



# HOKKAIDO UNIVERSITY

|                  |                                                                                                                                                                                                          |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title            | Artificial intelligence challenge of discriminating cutaneous arteritis and polyarteritis nodosa based on hematoxylin-and-eosin images of skin biopsy specimens                                          |
| Author(s)        | Kashiwa, Wataru; Hirata, Kenji; Endo, Hiroki et al.                                                                                                                                                      |
| Citation         | Pathology – Research and Practice, 269, 155915<br><a href="https://doi.org/10.1016/j.prp.2025.155915">https://doi.org/10.1016/j.prp.2025.155915</a>                                                      |
| Issue Date       | 2025-03-19                                                                                                                                                                                               |
| Doc URL          | <a href="https://hdl.handle.net/2115/97467">https://hdl.handle.net/2115/97467</a>                                                                                                                        |
| Rights           | © 2025. This manuscript version is made available under the CC-BY-NC-ND 4.0 license<br><a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a> |
| Rights(URL)      | <a href="https://creativecommons.org/licenses/by-nc-nd/4.0/">https://creativecommons.org/licenses/by-nc-nd/4.0/</a>                                                                                      |
| Type             | journal article                                                                                                                                                                                          |
| File Information | Manuscript_Kashiwa et al_accepted version.pdf                                                                                                                                                            |



# **Artificial intelligence challenge of discriminating cutaneous arteritis and polyarteritis nodosa based on hematoxylin-and-eosin images of skin biopsy specimens**

Wataru Kashiwa<sup>1</sup>, Kenji Hirata<sup>1</sup>, Hiroki Endo<sup>1</sup>, Kohsuke Kudo<sup>1</sup>, Chietsugu Katoh<sup>2</sup>, Tamihiro Kawakami<sup>3,4</sup>, Hiroyuki Kanno<sup>5</sup>, Kei Takahashi<sup>6</sup>, Tatsuhiko Miyazaki<sup>7</sup>, Eiji Ikeda<sup>8</sup>, Toshiaki Oharaseki<sup>6</sup>, Yayoi Ogawa<sup>9</sup>, Mitsuho Onimaru<sup>10</sup>, Mie Kurata<sup>11</sup>, Daigo Nakazawa<sup>12</sup>, Eri Muso<sup>13</sup>, Yuka Nishibata<sup>14</sup>, Sakiko Masuda<sup>14</sup>, Utano Tomaru<sup>15</sup>, Yoshihiro Matsuno<sup>15</sup>, Shunsuke Furuta<sup>16</sup>, Yoshiyuki Abe<sup>17</sup>, Naoto Tamura<sup>17</sup>, Masayoshi Harigai<sup>18</sup>, Akihiro Ishizu<sup>14</sup>

<sup>1</sup>Department of Diagnostic Imaging, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo, Japan

<sup>2</sup>Department of Biomedical Science and Engineering, Faculty of Health Sciences, Hokkaido University, Sapporo, Japan

<sup>3</sup>Division of Dermatology, Tohoku Medical and Pharmaceutical University, Sendai, Japan

<sup>4</sup>Kawakami Dermatology Clinic, Chiba, Japan

<sup>5</sup>Department of Pathology, Shinshu University School of Medicine, Matsumoto, Japan

<sup>6</sup>Department of Pathology, Toho University Ohashi Medical Center, Tokyo, Japan

<sup>7</sup>Department of Pathology, Gifu University Hospital, Gifu, Japan

<sup>8</sup>Department of Pathology, Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan

<sup>9</sup>Hokkaido Renal Pathology Center, Sapporo, Japan

<sup>10</sup>Division of Pathophysiological and Experimental Pathology, Department of Pathology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

<sup>11</sup>Department of Analytical Pathology, Ehime University Graduate School of Medicine, Ehime, Japan

<sup>12</sup>Department of Rheumatology, Endocrinology, and Nephrology, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo, Japan

<sup>13</sup>Department of Nephrology and Dialysis, Medical Research Institute Kitano Hospital, PIIF Tazuke Kofukai, Osaka, Japan

<sup>14</sup>Department of Medical Laboratory Science, Faculty of Health Sciences, Hokkaido University, Sapporo, Japan

<sup>15</sup>Department of Surgical Pathology, Hokkaido University Hospital, Sapporo, Japan

<sup>16</sup>Department of Allergy and Clinical Immunology, Chiba University Hospital, Chiba, Japan

<sup>17</sup>Department of Internal Medicine and Rheumatology, Juntendo University School of Medicine, Tokyo, Japan

