



Deep Learning for Lymphoma Detection on Microscopic Images

Ammar Ammar¹(✉), Irfan Tito Kurniawan¹, Resfyanti Nur Azizah¹, Hafizh Rahmatdianto Yusuf¹, Antonius Eko Nugroho¹, Ghani Faliq Mufiddin¹, Isa Anshori¹, Widyawardana Adiprawita¹, Hermin Aminah Usman², and Okky Husain²

¹ Biomedical Engineering, Bandung Institute of Technology, Bandung, Indonesia
ammam.chalifah@gmail.com

² Pathological Anatomy, Padjajaran University, Hasan Sadikin Education Hospital, Bandung, Indonesia

Abstract. Early lymphoma diagnosis is essential to improve the patients' survival rate and avoid irreversible damage. Immunohistochemistry-based lymphoma diagnostics is an expensive and time-consuming process, especially in developing countries with limited resources. Image-based lymphoma diagnostics might serve as an inexpensive, yet less accurate alternative to immunohistochemistry-based methods. One challenge in image-based methods is that carcinoma can occur in the same organ as lymphoma, thus making it hard to differentiate the two types of cancer. To assist lymphoma diagnostics, this study proposes a deep learning-based method to classify nasopharyngeal microscopic biopsy images into one of three classes: lymphoma, carcinoma, and benign lesion. The method works by splitting the images into patches, classifying each patch using a deep learning model, and taking the average confidence score of each patch. We compared three deep learning-based feature extractor architectures and studied the effects of three image color preprocessing techniques on classification performance. We reached 88.7% sensitivity and 91.3% specificity in differentiating lymphoma on 400x magnification CLAHE-enhanced microscopic images using the InceptionResNetV2 model. We also reached 87.0% three-class classification accuracy using the same model.

Keywords: Lymphoma · InceptionResNetV2 · Deep learning

1 Introduction

Lymphoma is a neoplasm or abnormal tissue that grows on lymph nodes and lymph tissues. Generally, there are two kinds of lymphoma: Hodgkin Lymphoma (HL) and non-Hodgkin Lymphoma (NHL) [1]. In HL cases, neoplasm growth occurs localized in lymphoid cells. This disease also has a relatively high survival rate [2]. On the contrary, neoplasm growth in NHL happens in the majority of lymphoid tissues. NHL can propagate to lymphadenopathy surface, digestive system, and even central nervous system

[3]. According to the Global Cancer Observatory (GLOBOCAN) 2018 data, the number of new cases of NHL lymphoma reached 14,164, ranked seventh in the standings of new cancer cases in Indonesia. The death rate caused by NHL lymphoma was also ranked seventh in the cancer death toll, with a mortality of 565 [4]. A study on 25 NHL patients held at RSUP Sanglah Denpasar in 2014 showed that 68% of NHL patients had a poor prognosis with a two-year and five-years life expectancy of 34% and 26% respectively. This low life expectancy was caused by the patient's non-compliance to the treatment, the availability of chemotherapy drugs, poor hygiene, and inadequate observation time of the NHL cases [5].

National guidelines for cancer treatment state that the procedure for diagnosing lymphoma is based on anamnesis, physical observation, and supportive observation [6]. Supportive observation in histopathology and immunohistochemistry was performed to determine the further appropriate treatment for the patient [7]. Histopathology examination was performed by biopsy staining using haematoxylin and eosin stain or H&E stain [7]. The samples were analyzed on the architectural, cellular, and subcellular levels, and it is crucial to obtain cell morphology such as epithelium, stroma, and lymphoid from segmentation on an architectural level so that the malignancy of lymphoma can be determined considering the quantity, density, diameter, shape, and chromatism distribution of the germinal center and also area, thickness, and chromatism of the marginal zone. Malignancy of lymphoma is also judged by the ratio of germinal center and marginal zone area.

In some cases, those parameters are biased, so malignant lymphoma looks like benign follicular hyperplasia [8]. In addition to the benign lesion, carcinoma also can not be ignored. Carcinoma is the most common cancer which begins in epithelial tissue [9]. Most of the organs in the respiratory tract have an epithelial layer and adenoid layer, which is part of the lymph system [10]. On nasopharyngeal tissue, carcinoma is often confused with lymphoma due to its rather similar growth pattern [11]. However, synchronous malignancy of lymphoma and carcinoma in the head and neck region is extremely rare [12]. Thus, the classification system must classify the H&E stain images into three classes: benign lesions, carcinoma, and lymphoma.

