. If network depth is increased to tackle this issue, gradient disappearance and network performance degradation issues will be encountered. For this reason, this study adopted the ResNet-152 architecture (Fig. 4a) proposed by He et al. [15] to construct a DL prediction module.

Compared with traditional convolutional neural networks, the main feature of ResNet-152 is the introduction of residual units (Fig. 4b), which are composed of the convolution (Conv), batch normalization (BN), and ReLU activation functions. By adding identity mapping, the residual mapping function $F(x) = H(x) - x$ is fitted, which mitigates the problems of gradient disappearance and network performance degradation, accelerates the convergence speed of the network's training, and dramatically improves the recognition ability of deep networks.

The input of the network is a sample of an image patch processed by the stain normalization method, with a size of $224 \times 224 \times 3$. First, after the convolution layer (Conv1), the input image patch is subjected to conventional feature extraction to reduce the feature size, and then, higher-level features are extracted by four residual blocks (Resblock 2–5), followed by inputting the extracted high-dimensional features into the fully connected layer (Fc6) for the classification output. The classifier in Fig. 4a is a SoftMax classifier, which ultimately outputs the probability of three melanocytic lesion types (malignant, atypical, and benign) for each image patch.



(a) **(b)**

Fig. 4 *Network architecture of the DL prediction module*

3.5 Aggregation of the Prediction Results

To diagnose a patient's type of melanocytic lesion, this study adopted a decision fusion strategy to aggregate the prediction results of all image patches of each melanocytic lesion patient. The predicted probabilities of the three lesion types of all image patches in each WSI were calculated by the statistical mean, which was considered the predicted probability of the three lesion types of the WSI. This was inspired by the clinical way through which pathologists tentatively classify cases with doubtful diagnosis of the microscopic pathological slides as suspicious cases [16] and subsequently integrate the patient's clinical information for further diagnosis. In this study, we determined the WSIs for which the predicted probabilities of the three lesion types were lower than 0.6 as suspicious types. We categorized the melanocytic lesion patients whose all WSIs were suspicious as suspicious cases, which required further diagnosis by a senior pathologist in combination with a review of the other clinical information of the patient. Excluding the suspicious cases, patients' WSIs were judged according to the highest degree of malignancy among all WSIs, in the order of malignant > atypical > benign.

3.6 Staining Normalization Comparison Method

To validate the effectiveness of the proposed stain normalization method, a comparative analysis was conducted between CycleGAN-Stain and two other histopathological image stain normalization methods: the color-based method [17] and stain-based method [18]. Following this, the classification performance for the melanocytic lesion pathological WSIs was compared across these methods.

The color-based normalization method [17] begins with converting both the unnormalized image and a template image from the RGB color space to the orthogonal lab color space. The mean and standard deviation of the pixel intensities within the template image were computed. A per-pixel least squares adjustment was then applied to the target image to revert the image to the RGB color space.

In the stain-based normalization method [18], by combining Beer–Lambert's spectral absorption law [19]. The pathological images are converted from the RGB color space to the optical density space [19]. This operation separates the stain concentration matrix A and the absorbance coefficient matrix C of both the target image and a template image. Subsequently, the stain concentration matrix A of the template image replaces that of the unnormalized image, and an inverse Beer–Lambert transformation is implemented. The mathematical expression for Beer–Lambert transformation is as follows:

$$OD = -\,ln\left(\frac{I}{I_0}\right) = A \cdot C, \tag{3}$$

where $I \in \mathbb{R}^{m\times n}$ denotes the transmitted light intensity of the pixels, with n denoting the number of pixels in the pathological image, and $m = 3$ denoting the three color channels (red, green, and blue). I_0, generally assumed as 255, denotes the incident light intensity. $A \in \mathbb{R}^{m\times r}$ denotes the stain concentration matrix; $r = 2$ denotes the H and E stains; and $C \in \mathbb{R}^{r\times n}$ denotes the absorbance coefficient matrix of the stains.