<sup>18</sup>Division of Rheumatology, Department of Internal Medicine, Tokyo Women's Medical University School of Medicine, Tokyo, Japan

Correspondence: Akihiro Ishizu, MD, PhD, Department of Medical Laboratory Science, Faculty of Health Sciences, Hokkaido University, Kita-12, Nishi-5, Kita-ku, Sapporo 0600812, Japan. E-mail: aishizu@med.hokudai.ac.jp

## Abstract

Diseases that develop necrotizing vasculitis of cutaneous muscular arteries include cutaneous arteritis (CA) and polyarteritis nodosa (PAN). It is difficult to distinguish them based on skin biopsy findings alone. This study demonstrated that artificial intelligence (AI) can discriminate them based on skin biopsy findings and revealed where AI focuses on the image. Ninety-three hematoxylin-and-eosin images of CA and 19 PAN images were used. Among them, 85 CA and 17 PAN images were used to train AI; thereafter, AI was challenged to classify the remaining images. The same test images were evaluated by 26 pathologists with different years of experience. AI accuracy was 75.2%, whereas that of pathologists was 42.8%. Gradient-weighted class activation mapping (Grad-CAM) indicated that AI focused on connective tissues around the affected vessels rather than the affected vessels. Twenty-two of the 26 pathologists were randomly divided into two groups of 11 each, one of which referred to Grad-CAM images and was challenged in the second-round test of images different from the first round. The accuracy significantly improved after referring to Grad-CAM images, whereas it was equivalent to the first round without referring to Grad-CAM images. In the survey after the second-round test, pathologists who referred to Grad-CAM images suggested that inflammation and fibrosis in the surrounding connective tissues in PAN might be abundant compared to CA. AI may be useful for histological differentiation between CA and PAN and can help pathologists improve the ability of discriminating CA and PAN based on histological findings of skin biopsy specimens.

**Keywords:** Cutaneous arteritis; Polyarteritis nodosa; Skin biopsy specimens; Artificial intelligence; Grad-CAM

### **List of Abbreviations**

AI: artificial intelligence

AUC: area under the curve

CA: cutaneous arteritis

CNN: convolution neural network

Grad-CAM: gradient-weighted class activation mapping

H&E: hematoxylin-and-eosin

PAN: polyarteritis nodosa

Resnet50: Residual Network 50

ROC: receiver operating characteristics

TN: true negative

TP: true positive

## 1. Introduction

Vasculitis is a disease in which the vascular wall is destroyed by primary inflammation<sup>1</sup>. Based on histopathological features, vasculitis can be broadly classified into necrotizing vasculitis, in which fibrinoid necrosis occurs in the vessel wall, and granulomatous vasculitis, in which granulomatous inflammation occurs in and around the vessel wall<sup>2</sup>. There are two diseases that develop necrotizing vasculitis of cutaneous muscular arteries: polyarteritis nodosa (PAN), which also affects systemic muscular arteries, and cutaneous arteritis (CA), which exclusively affects cutaneous muscular arteries.

The average age of PAN onset is 53 years, with a male-to-female ratio of 1:1<sup>3</sup>. The cause of the disease is unknown in most cases, although there are reports that the disease is triggered by hepatitis viruses or other viral infections and vaccinations<sup>4,5</sup>. Some studies have suggested that molecular mimicry between the viral antigens and autoantigens may be involved in the pathogenesis<sup>6-8</sup>. In severe cases, glucocorticoid pulse therapy may be required initially, followed by oral glucocorticoid administration<sup>9</sup>. Intravenous cyclophosphamide may be administered with glucocorticoids and is generally repeated four to six times. Oral glucocorticoids are tapered off as the disease improves, but small doses are often continued to prevent relapse. In mild cases, oral glucocorticoids alone may be the optional treatment. It is important to diagnose PAN early and initiate treatment when vascular lesions are nascent. Early treatment sometimes leads to complete remission<sup>10</sup>. In contrast, delayed initiation of treatment increases the frequency of cerebral hemorrhage, gastrointestinal bleeding and perforation, and death from renal failure<sup>11</sup>.

CA is defined as “vasculitis confined to the skin, with histopathological findings on skin biopsy similar to those of PAN but without organ involvement”<sup>12</sup> and is more common in women after the age of 40, with a peak in the 50s<sup>13</sup>. The cause of the

disease is unknown, its treatment has not been fully investigated, and there is no established therapy. Nonsteroidal anti-inflammatory drugs and drugs that improve peripheral circulation are prescribed for mild cases, and diaphenylsulfone and colchicine are considered for treatment-resistant cases<sup>14</sup>.