As such, immunohistochemistry or IHC was performed to gain a better diagnosis. However, the IHC method for lymphoma diagnosis requires various tools that are not widely accessible in Indonesia. Generally, the IHC method of diagnosis is available only in Type A hospitals. An alternative approach to solve this issue is by improving the quality of histopathology screening diagnosis based on sample images by utilizing an automated system to make an interpretable classification. This approach may tackle the hardships in identifying fine-grained morphological features [8] in histopathology examination. Feature identification is a type of pattern recognition, a task proven to be solvable by machines accurately. Furthermore, this alternative solution may reduce the volume of samples previously screened by immunohistochemistry, saving only precious resources for necessary cases.

Deep learning [13], is a machine learning method that uses a large amount of data to train a deep neural network to recognize patterns and acquire predictive properties. One of the deep learning architectures, the convolutional neural network (CNN), is an architecture that boosts deep learning's performance in image-related fields [14].

CNN-based deep learning models are already developed to solve various medical image classifications, i.e., melanoma classification [15], breast cancer, diabetic retinopathy classification [16–18], and lymphoma classification itself [8].

A deep learning-based system can decide whether the morphology change of the lymph cells recorded on the WSI is classified as follicular lymphoma (FL) or follicular hyperplasia (FH) [8]. This study processes H&E-stained images in various resolutions using a Bayesian Neural Network (BNN). The system starts the classification process by breaking down the images into small patches of lower resolutions. The system will then classify each patch separately as FL or FH. This method can perform automatic end-to-end lymphoma classification with an accuracy of 91% and an area under the curve of 0.99 [8]. A study by H. Miyoshi et al. shows that a deep learning-based classifier can perform lymphoma diagnosis with an accuracy of up to 97.0%, higher and faster than pathologists who can perform the same task with an average accuracy of 83.3% [19]. This result is obtained by using a deep neural network that consists of 11 layers, with 4 convolutional layers and 2 fully connected layers. This classifier takes an H&E-stained image as input and will locate and classify the lesions on the given image. Besides its use as an independent lymphoma diagnosis method, deep learning also finds use in processing images collected by devices. For example, a study by Hyungsoon *et al.* utilizes CNN to classify lymphoma on a contrast-enhanced micrography (CEM) device. The device yields holographic images, which are then used as input to the learning process. The system will then detect the existence of lymphoma based on the amount of B-cell on the image and the clonality value. By training and testing the system, it was found that the device can diagnose lymphoma with 91% sensitivity, 100% specificity, and 95% accuracy. It can also decide whether the lymphoma is aggressive or benign with an accuracy of 86% [20]. Aside from the studies above, a few other studies [21] investigate the application of machine learning, especially by deep learning and convolutional neural network, to aid in the lymphoma diagnosis process. Those studies developed classifiers that can perform lymphoma identification and classification with high accuracy and reliability.

In this study, we propose a deep learning-based method to classify H&E-stained nasopharyngeal microscopic biopsy images into one of three classes: lymphoma, carcinoma, and benign lesion.

2 Material and Methods

2.1 Dataset

The dataset used in this study was provided by Padjadjaran University Education Hospital. This dataset contains 430 H&E-stained histopathology microscopic images with the dimension of 1038×1376 pixels and 3 channels from 18 samples. Images in the dataset were taken using two magnification settings which are 100x and 400x. The average number of images for each sample is 23.8 images, with 11.7 standard deviations. There are 235, 99, and 96 images of lymphoma (LY), carcinoma (CA), and benign lesion (BL) samples respectively. Thus, the distribution of images is skewed towards lymphoma. Image samples from each class are shown in Fig. 1.

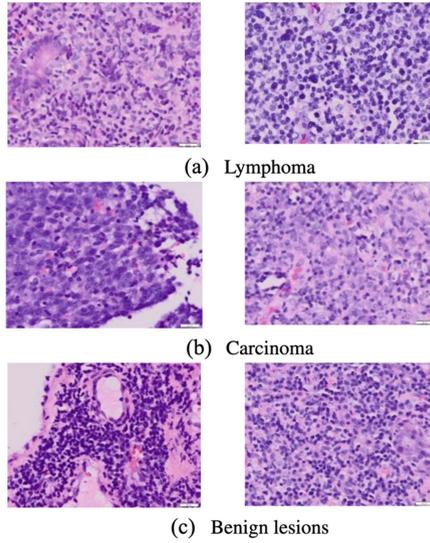


Fig. 1. Dataset visualization of (a) lymphoma, (b) carcinoma, and (c) benign lesion classes.

