Explainable Deep Feature Embedding using Multiple Instance Learning for Pathological Image Analysis

Kazuki Uehara[1], Wataru Uegami[2], Hirokazu Nosato[1], Masahiro Murakawa[1], Junya Fukuoka[2] and Hidenori Sakanashi[1]

[1]National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki, Japan
[2]Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki, Japan

Abstract

The development of computer-assisted diagnosis algorithms for analyzing pathological whole slide images (WSIs), consisting of giga-pixels, constitutes an important research topic. Such algorithms are required to be accurate and explainable for their decisions to ensure reliability. WSI classification can be formulated as multiple instance learning (MIL). The general approach of MIL is to train a model that embeds each image patch into feature space and then aggregates feature vectors of the image patches to classify WSIs. Recent MIL approaches have adopted convolutional neural networks (CNNs) with an attention mechanism to train feature embedding and localize the key image patches that trigger WSI classification. The key image patches are regarded as explainable for classification. However, it is unclear why these image patches are selected as important, and it is insufficient for the medical domain. Hence, this paper proposes a dictionary-based explainable CNN method using the MIL paradigm, which identifies some pathological findings in a target WSI and explains them by providing related dictionary items that are considered representative and useful to classifying WSIs. In addition, the method can learn the classifier and construct the dictionary based on the MIL scheme, thereby significantly alleviating the burden by exploiting diagnostic information that can be obtained through daily diagnosis instead of fine-grained annotations. The experimental results showed that the proposed method identified pathological features that contributed to the classification, with high accuracy for two pathological image datasets.

Keywords

Explainable AI, Multiple Instance Learning, Digital Pathology, Convolutional Neural Network

1. Introduction

Pathological image analysis plays an essential role in cancer diagnosis and treatment. To make diagnoses, pathologists inspect stained tissues on glass slides under high-powered magnification, which requires significant effort and time as they must look through large normal tissue regions to recognize the atypical cells and tissues. Thus, developing computer-assisted diagnosis (CAD)

In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, F. van Harmelen (Eds.), Proceedings of the AAAI 2022 Spring Symposium on Machine Learning and Knowledge Engineering for Hybrid Intelligence (AAAI-MAKE 2022), Stanford University, Palo Alto, California, USA, March 21–23, 2022.

k-uehara@aist.go.jp (K. Uehara); uegami.wataru@kameda.jp (W. Uegami); h.nosato@aist.go.jp (H. Nosato); m.murakawa@aist.go.jp (M. Murakawa); fukuokaj@nagasaki-u.ac.jp (J. Fukuoka); h.sakanashi@aist.go.jp (H. Sakanashi)

0000-0002-6628-6668 (K. Uehara); 0000-0002-0285-6270 (W. Uegami); 0000-0003-0332-7028 (H. Nosato); 0000-0002-8406-7426 (M. Murakawa); 0000-0001-8987-908X (H. Sakanashi)

algorithms for pathological image analysis, for example, detecting cancerous cells and atypical tissues, is required, which can significantly reduce the diagnostic workload and improve accuracy. These algorithms must be accurate and their decisions must be interpretable for humans because pathologists are ultimately responsible for their diagnoses. Thus, the trustworthiness of the predictions of the CAD systems must be confirmed for reliability.

Recent developments in whole slide image (WSI) systems using slide scanners enable CADs using machine-learning techniques. Because of their huge image size, WSIs are usually divided into small image patches, and most approaches for classifying WSIs adopt patch-based analysis. However, these approaches are costly as they involve expert pathologists making labels to many image patches.

To deal with these issues, weakly supervised learning for WSI classification is being actively studied [1, 2, 3]. Most of previous approaches on weakly supervised methods investigate a multiple instance learning (MIL), where each WSI is considered as a bag that contains multiple instances of image patches. WSIs are labeled as positive if any of their patches are positive, otherwise negative. The general approach of MIL is to train a model that embeds each image patch into a feature space and then aggregates feature vectors of the image patches to predict WSIs. Recent MIL approaches have adopted convolutional neural networks (CNNs) owing to the recent CNNs achievements in various tasks [4, 5, 6]. CNNs are beneficial to learn feature space embedding and can localize the key instances that trigger bag prediction. Localizing such instances are considered explainable because these instances sometimes correspond to the cancer region [7, 8, 9]. However, it is not clear why these instances are important in the decision-making process because their importance is evaluated in an embedded space that humans cannot understand.