3.7 Model Evaluation Criteria

To compare the performance of the stain normalization methods, this study adopted the structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) metrics for quantitative evaluation. The larger their value, the better the stain normalization method is considered to maintain the consistency of the image structure. Since different stain styles of the images before and after stain normalization affect the SSIM and PSNR, grayscale transformation was performed first, and then, the SSIM and PSNR are calculated. The calculation of the SSIM and PSNR was as follows:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - I^{norm}(i,j)]^2\,, \tag{4}$$

$$PSNR = 10 \cdot log_{10}\left(\frac{MAX_I^2}{MSE}\right), \tag{5}$$

$$SSIM(I, I^{norm}) = \frac{(2\mu_1\mu_2+C_1)(\sigma_{12}+C_2)}{(\mu_1^2+\mu_2^2+C_1)(\sigma_1^2+\sigma_2^2+C_2)}, \tag{6}$$

where I denotes an unnormalized image of size $m \times n$, and I^{norm} denotes the normalized image. MAX_I^2 is the maximum pixel value of image I, which is usually 255. μ_1 and μ_2 denote the pixel means of I and I^{norm}, respectively. σ_1 and σ_2 are the standard deviations of I and I^{norm}, respectively. σ_{12} is the covariance of I and I^{norm}. C_1 and C_2 are constants to avoid division by zero.

To evaluate the classification performance of the diagnosis models, this study used the classification accuracy (ACC) and macro F1 score for quantitative evaluation. The higher their values, the better is the diagnosis performance of the model. The calculation of the ACC and macro F1 score was as follows:

$$Acc = \frac{TP+TN}{TP+FP+TN+FN}, \tag{7}$$

$$Recall = \frac{TP}{TP+FN}, \tag{8}$$

$$Precision = \frac{TP}{TP+FP}, \tag{9}$$

$$F1\ score = 2\frac{Recall \times Precision}{Recall + Precision}, \tag{10}$$

$$macro\ F1 = \overline{F1\ score}, \tag{11}$$

where true positive (TP) represents the number of positive samples correctly classified; false positive (FP) represents the number of negative samples misclassified as positive; true negative (TN) represents the number of negative samples correctly classified; and false negative (FN) represents the number of positive samples misclassified as negative.

4. Results

4.1 Data Sources

A total of 218 malignant, 119 atypical, and 374 benign melanocytic lesion patients diagnosed between 2001 and 2018 were enrolled from the Shanghai Jiao Tong University School of Medicine Affiliated Ninth People's Hospital ("Center 1"), Shanghai Jiao Tong University School of Medicine Affiliated Ninth People's Hospital North Branch ("Center 2"), and Shanghai First People's Hospital Baoshan Branch ("Center 3"). The hematoxylin & eosin (H&E) stain slides of all patients were retrieved. All of them had definitive diagnosis results (benign, atypical, or malignant), mainly determined by pathologists through microscopic observation of cellular structures and morphology. If the H&E diagnosis was uncertain, the patients' clinical manifestations and immunohistochemical/molecular detection results were combined for the diagnosis. All H&E slides included in this study were re-evaluated by a senior pathologist with 30 years of pathological diagnosis experience in dermatologic diseases.

All H&E slides were scanned by a Hamamatsu NanoZoomerS60 scanner at 40× magnification, generating 457 malignant, 142 atypical, and 382 benign melanocytic lesion WSIs. All WSIs were annotated by two pathologists with eight and 15 years of experience in pathological diagnosis of dermatologic diseases. Two pathologists used the medical image processing software NDP.view2 (version 2.6.13) to annotate each lesion region on the premise that the lesion types of the WSI were known. To resolve inconsistencies in the annotation results, the pathologists reached a consensus by considering the patient's clinical information (gender, age, anatomical location, etc.) and through group discussion.

The WSIs of the patients with melanocytic lesions in Center 1 were divided into a training set and internal testing set using random stratified sampling at a ratio of 8:2, and the WSIs of the patients with melanocytic lesions in Center 2 and Center 3 were used as an external testing set (Table 1).