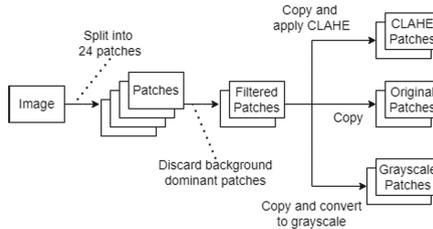


Fig. 2. Dataset preprocessing diagram flowchart, this pipeline is applied to every image in the training and testing set.

2.2 Data Preprocessing

The dataset used in this study consists of two sets: 100x magnification and 400x magnification. The dataset of 100x magnification was split into training and testing sets with a ratio of 4:1. As for the 400x magnification dataset, we reduced the number of images with lymphoma class to balance the distribution of each class. Each image in the training and testing set is preprocessed according to the pipeline illustrated in Fig. 2. Each original image was segmented into 24 patches of $224 \times 224 \times 3$ pixels as shown in Fig. 3. However, we discard the 24th image located in the bottom-right as it contains the magnification annotation. We also discard image patches that have more background pixels than foreground pixels as it contains little to no information. Background pixels are defined as pixels that have saturation values in the range of 0 to 25. Finally, image color preprocessing techniques are applied to the patches. The final composition of the two datasets used in this study is shown in Table 1.

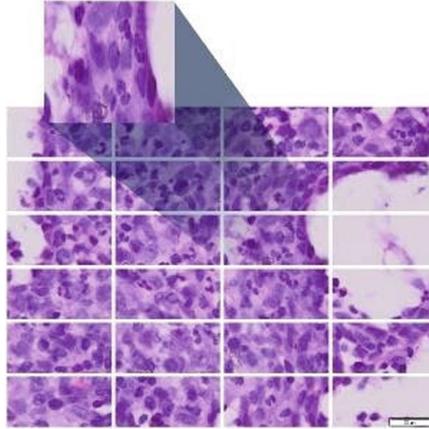


Fig. 3. Image patch segmentation and resizing to $224 \times 224 \times 3$.

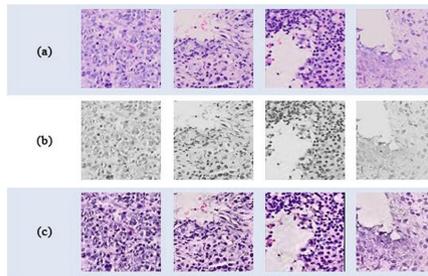


Fig. 4. Color preprocessing scenarios. (a) no preprocessing, (b) grayscale, and (c) image enhancement with CLAHE.

Table 1. Dataset distribution

Magnification	Class	Training	Test
100x	LY	666	890
	CA	653	188
	BL	582	153
400x	LY	1141	2629
	CA	893	245
	BL	1094	276

Image normalization was used to scale image pixel values to float numbers between 0 and 1 inclusively. In addition, the dataset was augmented with a random horizontal flip, random vertical flip, random rotation, and random zoom to increase the model's

generalization ability. In this study, we use three scenarios of image color preprocessing: (1) no color preprocessing, (2) grayscale, and (3) image enhancement with Contrast Limited Adaptive Histogram Equalization (CLAHE). The visualization of each image's color preprocessing result is shown in Fig. 4.

2.3 Model's Architectures

We compare three kinds of feature extractor architecture in this study: (1) NASNetLarge [22], (2) InceptionResNetV2 [23], and (3) Xception [24]. Our choice was motivated by the top-5 accuracy on the ImageNet dataset, where the models above achieved the highest scores and were readily available for use via Keras API. We use transfer learning to install each architecture with pre-trained weights trained on ImageNet.

The NASNetLarge architecture [22] currently tops the ImageNet dataset validation accuracy chart, beating all other readily available architectures on Keras API with a top-5 accuracy of 96.0%. The NASNetLarge architecture was designed by employing reinforcement learning search on the ImageNet dataset to infer the optimal architecture to perform classification on the said dataset. The resulting architecture is a transferable CNN of 88 million parameters which receives a 331×331 pixels image as input.