Therefore, we propose a dictionary-based CNN that can explain its decision by presenting images of the findings. The main idea of the method is to provide examples that can be understandable by human users. The examples are selected from a dictionary comprised of a set of representative patterns (e.g., pathological findings) that can influence the classification in a dataset learned by the CNN. (Fig. 1).

Specifically, the proposed method makes decisions to the input WSI, summarizing the results of the comparison between the local region and the dictionary items constructed to have representative findings useful for pathological diagnosis during the model's training process. The dictionary consists of the feature vectors of the images of the findings. The feature vectors are allocated weights indicating their contribution to discriminate normal and lesion areas in the deep feature space using CNN. An area where a local region meets a dictionary item with a large weight is regarded as a notable area. The proposed method presents the decision via visualized notable areas and the dictionary items corresponding to them. In this way, the users understand the diagnosis and how it was generated. To locate notable regions in WSIs, we adopt attention-based MIL [7] that exploits diagnosis information assigned to WSIs. The attention mechanism calculates the relative importance of each instance, that can be flexibly assigned depending on input images. In addition, the use of MIL is practical because it can reduce users' burden for creating a training dataset.

The contributions of this paper are as follows: (i) This paper proposes a dictionary-based CNN that can learn discriminative pathological features as dictionary items for explainability without fine-grained annotations. (ii) The method is verified using two types of pathological image

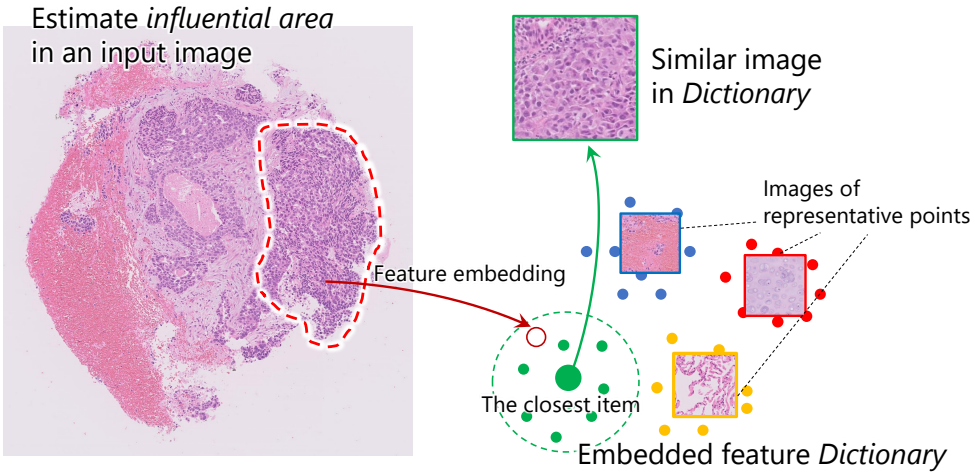

Figure 1: Illustration of the key idea to improve interpretability of the notable region for the decision made by CNNs.

datasets: one is "NCT-CRC-HE-100K" [10], which is a publicly available patch-based dataset for colorectal cancer classification, and the other is a private dataset, which was constructed from WSIs of biopsied lung tissues obtained from Nagasaki University Hospital, Japan. The experimental results showed that the proposed method yielded high classification accuracy while providing the basis behind the decision regardless of the differences in organs.

2. Related Work

2.1. Explainability for CNNs

Explainability or interpretability for CNNs has been actively studied in recent years owing to their importance. Various highlight-based methods that suggest class discriminative pixels in an image have been proposed [11, 12] and popularly used [13]. However, these methods can yield misleading results [14] because they do not explain the basis of the decision made by the CNNs. Several studies have attempted to provide the basis for CNN's decisions [15, 16, 17, 18]. These methods provide evidence for CNN's decisions by comparing observations and representative features in the training dataset. However, they cannot be directly applied to WSI classification because they are designed to classify ordinary-sized images, namely, patch-level classification.

2.2. Multiple instance learning

Several studies have adopted CNNs in pathological image analysis for patch-based classification [19, 20, 21] or semantic segmentation [22, 23]. However, these works require fine-grained annotations, which are pretty expensive because it involves pathologists to make the annotations.