Most patients with CA have a good prognosis despite a long course with recurrent cutaneous symptoms, but there are a few cases that transform to PAN after a long period of the disease<sup>13</sup>. Therefore, when necrotizing vasculitis is discovered in skin biopsy specimens, it is important to consider whether it is a lesion of CA, a dermal lesion involved in PAN, or a lesion that will transform to PAN. However, there are no histopathological criteria to distinguish PAN from CA, and they are currently differentiated each other clinically by systemic examinations and careful follow-up.

With the rapid development of artificial intelligence (AI) in recent years, there are high expectations for the early detection of diseases and suggestion of optimal treatment by AI in the medical field, and research and development of medical AI is actively pursued<sup>15</sup>. AI is particularly strong when analyzing medical images with huge data, such as computed tomography, magnetic resonance imaging, ultrasound, endoscopy, and pathology images. Deep learning is a core technology for AI-based image diagnosis, which deepens the hierarchy of artificial neural networks<sup>16</sup>. The principle began to attract attention when it dramatically improved the accuracy of image recognition up to that point; in 2015, it exceeded the general human image recognition rate (95%)<sup>17</sup>.

This study aimed to demonstrate that AI can discriminate CA and PAN based on skin biopsy findings alone and to reveal where AI focuses on the image for discrimination. In addition, this study asked whether pathologists supported by AI could improve the ability of discriminating CA and PAN based on histological findings of skin biopsy specimens.

## 2. Materials and Methods

### 2.1.Images

Hematoxylin-and-eosin (H&E)-stained skin biopsy tissue images ( $\times 40$ ;  $1360 \times 1024$  pixels) labeled with clinical diagnosis of CA or PAN were collected; these included 93 CA and 19 PAN images.

### 2.2.Dataset

Eight CA and two PAN images were randomly selected as test data, and the remaining 85 CA and 17 PAN images were further split in a random manner into training data (80%, 77 CA and 15 PAN images) and validation data (20%, 8 CA and 2 PAN images).

### 2.3.Image preprocessing

After masking background with black manually, all pixels were divided by the maximum value of all images and normalized to (0,1). Because it was difficult for the computer to operate  $1360 \times 1024$  pixel images due to memory capacity, each image was divided into  $340 \times 256$  pixel images. Brightness adjustment and rotation were performed to amplify CA images to 5440 and PAN images to 5100 finally in order to minimize the imbalanced distribution between CA and PAN.

### 2.4.Architecture

Residual Network 50 (Resnet50), a convolution neural network (CNN), was used to discriminate CA and PAN. CNN is a computational system that simulates brain neurons and consists of an input layer, a hidden layer, and an output layer, each consisting of multiple nodes connected by edges. Resnet50 is a CNN architecture built using residual learning<sup>18</sup>. This model comprises 50 convolutional and all-joining layers and is mainly

used for image recognition tasks. It can train very deep networks by reducing the gradient loss problem, in which the gradient becomes zero as the layers get deeper and learning does not proceed. Resnet50 also performs well on large datasets, such as ImageNet, which is a large supervised labeled image database of >14 million color photographs<sup>19</sup>. In this study, Resnet50, which has been pretrained with ImageNet, was used to prevent overtraining.

### *2.5.Learning*

Learning was performed using training data, with images as input and diagnosis as output. Learning was done in 100 epochs of mini-batch training with a batch size of 16, and Adam was used as the Optimizer. Cross-entropy was used as the loss function. During learning, the pretrained Resnet50 was fine-tuned to adapt to this task. Fine tuning is the process of adjusting machine learning parameters for a new task by retraining the same data or using a model pretrained with a different dataset. It is considered effective when the dataset is small.

### *2.6.Evaluation of AI*

Accuracy was evaluated by inputting test data to the trained AI. For evaluation, CA and PAN were defined as positive and negative, respectively. When CA is correctly discriminated as CA, the classification is true positive (TP); when CA is incorrectly discriminated as PAN, the classification is false negative; when PAN is incorrectly discriminated as CA, the classification is false positive; and when PAN is correctly discriminated as PAN, the classification is true negative (TN). From the mixture matrix of discrimination results, accuracy, sensitivity, and specificity were calculated. A receiver operating characteristics (ROC) curve was generated from discrimination probabilities of the images, and area under the curve (AUC) was calculated.

### *2.7.Gradient-weighted class activation mapping (Grad-CAM)*

Grad-CAM is a visualization method that extends the class activation map, in which weights connecting the global average pooling calculation values and the neural network output layer are applied to a feature map to visualize the regions in the image that provide the basis for decisions<sup>20</sup>. Grad-CAM images were created by overlying the input H&E and heatmap images using training data.