The InceptionResNetV2 [23] is an Inception and ResNet hybrid architecture, boasting a top-5 accuracy of 95.3% on the ImageNet dataset validation accuracy chart with only 55 million parameters. The combination lies in the residual connection that replaces the filter concatenation stage of the Inception architecture, making it more efficient [25]. The final architecture of this hybrid model contains 572 layers and classifies input images of 224×224 pixels size.

The Xception model is also available on Keras API with a top-5 accuracy of 94.5% on the ImageNet dataset validation accuracy chart. Furthermore, the amount of parameters required to classify the image is only 22 million. The architecture of the Xception model consists of a linear stack of 36 depthwise separable convolution layers. The residual connection helps the Xception model improve its speed and final classification performance [24]. In total, the Xception architecture is built of 126 layers. It receives 224×224 pixels-sized images as classification input.

The classifier used in this study is a global average pooling 2D layer connected to fully connected layers with dropouts, ending in a three-node dense layer with a softmax activation function. The classifier is appended to each CNN model's output to process the feature maps into class predictions. The predicted value is a class probability score ranging from 0 to 100 percent, giving confidence information and class prediction.

2.4 Classification Pipeline

The inputs of this classification system are images of both 100x and 400x magnification from the same specimen. Each image is segmented into 23 patches of $224 \times 224 \times 3$ pixels images. Each patch is fed into the respective CNN model which will return the prediction score for each class. Finally, we take the average prediction score value over all the patches as our prediction result.

2.5 Hyperparameters

Hyperparameters set parameter values before the training process begins, affecting the model's performance later on [26]. Hyperparameters used in this study include the batch size of 16 and epochs of 300. Other hyperparameters used are learning rate, loss function, metric, and optimizer. The learning rate is a step size of each iteration to achieve the global cost minimum of a loss function [27]. The loss function is a measure of any misclassified data from the model [28]. On the other hand, a metric measures the model's performance during training and test [29]. An optimizer is an algorithm used to find a neural network's optimal weights and learning rate to achieve as minimal loss function as possible, which means the model could perform better. The loss function, metric, and optimizer used in the training process were sparse categorical cross-entropy, accuracy, and Adam (with 0.03 learning rate and learning rate scheduler to divide the learning rate by the factor of $e^{0.001}$ for every epoch equal or greater than 25). We used sparse categorical cross-entropy as the loss function and accuracy as the metric, considering the main objective of the algorithm is to classify dataset images into three classes. Adam optimizer was chosen because of its computational efficiency and less tuning for the hyperparameters [30].

The training was done in 300 epochs with three callbacks functions: (1) early stopping to stop training if the validation loss is not improving in 50 epochs, (2) model checkpoint to save the model's weight for every epoch resulting in the best model performance in terms of validation loss, and (3) learning rate scheduler as mentioned before. We train our model while freezing the feature extractor layers to preserve the weight from ImageNet. Each training batch consisted of 32 images. The training was done by using TensorFlow 2.3. While training, the training set was split into a training and validation set with a ratio of 4 to 1.

2.6 Metrics

We divide the metrics into two categories: metrics used in the CNN classification model and the metrics used in the final pipeline. Metrics used in the CNN classification are average accuracy, per class sensitivity, per class specificity, and area under the curve. Per class, sensitivity and specificity were calculated in a one-*vs*-all manner. Metrics used in the final pipeline are accuracy, sensitivity, and specificity.

2.7 Activation Heatmap

To assist the pathologist further, we opt to provide an activation heatmap for each image patch. To do so, we use Gradient-weighted Class Activation Map (Grad-CAM) [31] to process the gradients of the classification result found in the final convolutional layer of the model. The Grad-CAM finally yields a localization map highlighting the important regions in the input image, leading to the classification decision. Thus, by providing an activation heatmap to the pathologist, the pathologist can verify the classification result more easily.

2.8 Experiment Environment

The experiment was performed in the Google Colaboratory platform that uses $1 \times$ Tesla K80 GPU, $1 \times$ single-core Intel® Xeon® CPU @2.30 GHz, 13 GB RAM, and 38 GB disk size.