To address this problem, methods based on weakly supervised learning have been actively studied. MIL is a weakly supervised learning method that can train machine learning models

using only the weak labels [24, 25, 26]. Specifically, data called a bag consists of multiple instances or feature vectors and labels are assigned to only entire bags. Learners cannot access to each label of instance in the learning process. The MIL assumption is suitable for medical imaging [27, 28]. Because fine-grained annotations are not required, it has attracted particular attention in the field of digital pathology. Ilse et al. [7] proposed a neural-network-based weighted average pooling that corresponds to the attention mechanism for improving interpretability and flexibility. Chikontwe et al. [8] proposed a framework that can train both instance and bag classifiers with center loss to concentrate the distributions of embedded features extracted from the same bag. Campanella et al. [9] conducted cancer classification using a deep neural network model trained in the MIL framework. Their diagnoses are based on a recurrent neural network that receives top-ranked suspicious samples selected by an instance predictor trained using MIL. Yao et al. [29] efficiently combined multiple-instance segmentation and attention-based pooling for survival period prediction. These MIL methods can identify highly influential instances for making decisions. Often these decisions are interpretable because these instances sometimes correspond to the cancer region. However, it is unclear why these instances are important in the decisions, especially when these instances were selected from irrelevant regions. In contrast, our proposed method enhances explainability by providing notable instances and the dictionary items that are most similar to those instances.

3. Proposed Method

3.1. Problem Formulation

We formulate the WSI classification problem as a MIL classification problem. In MIL, the training dataset $D = \{S_i : i = 1, 2, ..., N\}$ is considered as a bag consisting of multiple instances (WSIs consisting of multiple image patches). For binary classification, let $S_i = \{(x_{i,j}, y_{i,j}) : j = 1, 2, ..., M_i\}$ be a WSI, where $x_{i,j}$ is an image patch in the WSI, and $y_{i,j} \in \{0, 1\}$ is an image patch label that is an unknown. The number of image patches M_i varies for different WSIs. Each WSI has a label $Y_i \in \{0, 1\}$, which is positive ($Y_i = 1$) if the WSI has at least one positive image patch, otherwise it is negative ($Y_i = 0$) as follows:

$$Y_i = \begin{cases} 1, & \sum y_{i,j} > 0 \\ 0, & otherwise, \end{cases} \tag{1}$$

where i is an index of the WSI, while j is an index of the image patch in the S_i.

3.2. Dictionary-based Explainable CNN

Figure 2 shows an overview of the proposed framework. The framework consists of the explainable feature representation and the feature integration. The explainable feature representation is a mapping of patch-level feature vectors to a feature dictionary that can be easily interpreted by humans, whereas the feature integration summarizes all patch-level features in one feature vector for WSI classification.

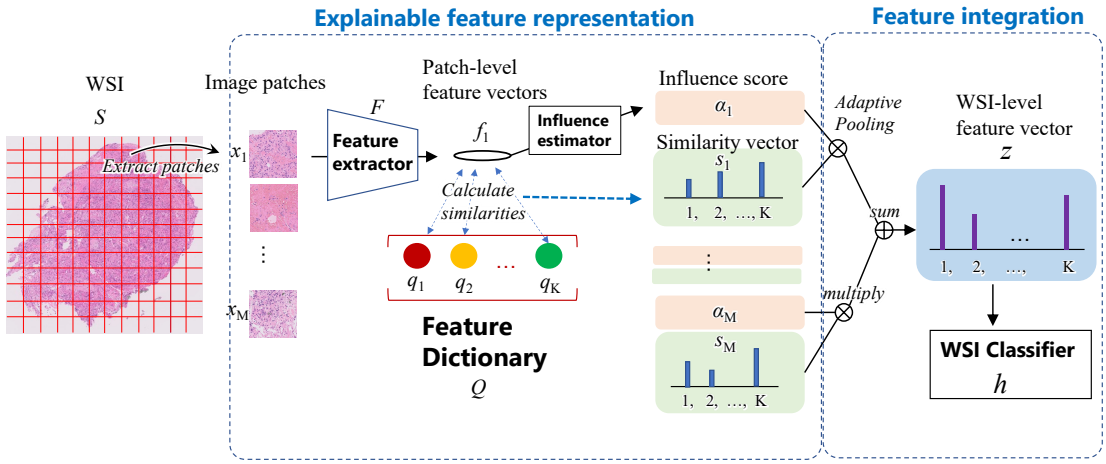

Figure 2: Overview of the proposed framework.

3.2.1. Explainable feature representation

The feature extractor F converts all image patches $\{x_{i,1}, ..., x_{i,M_i}\}$ in a WSI S_i to multiple feature vectors $\{f_{i,1}, ..., f_{i,M_i}\}$. Thereafter, each vector is used for calculating two values: the influence score $a_{i,j}$ and the similarity vector $s_{i,j}$. The influence score determines the degree to which an image feature impacts WSI classification, and the patch-level feature vector with a large influence score has a greater impact on the decision. The influence estimator calculates the scores using the soft-max method to present relative importance among image patches in a WSI as follows:

$$a_{i,j} = \frac{\exp\{w^\intercal \tanh(V f_{i,j}^\intercal)\}}{\sum_{l=1}^{M_i} \exp\{w^\intercal \tanh(V f_{i,l}^\intercal)\}}, \tag{2}$$

where j and l denote an index of vectors, and both w and V are the trainable parameters.