Table 1 *Number of patients and WSIs in the training and testing sets*

Lesion Type	Training Set		Internal Testing Set		External Testing Set	
	Patients	WSIs	Patients	WSIs	Patients	WSIs
Benign	282	290	71	71	21	21
Atypical	88	107	23	26	8	9
Malignant	162	348	42	85	14	24
Total	532	745	136	182	43	54

4.2 Experimental Environment and Hyperparameter Settings

The intelligent pathological diagnosis model was built using the open-source Python machine learning library PyTorch and an NVIDIA GTX 2080Ti GPU. The template image (Fig. 5a) for the color-based and stain-based stain normalization methods was selected by the pathologists based on the staining effect of the pathological tissues. Through experimental analysis, the network training hyperparameter settings for the CycleGAN-Stain stain normalization method and the DL prediction module were determined; they are shown in Table 2. The DL prediction module adopted a multi-class cross-entropy loss function and initialized the ResNet-152 network with pre-trained weights on the ImageNet dataset.

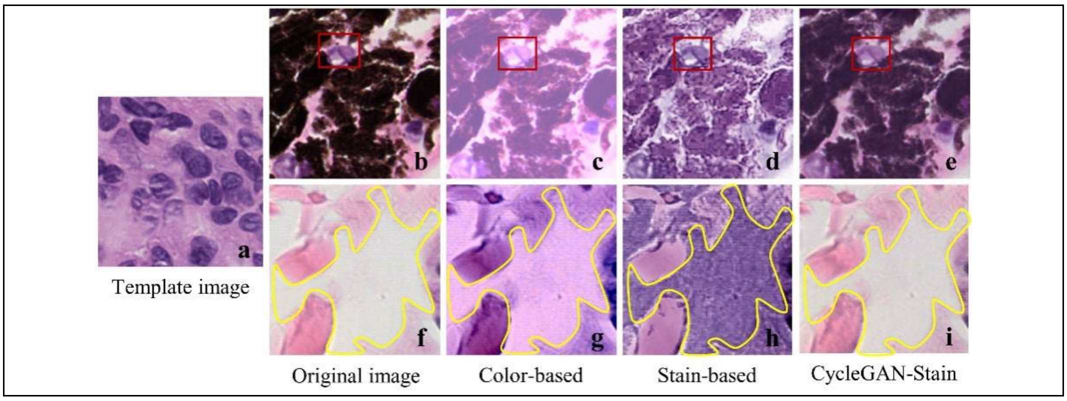


Fig. 5 *Stain normalization results of the pathology image*

Table 2 *Intelligent pathological diagnosis model training hyperparameter settings*

Network Architecture	Optimizer	Learning Rate	Batch Size	Epoch
CycleGAN	Adam	0.0002	2	200
ResNet-152	Adam	0.00001	16	400

4.3 Comparison of Stain Normalization and Disease Diagnosis Performance

This study randomly extracted 300 unnormalized image patches from the dataset and processed them using three stain normalization methods. Finally, the mean SSIM and PSNR values were calculated for the 300 pairs of image patches before and after stain normalization, as shown in Table 3.

Table 3 *Image quality evaluation after stain normalization*

Stain Normalization Method	Color-based Method	Stain-based Method	CycleGAN-Stain Method
SSIM	0.902	0.880	**0.963**
PSNR (dB)	20.024	22.821	**32.622**

To visually demonstrate the stain normalization capabilities of the different methods, two pathological image patches of melanocytic lesions and their stain normalization results using the three methods were randomly selected, as shown in Fig. 5.

The ACC and macro F1 scores of the intelligent pathological diagnosis model on the unnormalized dataset and the datasets processed by the three stain normalization methods are shown in Table 4.