### *2.8.Hardware and software environment*

This study was conducted in the following environments: OS, Windows 11 64-bit; CPU, AMD Ryzen5 5600; memory, 48GB; GPU, NVIDIA GeForce RTX 3060 16GB; framework, Pytorch 2.0.1; language, Python 3.7.7.

### *2.9.Discrimination of CA and PAN by pathologists*

To compare AI accuracy to that of the pathologists, 15 images ( $340 \times 256$  pixels) from test data used for AI evaluation were randomly selected to contain CA and PAN. Regarding these images, 26 pathologists with different years of experience were challenged to discriminate CA and PAN. Twenty-two pathologists, excluding the 4 who withdrew, were randomly divided into two groups of 11 each, one of which referred to Grad-CAM images. Pathologists with or without referring to Grad-CAM images were challenged again to discriminate another 15-image set, including CA and PAN.

### *2.10. Statistics*

Accuracy was compared between the first- and second-round challenges using a paired Student's *t*-test and then confirmed by Wilcoxon signed rank sum test.  $P < 0.05$  was considered significant.

### 3. Results

#### 3.1. Deep learning

During deep learning up to 100 epochs using training data, AI accuracy of discriminating CA and PAN improved up to 99.2% (**Figure 1a**, blue line). Regarding validation data, AI discriminated CA and PAN with a maximum accuracy of 83.5% (**Figure 1a**, orange line).

#### 3.2. Evaluation of AI

The performance of AI after deep learning was evaluated using test data. A ROC curve was drawn by defining TP and TN as correctly diagnosed CA and PAN, respectively. AUC calculated from the ROC curve was 0.800, and accuracy, sensitivity, and specificity were 75.2%, 80.8%, and 53.1%, respectively (**Figure 1b**).

#### 3.3. Grad-CAM

Grad-CAM images were created by overlying the input H&E and heatmap images using training data (**Figure 2**). Grad-CAM images indicated that AI focused on connective tissues around the affected vessels rather than the affected vessels themselves when diagnosing both CA and PAN.

#### 3.4. Comparison of AI and pathologists

Twenty-six pathologists with different years of experience were challenged to discriminate CA and PAN regarding 15 images ( $340 \times 256$  pixels) randomly selected from test data containing CA and PAN images. The accuracy of pathologists was 42.8% on average, regardless of years of experience (**Figure 3**).

### *3.5.Improved accuracy of discriminating CA and PAN after referring to Grad-CAM images*

Twenty-two pathologists were randomly divided into two groups of 11 each, and one group referred to Grad-CAM images and was challenged in the second-round test, which included CA and PAN images different from those in the first-round test. The other group was challenged the same second-round test without referring to Grad-CAM images. In the former group, the accuracy significantly improved after referring to Grad-CAM images (from 43.0% to 55.2%,  $p=0.0011$ ), whereas it was equivalent to the first-round test in the latter group without referring to Grad-CAM images (from 42.4% to 44.8%,  $p=0.5625$ ; **Figure 4**).

#### 4. Discussion

In this study, AI was trained using skin biopsy tissue images labeled with clinically confirmed diagnosis of CA and PAN and discriminated these two diseases from skin biopsy tissue images alone with 75.2% accuracy. Because the diagnostic accuracy of pathologists was 42.8%, AI could be effective in histopathological differentiation between CA and PAN.

When CA was considered the correct answer, the AUC calculated from the ROC curve was 0.800, and the sensitivity and specificity of AI discrimination were 80.8% and 53.1%, respectively. Although CA skin biopsy tissue images can be distinguished from PAN with relatively high probability, improving the ability to accurately diagnose PAN is a future challenge. The reason for the poor discrimination of PAN (53.1%) compared to CA (80.8%) in this study is the smaller number of PAN images (n=19) than CA images (n=93). Skin biopsy tissue images from patients with PAN are difficult to collect because it is a rare disease. Further improvement of AI performance is expected by collecting more PAN skin biopsy tissue images.