3 Results and Discussion

The training set was divided into two scenarios: (1) 100x magnification set; and (2) 400x magnification set. For each CNN feature extractor architecture, six models will be trained, each one using a different magnification setting and image color preprocessing scheme. The trained model was then used to classify image patches from the test set. The average classification accuracies of each image patch of each model are shown in Table 2.

From the data in Table 2, it can be seen that all the models performed better on the 400x than the 100x magnification. Among all the models on the 400x magnification, InceptionResNetV2 performed the best overall, reaching 80.3% average patch classification accuracy on the CLAHE-enhanced image patches.

The previous experiment classified each image patch separately, as opposed to classifying a whole image. To improve the classification performance of each image, we proposed an image classification strategy based on the average score of each patch in the image. Our classification pipeline strategy, as mentioned in the previous section, consists of three steps: (1) image patching to 23 patches, (2) classification by either 100x or 400x magnification model, and (3) taking the average score of each class over all patches as the final score. Then we take the class with the highest score as our prediction.

We tested this approach by using the three best models we acquired based on the previous experiment. It is known from the previous experiment that the best way to classify 100x magnification images is by using the Xception (grayscale), NASNetLarge (CLAHE), and InceptionResNetV2 (normal) models. As for 400x magnification images, the three best models are InceptionResNetV2 (CLAHE), NASNetLarge (normal), and InceptionResNetV2 (normal).

Table 2. Experiment results using (a) the normal scenario, (b) grayscale image, and (c) CLAHE-enhanced image, with the best result of each criterion written in boldface

Set	Model	Average accuracy		
		(a)	(b)	(c)
100x	NASNetLarge	40.5%	29.1%	46.7%
	InceptionResNetV2	42%	35.3%	38.1%
	Xception	26.2%	48.5%	37.9%
400x	NASNetLarge	77.1%	50.3%	59.0%
	InceptionResNetV2	73.8%	65.2%	80.3%
	Xception	67.5%	61.1%	60.1%

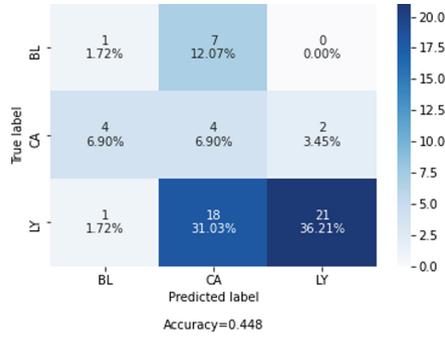


Fig. 5. Confusion matrix of final classification pipeline using Xception on 100x test dataset with grayscale preprocessing.

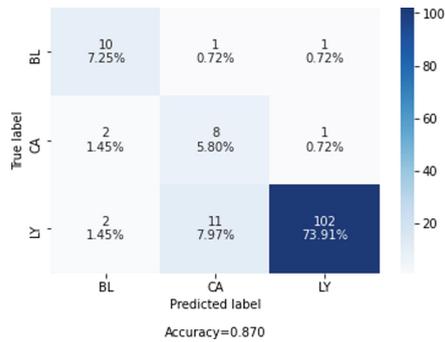


Fig. 6. Confusion matrix of final classification pipeline using InceptionResNetV2 on 400x test dataset with image enhancement using CLAHE.

The results of this experiment are shown in Fig. 5 and Fig. 6. Figure 5 and Fig. 6 show the confusion matrix of the final classification pipeline of Xception (grayscale) and InceptionResNetV2 (CLAHE), the best performing model in 100x and 400x magnification.

The confusion matrix in Fig. 5 shows that the Xception (grayscale) model often misclassified lymphoma as carcinoma. This phenomenon is less apparent in the InceptionResNetV2 (CLAHE)'s confusion matrix shown in Fig. 6, which works with 400x magnification images. However, relatively speaking, the misclassified lymphoma samples are most often classified as carcinoma. This shows that the phenomenon in Xception (grayscale) also exists in InceptionResNetV2 (CLAHE), although on a much lower scale.