Table 4 *Comparison of the intelligent pathological diagnosis results*

Stain Normalization Method	Internal Testing Set		External Testing Set	
	ACC	Macro FI	ACC	Macro FI
Unnormalized	93.38%	0.933	76.74%	0.843
Color-based Method	91.18%	0.889	88.37%	0.901
Stain-based Method	92.65%	0.902	88.37%	0.859
CycleGAN-Stain Method	94.12%	0.938	90.70%	0.949

The results in Table 3 show that the CycleGAN-Stain method excels in preserving the structural consistency of the images. As shown in Fig. 5b, the tissues within the red rectangular region differ in stain properties from the surrounding tissues due to them being distinct types, leading to varied stain outcomes. The CycleGAN-Stain method effectively retains the color differences caused by tissue variation (Fig. 5e), whereas the color-based and stain-based methods lose this critical discriminatory stain information, resulting in over-normalization (Fig. 5c and Fig. 5d). Figs. 5f, 5g, and 5h together illustrate that the color-based and stain-based methods introduce normalization artefacts (in the yellow regions) that incorrectly convert the non-diagnostic background regions into diagnostic tissue regions. This directly leads to erroneous diagnostic evaluations. Overall, CycleGAN-Stain surpasses the color-based and stain-based methods in maintaining the consistency of the histological structures and suppressing the artefacts.

Moreover, CycleGAN-Stain performs style mapping between the image domains rather than between images, which can maximally preserve all pathological tissue structural information, thus laying the foundation for building an intelligent pathological diagnosis model with strong generalization performance. As shown in Table 4, the intelligent pathological diagnosis model using the CycleGAN-Stain method has the strongest diagnostic performance, with ACC exceeding 90.00% on the external testing set. Of note, the ResNet-152 network used in this study is a commonly used architecture for medical image processing tasks, with solid feature extraction capabilities. It can fully use the morphological features that have low discriminability and invisibility to the naked eye while classifying melanocytic lesions, thus providing strong support for the precise diagnosis of melanocytic lesions. The decision fusion strategy that this study employed allowed the input to be the pathological WSI data of melanocytic lesion patients and the output to be the lesion type of the melanocytic lesion patients, achieving an end-to-end intelligent pathological diagnosis of melanocytic lesions.

However, a limitation of this study is that the external testing set in this study comprised a relatively small dataset, and the intelligent diagnosis model was trained solely on patients' histological slide data. Future studies should aim to collect more melanocytic lesion patient data from multiple medical centers to validate the diagnostic performance of our intelligent model and explore methods for integrating clinical data into the construction of such models in order to enhance diagnostic intelligence and practical applicability.

5. Conclusion

This study proposed an intelligent pathological diagnosis method based on DL for classifying melanocytic lesions. Our CycleGAN-Stain method for melanocytic pathological WSI stain normalization eliminates differences in the stain styles between different WSIs. Then, the melanocytic lesion classification network that we constructed based on the ResNet-152 architecture and a decision fusion strategy that we employed to aggregate the melanocytic lesion types of all WSIs for each patient helped obtain the final diagnosis result of the patient. We anticipate that this approach will enhance the diagnostic efficiency of melanocytic lesions.

Acknowledgement

This work was supported by the Shaanxi Province Education and Teaching Reform Key Project (Grant Number 23BZ082). We thank LetPub (www.letpub.com.cn) for its linguistic assistance during the preparation of this manuscript.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm the contributions to the paper as follows. ***Study conception and design:*** *Qian Bian, Jiayi Zhang, and ELcid A. Serrano;* ***data collection:*** *Qian Bian;* ***analysis and interpretation of results:*** *Qian Bian and ELcid A. Serrano;* ***draft manuscript preparation:*** *Qian Bian, Jiayi Zhang, and ELcid A. Serrano. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Piepkorn, M. W., Barnhill, R. L., Elder, D. E., Knezevich, S. R., Carney, P. A., Reisch, L. M., & Elmore, J. G. (2014). The MPATH-Dx reporting schema for melanocytic proliferations and melanoma. *Journal of the American Academy of Dermatology, 70*(1), 131-141. doi: 10.1016/j.jaad.2013.07.027.

[2] Sarah, H., Roman, C. M., Achim, H., Jochen, S. U., Catarina, B., Raymond, L. B., . . . Titus, J. B. (2021). Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. *European Journal of Cancer.* doi: 10.1016/j.ejca.2021.06.049.