In this study, Grad-CAM analysis was performed to visualize the areas of interest when AI discriminated CA and PAN. Grad-CAM images indicated that AI focused on connective tissues around the affected vessels as the basis for discriminating CA and PAN rather than the affected vessels themselves. Collective findings suggest a certain histopathological difference between CA and PAN. As proof, pathologists who referred to Grad-CAM images showed a significant improvement in their ability to discriminate CA and PAN, whereas those without referring to Grad-CAM images did not. It is possible that pathologists who referred to Grad-CAM images may have gained some insights into the condition of perivascular tissues that may provide a basis for discriminating CA and PAN. Therefore, we sent them a questionnaire asking the

difference in connective tissues around the affected vessels between the two diseases. The most agreed viewpoint was the degree of inflammatory cell infiltration and fibrosis. They pointed out that “AI seems to focus on fibrosing connective tissues with inflammatory cell infiltration around the affected vessels when diagnosing PAN, whereas it focuses on the surrounding adipose tissues without inflammation and fibrosis when diagnosing CA.”

Although further studies are needed to verbalize the difference, these findings may reflect the difference in the quantity and/or quality of inflammation between PAN and CA, consistent with a previous study demonstrating that the serum basic fibroblast growth factor and vascular endothelial growth factor levels in patients with PAN are higher than those in patients with CA<sup>21</sup>.

There is another limitation in this study: an uncertainty of clinical diagnosis of CA. CA rarely transforms to PAN after a long clinical course of the disease. Therefore, it cannot deny that CA images used in this study include those with future possibility of transition to PAN, although CA was diagnosed after a considerable period of clinical observation. A long-term prospective study will solve this problem.

## **5. Conclusion**

This study demonstrated that trained AI could discriminate CA and PAN based on skin biopsy histological findings with 75.2% accuracy, which is superior than pathologists' accuracy (42.8%). Grad-CAM suggests that the histopathological difference between CA and PAN may be present in connective tissues surrounding the affected vessels rather than the affected vessels themselves. The difference should be identified precisely in future studies. Although the AI accuracy of 75.2% is still insufficient for clinical practice, AI can help pathologists improve the ability of discriminating CA and PAN based on histological findings of skin biopsy specimens by pointing out the clue of differentiation between the two diseases. Pathologists and AI can work together to improve diagnostic accuracy.

## **Acknowledgments**

We thank all pathologists who participated in this study. This work was supported by the Japan Research Committee of the Ministry of Health, Labour and Welfare for Intractable Vasculitis (20FC1044 and 23FC1019).

## **Author Contributions**

T.K., H.K., K.T., T.M., E.I., T.O., Y.O., M.O., M.K., D.N., E.M., and A.I. designed the study. T.K., S.F., Y.A., N.T., and M.H. collected the skin biopsy tissue CA and PAN images. W.K., K.H., Y.K., C.K. and A.I. performed experiments. W.K., K.H., Y.K., C.K., T.K., H.K., K.T., T.M., E.I., T.O., Y.O., M.O., M.K., D.N., E.M., Y.N., S.M., U.T., Y.M., and A.I. analyzed data. W.K., Y.N., S.M., and A.I. wrote the manuscript. All authors read the final version of the manuscript and provided critical feedback.

## **Ethics approval**

This study was approved by the Ethics Committee of the Faculty of Health Sciences of Hokkaido University (approval no. 20-31).

## **Data availability statement**

The data underlying this article will be shared at reasonable request to the corresponding author.

## **Declaration of competing interest**

The authors report no declarations of interest.