As shown in Table 2, the lymphoma classification system using 400x magnification images yields better accuracy, sensitivity, and specificity than the system using 100x magnification images. This result is in line with the previous experiments, where the models work better on 400x than 100x magnification images. Carcinoma is different from lymphoma and benign lesion due to its morphology which the cells can be syncytial and form some island-like structure [32]. On the contrary, lymphoma and benign lesion are more diffuse. Then lymphoma and benign lesions are different in the appearance

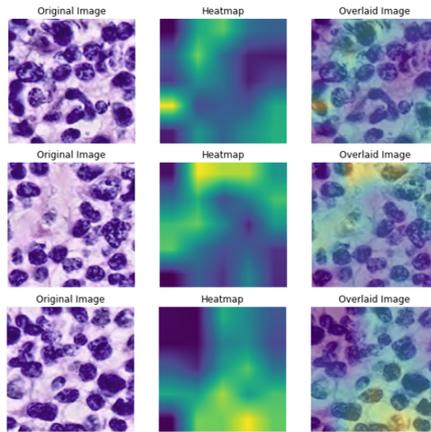


Fig. 7. The activation heatmap of an image patch of a 400 times magnified lymphoma image processed by InceptionResNetV2.

of the nucleolus. Lymphoma cell has lobated nuclei and forms centroblasts which are multiple nuclei with irregular round shape and dispersed chromatin, but benign lesion does not [33–35]. These differences are observed better in higher magnification. Thus, the classification system using 400x magnification images has better performance.

The proposed image classification strategy performed better than classifying each patch separately on 400x magnification. InceptionResNetV2 (CLAHE) saw a 6.7% increase in average accuracy, improving its previous 80.3% average accuracy to 87%. In similar fashion, NASNetLarge (normal) and InceptionResNetV2 (normal) also saw a 2.9% and 11.0% increase in average accuracy. This shows that the proposed strategy is able to counter outlier patches and make the whole image classification more robust and accurate, improving the overall performance.

Interestingly, the proposed strategy shows to be lowering the accuracy on 100x magnification. This may be caused by the low, under 50%, average accuracy of the models when working on 100x magnification images. This means that the misclassified patches outnumber the correctly classified ones. Thus, when the proposed strategy is used, the correctly classified patches are considered outliers instead. This, in turn, hampered the models' performance.

Figure 7 shows the activation heatmap of three lymphoma image patches of 400x magnification, each enhanced by CLAHE and predicted by InceptionResNetV2. The brighter a pixel is, the more important that pixel is to the classifier decision. Conversely, the darker a pixel is, the less important that pixel is to the classifier decision. Although the significance of the highlighted regions is not immediately clear to an untrained human's eye, it might help a pathologist verify the classifier decision because this classifier model has been used in microscopic image analysis [36].

The best nasopharyngeal microscopic biopsy images classification system in this research is the one that uses 400x magnification images and the InceptionResNetV2 model with image enhancement using CLAHE. InceptionResNetV2 with CLAHE yields 87% accuracy. Its performance in differentiating lymphoma images over carcinoma and

benign lesions images are outperforming other methods with 88.7% sensitivity and 91.3% specificity. On the other hand, the classification performance on 100x magnification images is inadequate. The proposed image classification strategy can increase the models' average image classification accuracy up to 11.0%, as long as it has a good enough per patch accuracy. Further research with a larger and better-annotated dataset is essential to verify and improve this method. Aside from using popular deep learning architectures, research on customized architecture is also recommended.

Acknowledgments. This research was supported by the Department of Pathological Anatomy Padjajaran University at Hasan Sadikin Hospital, and the Department of Biomedical Engineering at Bandung Institute of Technology. We are grateful for the lymphoma dataset provided.

Authors' Contributions. HAU and OH collected lymphoma datasets from patients in Hasan Sadikin Hospital. A developed experimental design. A, ITK, RNA, HRY, AEN, and GFM conducted the experiment and interpret the experimental result. A, ITK, RNA, HRY, AEN, and GFM wrote the manuscript. IA and WA contributed to the revision of the manuscript. All authors read and approved the final manuscript.