The similarity vector $s_{i,j}$ represents similarity between the patch-level image feature and all items in the dictionary $Q = \{q_k : k = 1, ..., K\}$ consisting K items. Thus, the vector consists of K values and each value in the vector $s_{i,j,k}$ is calculated based on the distance between embedded features of image patches and an item q_k in the dictionary. The distance is transformed similar to [16] as follows:

$$s_{i,j,k} = \log\left(\frac{||f_{i,j} - q_k||_2^2 + 1}{||f_{i,j} - q_k||_2^2 + \epsilon}\right), \tag{3}$$

where k is an indicator of the dictionary items and ϵ is a small positive number.

3.2.2. Feature integration

After calculating the influence scores (equation 2) and the similarity vectors (equation 3), we aggregate the vectors weighted by the influence scores as a WSI-level feature vector z_i for

representation of a WSI S_i so that the diagnosis is robust to the irrelevant image patches. This is similar to the adaptive pooling proposed in [7], and corresponds to the attention mechanism [30] as follows:

$$z_i = \sum_{j=1}^{M_i} a_{i,j} s_{i,j}. \tag{4}$$

Finally, the last layer of the framework, which is the classifier h, makes a decision based on WSI-level features. We adopt single layer perceptron as the classifier. Because each value in the WSI-level features is calculated based on similarities to the input image patches and dictionary items in the embedding space that are non-negative, the dictionary items assigned positive weights to the class c would belong to class c, whereas the items assigned negative wights have a lower probability of being in class c. This property allows easy interpretation of the decisions made by this framework.

3.2.3. Training objectives

We train the network to achieve two objectives (i.e., classification and dictionary construction) as follows:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\mathrm{CE}} + \lambda_2 \mathcal{L}_{\mathrm{DICT}}, \tag{5}$$

where λ_1 and λ_2 are hyper-parameters to balance the loss function. The parameter λ_1 is responsible for classification, while λ_2 is responsible for the similarity of embedded image features and dictionary items in the feature space. We set λ_1 to 1.0 and λ_2 to 0.5.

The classification loss is the cross-entropy function that penalizes misclassification of the WSI based on WSI-level feature vectors. The loss is calculated as follows:

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{i=1}^{N}\sum_{c=1}^{C} Y_{ic} \cdot \log(h \circ z_{ic}), \tag{6}$$

where Y_{ic}, corresponding to WSI S_i is a binary indicator for class c.

The second objective is to construct a dictionary of representative pathological features. It consists of two types of losses and is calculated as follows:

$$\mathcal{L}_{DICT} = \mathcal{L}_{DICT_1} + \mathcal{L}_{DICT_2}, \tag{7}$$

Minimizing L_{DICT_1} requires each dictionary item to be close to the closest patch-level feature in the embedding space. In contrast, minimizing L_{DICT_2} requires every patch-level feature in a mini-batch to be close to one of the closest item in the dictionary, causing the patch-level features to construct clusters around the dictionary item in the embedding space. The losses are calculated as follows:

$$\mathcal{L}_{\mathrm{DICT}_1} = \frac{1}{c}\sum_{k=1}^{c} \min_{j\in[1,M_i]} ||q_k - f_{i,j}||_2^2, \tag{8}$$

$$\mathcal{L}_{\mathrm{DICT}_2} = \frac{1}{M_i}\sum_{j=1}^{M_i} \min_{k\in[1,c]} ||f_{i,j} - q_k||_2^2. \tag{9}$$

4. Experiment

4.1. CNN architecture

The network architecture of the feature extractor in the proposed method consists of three convolutional layers. The kernel sizes for the layers were 4×4, and they were applied in two strides in one padding. After applying the convolutional operation, we adopted a rectified linear unit function and max pooling to promote parameter training.

To construct the dictionary, we set the number of items to be 30 and the dimension of each item to 128. We used the same network architecture for the following experiments.

4.2. CRC dataset

The CRC dataset is a set of image patches from pathological images of colorectal cancer and normal tissues. There are two types of datasets: "NCT-CRC-HE-100K" and "CRC-VAL-HE-7K". "NCT-CRC-HE-100K" contains 100,000 images, while "CRC-VAL-HE-7K" contains 7180 images. These datasets do not onverlap. All images contained in these datasets have the size of 224×224 pixels and are assigned class labels: adipose (ADI), background (BACK), debris (DEB), limphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR), and colorectal adenocarcinoma epithelium (TUM).