## References

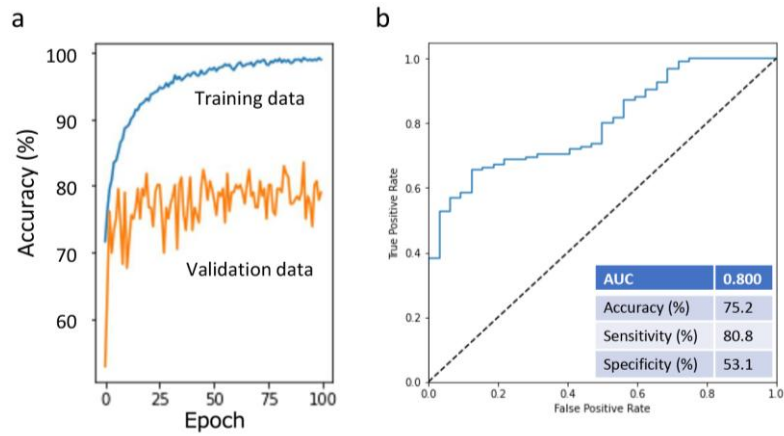
- 1 J. C. Jennette, R. J. Falk, P. A. Bacon, N. Basu, M. C. Cid, F. Ferrario, L. F. Flores-Suarez, W. L. Gross, L. Guillevin, E. C. Hagen, G. S. Hoffman, D. R. Jayne, C. G. Kallenberg, P. Lamprecht, C. A. Langford, R. A. Luqmani, A. D. Mahr, E. L. Matteson, P. A. Merkel, S. Ozen, C. D. Pusey, N. Rasmussen, A. J. Rees, D. G. Scott, U. Specks, J. H. Stone, K. Takahashi, and R. A. Watts, 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides, *Arthritis Rheum.* 65 (1) (2013) 1-11.
- 2 A. Ishizu, T. Kawakami, H. Kanno, K. Takahashi, T. Miyazaki, E. Ikeda, T. Oharaseki, Y. Ogawa, M. Onimaru, M. Kurata, D. Nakazawa, E. Muso, and M. Harigai, Expert perspectives on pathological findings in vasculitis, *Mod. Rheumatol.* 33 (1) (2023) 1-11.
- 3 M. Kawazoe, T. Nanki, N. Hagino, N. Ikegaya, S. Ito, M. Koderu, N. Nakano, M. Suzuki, S. Kaname, and M. Harigai, Clinical characteristics of patients with polyarteritis nodosa based on a nationwide database in Japan, *Mod. Rheumatol.* 32 (3) (2022) 598-605.
- 4 L. Wolff, A. Horisberger, L. Moi, M. P. Karampetsou, and D. Comte, Polyarteritis Nodosa: Old Disease, New Etiologies, *Int. J. Mol. Sci.* 24 (23) (2023) 16668.
- 5 A. Makiyama, Y. Abe, H. Furusawa, M. Kogami, T. Ando, K. Tada, M. Onimaru, A. Ishizu, K. Yamaji, and N. Tamura, Polyarteritis nodosa diagnosed in a young male after COVID-19 vaccine: A case report, *Mod. Rheumatol. Case Rep.* 8 (1) (2023) 125-132.
- 6 Y. Segal and Y. Shoenfeld, Vaccine-induced autoimmunity: the role of molecular mimicry and immune crossreaction, *Cell Mol. Immunol.* 15 (6) (2018) 586-594.
- 7 Y. Chen, Z. Xu, P. Wang, X. M. Li, Z. W. Shuai, D. Q. Ye, and H. F. Pan, New-

- onset autoimmune phenomena post-COVID-19 vaccination, *Immunology* 165 (4) (2022) 386-401.
8. V. Sanker, M. Mylavarapu, P. Gupta, N. Syed, M. Shah, and V. V. K. Dondapati, Post COVID-19 vaccination medium vessel vasculitis: a systematic review of case reports, *Infection* 52 (4) (2024) 1207-1213.
  9. M. Isobe, K. Amano, Y. Arimura, A. Ishizu, S. Ito, S. Kaname, S. Kobayashi, Y. Komagata, I. Komuro, K. Komori, K. Takahashi, K. Tanemoto, H. Hasegawa, M. Harigai, S. Fujimoto, T. Miyazaki, T. Miyata, H. Yamada, A. Yoshida, T. Wada, Y. Inoue, H. A. Uchida, H. Ota, T. Okazaki, M. Onimaru, T. Kawakami, R. Kinouchi, A. Kurata, H. Kosuge, K. E. Sada, K. Shigematsu, E. Suematsu, E. Sueyoshi, T. Sugihara, H. Sugiyama, M. Takeno, N. Tamura, M. Tsutsumino, H. Dobashi, Y. Nakaoka, K. Nagasaka, Y. Maejima, H. Yoshifuji, Y. Watanabe, S. Ozaki, T. Kimura, H. Shigematsu, K. Yamauchi-Takihara, T. Murohara, S. I. Momomura, and J. C. S. Joint Working Group, JCS 2017 Guideline on Management of Vasculitis Syndrome - Digest Version, *Circ J* 84 (2) (2020) 299-359.
  10. S. M. Sanchez-Cubias, E. Martin-Nares, G. Hernandez-Molina, and A. Hinojosa-Azaola, Clinical Characteristics and Outcomes of Mexican Patients With Polyarteritis Nodosa: A Single-Center Study Over 40 Years, *J. Clin. Rheumatol.* 27 (6S) (2021) S259-S264.
  11. S. Jardel, X. Puechal, A. Le Quellec, C. Pagnoux, M. Hamidou, F. Maurier, O. Aumaitre, A. Aouba, T. Quemeneur, J. F. Subra, V. Cottin, J. Sibilia, P. Godmer, P. Cacoub, A. L. Fauchais, E. Hachulla, D. Maucort-Boulch, L. Guillevin, J. C. Lega, and Group French Vasculitis Study, Mortality in systemic necrotizing vasculitides: A retrospective analysis of the French Vasculitis Study Group registry, *Autoimmun. Rev.* 17 (7) (2018) 653-659.
  12. A. Bauza, A. Espana, and M. Idoate, Cutaneous polyarteritis nodosa, *Br. J. Dermatol.*

146 (4) (2002) 694-699.