References

1. M. R. Hongorjo and S. Syed, "Presentation, staging and diagnosis of lymphoma: a clinical persepective," *Journal of Ayub Medical College: JAMC*, vol. XX, pp. 100-3, 2008.
2. H. Kaseb and H. M. Babiker, *Hodgkin Lymphoma*, Treasure Island (FL): StatPearls Publishing, 2020.
3. S. Sapkota and H. Shaikh, *Non-Hodgkin Lymphoma*, Treasure Island (FL): StatPearls Publishing, 2020.
4. T. G. C. Observatory, "Indonesia," *World Health Organization: International Agency for Research on Cancer*, 2018.
5. I. P. D. P. Suryadiarsa, N. M. R. A. Rena and T. G. Dharmayuda, "Tingkat harapan hidup pasien limfoma non-hodgkin berdasarkan skor IPI yang mendapatkan kemoterapi lini pertama di RSUP Sanglah Denpasar tahun 2014," *E-Jurnal Medika*, vol. VIII, no. 5, 2019.
6. K. N. P. Kanker, "Panduan nasional penanganan kanker limfoma non-hodgkin," *Kementerian Kesehatan Republik Indonesia*, 2015.
7. D. R. Wahyuningrum, "Kesesuaian diagnosis pemeriksaan histopatologi pewarnaan hematok-silin eosin pada kasus limfoma maligna dengan standar baku pemeriksaan imunohidtokimia di instalasi patologi anatomi RSSA Malang," *Universitas Brawijaya: Sarjana Thesis*, 2016.
8. C. Sryrykh, A. Abreu, N. Amara, A. Siegfried, V. Masongrosse, F. X. Frenois, L. Martin, C. Rossi, C. Laurent and P. Brousset, "Accurate diagnosis of lymphoma on whole-side histopathology images using deep learning," *npj Digital Medicine*, vol. III, no. 63, 2020. doi: <https://doi.org/10.1038/s41746-020-0272-0>
9. Y. W. Caigan Du, "The immunoregulatory mechanisms of carcinoma for its survival and development," *Journal of Experimental & Clinical Cancer Research*, 2011. <https://doi.org/10.1186/1756-9966-30-12>
10. M. E. M. T. S. K. D. M. R. X. F. M. V. Claudia Gonzales, "Epithelial Cell Culture from Human Adenoids: A Functional Study Model for Ciliated and Secretory Cells," *BioMed Research International*, 2013. Doi: <https://doi.org/10.1155/2013/478713>

11. S. A. Hoda and A. Patelno, "Rosai and Ackerman's Surgical Pathology," *American Journal of Clinical Pathology*, vol. 149, no. 6, p. 548, April 2018.
12. N. Watanabe, H. Inohara, S. Akahani, Y. Yamamoto, K. Moriwaki and T. Kubo, "Synchronous squamous cell carcinoma and malignant lymphoma in the head and neck region," *Auris Nasus Larynx*, vol. 34, no. 2, pp. 273–276, 2007. <https://doi.org/https://doi.org/10.1016/j.anl.2006.07.002>
13. L. Yann, Y. Bengio and G. Hinton, "Deep learning," *Nature*, no. 521, pp. 436–444, 2015. doi: <https://doi.org/10.1038/nature14539>
14. F. Sultana, A. Sufian and G. Hinton, "Advancements in image classification using convolutional neural network," *Proc. - 2018 4th IEEE Int. Conf. Res. Compt. Intell. Commun. Networks, ICRCICN 2018*, pp. 122–129, 2018.
15. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. DXXIV, no. 7639, pp. 115–118, 2017. <https://doi.org/https://doi.org/10.1038/nature21056>
16. V. Gulshan, L. Peng and M. Coram, "Development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Journal of the American Medical Association*, vol. CCCXVI, no. 22, pp. 2402–2410, 2016. doi: <https://doi.org/10.1001/jama.2016.17216>
17. R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. CXXIV, no. 22, pp. 962–969, 2017. doi: <https://doi.org/10.1016/j.ophtha.2017.02.008>
18. D. S. W. Ting, C. Y. Cheung, G. Lim and G. Tan, "Development and validation of deep learning system for diabetic retinopathy ad related eye diseases using retinal images from multiethnic population," *Jornal of the American Medical Association*, vol. CCCVIII, no. 7, pp. 962–969, 2017. doi: <https://doi.org/10.1001/jama.2017.18152>
19. H. Miyoshi, K. Sato, Y. Kabeya, S. Yonezawa, H. Nakano, Y. Takeuchi, I. Ozawa, S. Higo, E. Yanagida, K. Yamada, K. Kohno, T. Furuta, H. Muta, M. Takeuchi, Y. Sasaki, T. Yoshimura, K. Matsuda, R. Muto, M. Moritsubo, K. Inoue, T. Suzuki, H. Sekinaga and K. Ohshima, "Deep learning shows the capability of high-level computer-aided diagnosis in malignant lymphoma," *Laboratory Investigation*, vol. 100, pp. 1300–1310, 2020. <https://doi.org/10.1038/s41374-020-0442-3>
20. H. Im , D. Pathania , P. J. McFarland, A. R. Sohani, I. Degani, M. Allen, B. Coble, A. Kilcoyne, S. Hong, L. Rohrer, J. S. Abramson, S. Dryden-Peterson, L. Fexon, M. Pivovarov and B. Chabner, "Design and clinical validation of a point-of-care device for the diagnosis of lymphoma via contrast-enhanced microholography and machine learning," *Nature Biomedical Engineering*, pp. 1–9, 2018. <https://doi.org/10.1038/s41551-018-0265-3>
21. J. S. Mohlman, S. D. Leventhal, T. Hansen, J. Kohan, V. Pascucci and M. E. Salama, "Improving Augmented Human Intelligence to Distinguish Burkitt Lymphoma From Diffuse Large B-Cell Lymphoma Cases," *American Journal of Clinical Pathology*, vol. 153, no. 6, pp. 743–759, 2020. doi: <https://doi.org/10.1093/ajcp/aqaa001>
22. V. V. J. Barret Zoph, "Learning Transferable Architectures for Scalable Image Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision*, 2018.
23. S. I. V. V. A. A. A. Chirstian Szegedy, "Inception-v4, inception-ResNet and the impact of residual connections on learning," 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 2017. doi: <https://doi.org/10.5555/3298023.3298188>
24. F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2014.
25. R. M. Kamble, G. C. Y. Chan, O. Perdomo, F. A. M. H. Gonzalez and F. Meriaudeau, "Automated Diabetic Macular Edema (DME) Analysis using Fine Tuning with Inception-Resnet-v2 on OCT Images," *IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 442–446, 2018. doi: <https://doi.org/10.1109/IECBES.2018.8626616>