[3] Saurabh, L., Sarika, S., Julide, T. C., & David, N. S. (2008). Discordance in the histopathologic diagnosis of difficult melanocytic neoplasms in the clinical setting. Journal of Cutaneous Pathology. doi: 10.1111/j.1600-0560.2007.00970.x.

[4] TJ, B. M., Schmitt EI, Krieghoff-Henning R. (2021). Diagnostic performance of artificial intelligence for histologic melanoma recognition compared to 18 international expert pathologists. Journal of the American Academy of Dermatology. doi: 10.1016/J.JAAD.2021.02.009.

[5] Jeroen van der, L., Geert, L., & Francesco, C. (2021). Deep learning in histopathology: the path to the clinic. Nature Medicine. doi: 10.1038/s41591-021-01343-4.

[6] Brinker et al., 2022Hekler, A., Utikal, J. S., Enk, A. H., Berking, C., Klode, J., Schadendorf, D., . . . Krahl, D. (2019). Pathologist-level classification of histopathological melanoma images with deep neural networks. European Journal of Cancer, 115, 79-83.

[7] Hekler, A., Utikal, J. S., Enk, A. H., Solass, W., Schmitt, M., Klode, J., . . . Bestvater, F. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. European Journal of Cancer, 118, 91-96.

[8] Brinker, T. J., Schmitt, M., Krieghoff-Henning, E. I., Barnhill, R., Beltraminelli, H., Braun, S. A., . . . Fraitag, S. (2022). Diagnostic performance of artificial intelligence for histologic melanoma recognition compared to 18 international expert pathologists. Journal of the American Academy of Dermatology, 86(3), 640-642.

[9] Li, T., Li, F., liu, J., & Zuo, K. (2022). Pathologist-Level classification of melanoma disease pathologies using a convolutional neural network: A retrospective study of Chinese. Paper presented at the International Conference on Image, Vision and Intelligent Systems (ICIVIS 2021).

[10] Bejnordi, B. E., Litjens, G., Timofeeva, N., Otte-Höller, I., Homeyer, A., Karssemeijer, N., & Van Der Laak, J. A. (2015). Stain specific standardization of whole-slide histopathological images. IEEE transactions on medical imaging, 35(2), 404-415.

[11] Howard, F. M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., . . . Kather, J. N. (2021). The impact of site-specific digital histology signatures on deep learning model accuracy and bias. Nature communications, 12(1), 4423.

[12] Stiff, K. M., Franklin, M. J., Zhou, Y., Madabhushi, A., & Knackstedt, T. J. (2022). Artificial intelligence and melanoma: A comprehensive review of clinical, dermoscopic, and histologic applications. Pigment Cell & Melanoma Research, 35(2), 203-211.

[13] Ianni, J. D., Soans, R. E., Sankarapandian, S., Chamarthi, R. V., Ayyagari, D., Olsen, T. G., . . . Cockerell, C. J. (2020). Tailored for real-world: a whole slide image classification system validated on uncurated multi-site data emulating the prospective pathology workload. Scientific Reports, 10(1), 3217.

[14] Shaban, M. T., Baur, C., Navab, N., & Albarqouni, S. (2019). *Staingan: Stain Style Transfer for Digital Histological Images.* Paper presented at the International Symposium on Biomedical Imaging.

[15] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

[16] Julia, H., Eva, K.-H., Tanja, B. J., Christof von, K., Jochen, S. U., Friedegund, M., . . . Titus, J. B. (2021). Combining CNN-based histologic whole slide image analysis and patient data to improve skin cancer classification. *European Journal of Cancer*. doi: 10.1016/j.ejca.2021.02.032

[17] Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. IEEE Computer Graphics and Applications. doi: 10.1109/38.946629

[18] Abhishek, V., Tingying, P., Amit, S., Shadi, A., Lichao, W., Maximilian, B., . . . Nassir, N. (2016). Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images. *IEEE Transactions on Medical Imaging*. doi: 10.1109/tmi.2016.2529665.

[19] Ruifrok, A. C., & Johnston, D. A. (2001). Quantification of histochemical staining by color deconvolution. Analytical and quantitative cytology and histology.