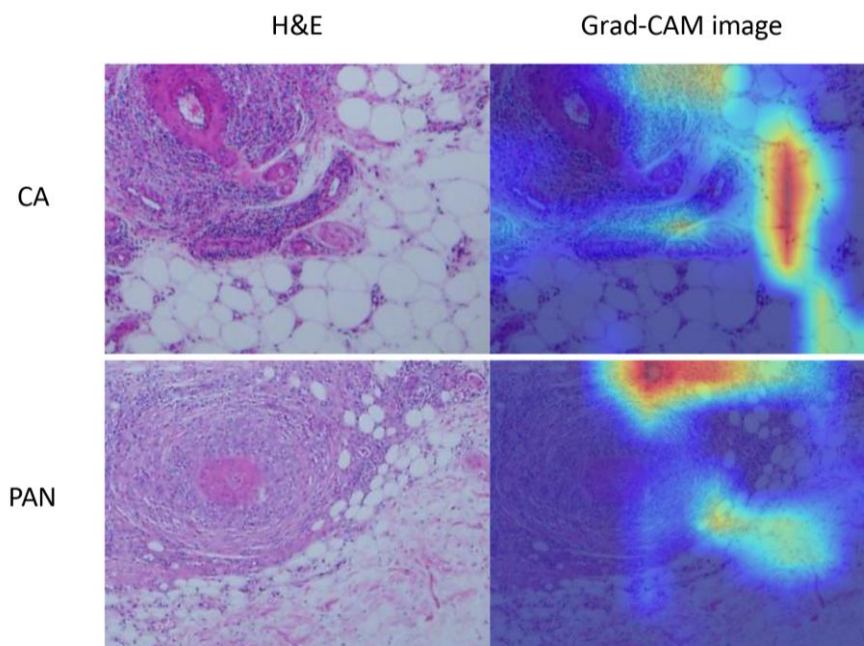
- 13 T. Ikeda, F. Furukawa, T. Kawakami, N. Ishiguro, M. Uzuki, S. Ozaki, K. Katsuoka, T. Kono, S. Kawana, M. Koderu, T. Sawai, Y. Sawada, M. Seishima, A. Tanikawa, K. R. Chen, M. Hasegawa, Vasculitis Committee for Guidelines for the Management of, and Association Vascular Disorders of The Japanese Dermatological, Outline of guidelines for the management of vasculitis and vascular disorders in Japan, 2016 revised edition, J. Dermatol. 45 (2) (2018) 122-127.
- 14 T. Ikeda, T. Kawakami, Y. Arimura, N. Ishiguro, A. Ishizu, F. Ito, T. Ito-Ihara, N. Okiyama, S. Ono, K. Suzuki, K. Sugawara, M. Seishima, M. Koderu, M. Tanaka, M. Hasegawa, F. Furukawa, Y. Yamaguchi, A. Yoshizaki, and vascular disorders of the Japanese Dermatological Association Revised Committee for guidelines for the management of vasculitis, Survey of Japanese dermatological vasculitis specialists on cases of cutaneous arteritis (cutaneous polyarteritis nodosa), J. Dermatol. 47 (5) (2020) 534-537.
- 15 P. Hamet and J. Tremblay, Artificial intelligence in medicine, Metabolism 69S (2017) S36-S40.
- 16 R. Salakhutdinov and G. Hinton, An efficient learning procedure for deep Boltzmann machines, Neural Comput. 24 (8) (2012) 1967-2006.
- 17 Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature 521 (7553) (2015) 436-444.
- 18 P. Korfiatis, T. L. Kline, D. H. Lachance, I. F. Parney, J. C. Buckner, and B. J. Erickson, Residual Deep Convolutional Neural Network Predicts MGMT Methylation Status, J. Digit. Imaging 30 (5) (2017) 622-628.
- 19 I. Ray, G. Raipuria, and N. Singhal, Rethinking ImageNet Pre-training for Computational Histopathology, Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 2022 (2022) 3059-3062.