26. V. Jatana, “Hyperparameter Tuning,” August 2019.
27. S. Raschka, Python Machine Learning, Packt Publishing, 2015.
28. C. M. Bishop, “Pattern recognition and machine learning,” 2006.
29. M. Hossin and S. M.N, “A Review on Evaluation Metrics for Data Classification Evaluations,” International Journal of Data Mining & Knowledge Management Process, vol. 5, p. 1–11, March 2015. doi: <https://doi.org/10.5121/ijdkp.2015.5201>
30. D. Kingma and D. Kingma, “Adam: A Method for Stochastic Optimization,” arXiv, 2014. <https://doi.org/10.48550/arXiv.1412.6980>
31. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” International Journal of Computer Vision, vol. 128, no. 2, pp. 336–359, 2020. doi: <https://doi.org/10.1109/ICCV.2017.74>
32. M.-Y. a. Y. L. a. J. X.-D. a. Z. Z. Ji, N. Zhan, P.-X. Huang, C. Lu and W.-G. Dong, “Nuclear shape, architecture and orientation features from H&E images are able to predict recurrence in node-negative gastric adenocarcinoma,” Journal of translational medicine, vol. 17, no. 1, p. 92, 2019. <https://doi.org/10.1186/s12967-019-1839-x>
33. T. A. A. Tosta, L. A. Neves and M. Z. do Nascimento, “Segmentation methods of H&E-stained histological images of lymphoma: a review,” Informatics in Medicine Unlocked, vol. 9, pp. 35–43, 2017. <https://doi.org/10.1016/j.imu.2017.05.009>
34. Loew, J. M., & Macon, W. R. (2010). Lymph nodes. In Differential Diagnosis in Surgical Pathology (pp. 745–789). Elsevier Inc. <https://doi.org/10.1016/B978-1-4160-4580-9.00014-9>
35. S. C. van der Putte, H. J. Schuurman, L. H. Rademakers, P. Kluin and J. A. van Unnik, “Malignant lymphoma of follicle centre cells with marked nuclear lobation,” Virchows Archiv. B, Cell pathology including molecular pathology, vol. 46, pp. 93–107, 1984. doi: <https://doi.org/10.1007/BF02890299>
36. L. D. Nguyen, D. Lin, Z. Lin and J. Cao, “Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation,” in 2018 IEEE International Symposium on Circuits and Systems (ISCAS), 2018. doi: <https://doi.org/10.1109/ISCAS.2018.8351550>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