- 20 Z. Miao, K. M. Gaynor, J. Wang, Z. Liu, O. Muellerklein, M. S. Norouzzadeh, A. McInturff, R. C. K. Bowie, R. Nathan, S. X. Yu, and W. M. Getz, Insights and approaches using deep learning to classify wildlife, *Sci. Rep.* 9 (1) (2019) 8137.
- 21 K. Kikuchi, T. Hoashi, S. Kanazawa, and K. Tamaki, Angiogenic cytokines in serum and cutaneous lesions of patients with polyarteritis nodosa, *J. Am. Acad. Dermatol.* 53 (1) (2005) 57-61.



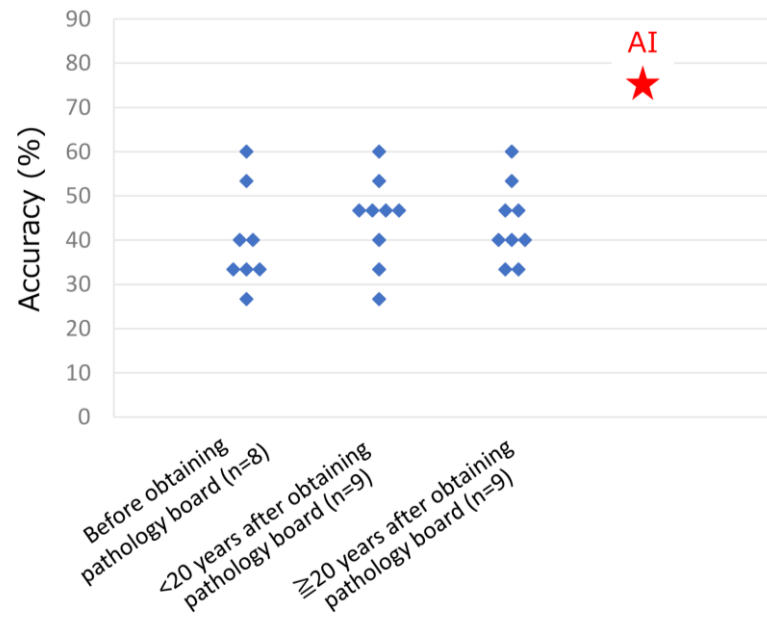
**Figure 1. Learning progress and evaluation of AI.**

(a) Learning progress of AI. As the epoch number increases, the accuracy improves. Blue line, deep learning using training data; orange line, deep learning using validation data. (b) Evaluation of AI. AI after deep learning was challenged to discriminate CA and PAN using test data. A ROC curve was drawn by defining TP and TN as correctly diagnosed CA and PAN, respectively.



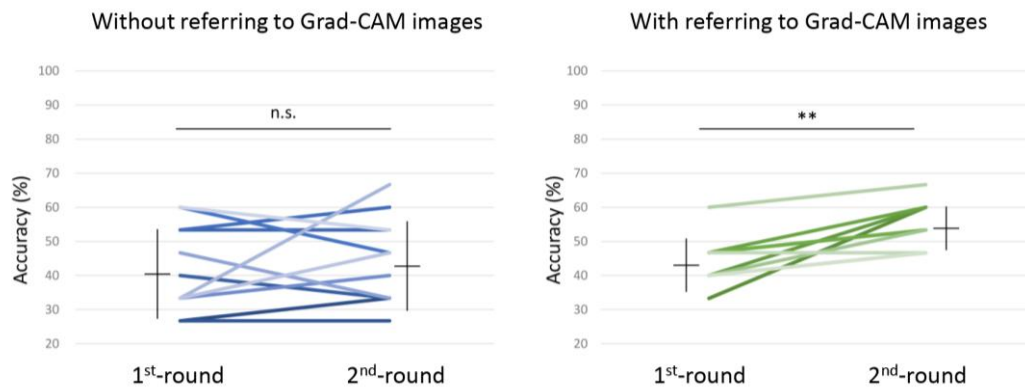
**Figure 2. Grad-CAM analysis.**

Grad-CAM images were created by overlying the input H&E and heatmap images using training data.



**Figure 3. Comparison of AI and pathologists.**

Twenty-six pathologists with different years of experience were challenged to discriminate CA and PAN regarding 15 images ( $340 \times 256$  pixels) randomly selected from test data containing CA and PAN images.



**Figure 4. Improved accuracy of discriminating CA and PAN after referring to Grad-CAM images.**

Among the 26 pathologists who were challenged in the first-round test, 22, excluding the 4 who withdrew, were randomly divided into two groups of 11 each, and one group referred to Grad-CAM images and was challenged in the second-round test, containing CA and PAN images different from those in the first-round test. The other group was challenged in the second-round test without referring to Grad-CAM images.  $**p < 0.01$ ; n.s., nonsignificant.