We created a new dataset from the CRC dataset for the MIL problem similar to [7]. A bag was made up of randomly selected image patches from the dataset. The number of image patches in the bag was determined based on the Gaussian distribution. We defined that a bag containing at least one "TUM" class of image patches was positive, and otherwise negative. Bags for training and test were selected from "NCT-CRC-HE-100K" and "CRC-VAL-HE-7K", respectively. The numbers of training and test bags were 1,000 and 200. Each bag contained approximately 100 image patches.

4.2.1. Results and discussion

We compared classification accuracy of our method with that of the baseline method proposed by Ilse et al. [7]. The CNN architecture for its feature extractor was the same as that of the proposed method. For the classification of this dataset, our method yielded an accuracy of 92%, whereas the comparison method yielded an accuracy of 90%. Although our method was constrained to provide the rationales of its decisions, its performance was slightly higher than that of the baseline model.

For providing the basis behind the decision as explanation, an example of identified key image patches and corresponding dictionary items are shown in figure 3. The method could localize image patches that are "TUM" class for classifying the bag. The corresponding items denote the most similar pathological feature in the embedding space. The method learned two types of pathological features as the dictionary items that can contribute to predicting bags as positive. The contained image patches in the dictionary items actually belong to the target class.

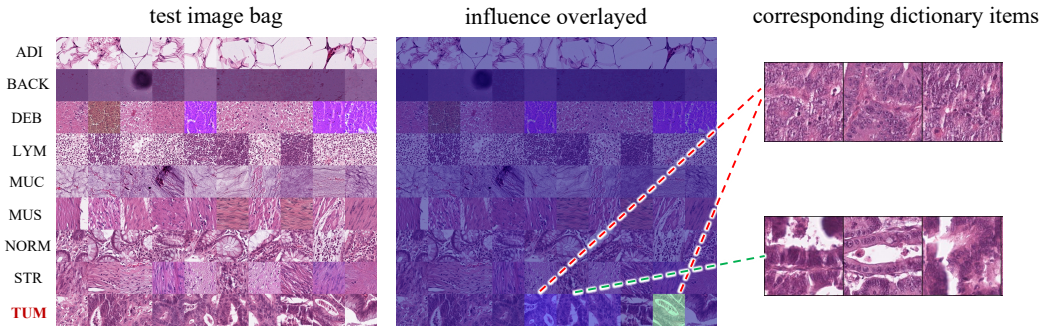

Figure 3: Influential instances and corresponding dictionary items.

4.3. Lung biopsy dataset

We conducted tissue classification in WSIs of a lung obtained by biopsy, provided by Nagasaki University Hospital, Japan. All diagnostic labels, benign or adenocarcinoma (ADC), were provided by pathologists at the university. This study was approved by an institutional review board (ethics committee), and complied with all relevant ethical regulations.

We cropped the tissue images in the WSIs to the size of 2500 $\times$ 2500 pixels to create bag dataset. The total number of benign and ADC images were 323 and 113, respectively. These tissue images were divided into a ratio of 4:1 for training and test datasets. The 20% of training data was used as validation dataset to select the model.

4.3.1. Result and discussion

We compared our method with the method [7] as in the subsection 4.2.1. Figure 4 shows the receiver operating characteristics curve (ROC) for each method. These methods have similar classification performance in terms of the area under curve (AUC) of the ROC. The proposed method showed slightly higher accuracy than that of the baseline model similar to the subsection 4.2.1. Thus, our method has more advantages than the conventional method because it provides explainability with the high classification accuracy.

Figure 5 shows an example of the factors learned by the proposed network as pathological features that contributed to the classification in the dataset. The left and right sides of the figure show the dictionary items that contributed to the classification of the tissue images as ADC, and benign, respectively. The items assigned the highest contribution value to each classification are listed on the top.

Figure 6 compares the dictionary items with high contribution to the classification and the areas annotated by the pathologist as cancer areas. The red region highlights the area annotated by the pathologist. The dictionary items constructed by the proposed method as contributing to the classification of cancer were composed of image patches containing many cancer cells, while the items that contributed to the classification of benign contained few image patches of cancer cells within the instances of the dictionary items.

Figure 7 shows estimated influenced areas for decision making by the proposed method and the areas contained cancer cells and were annotated by the pathologist. In the influence map,

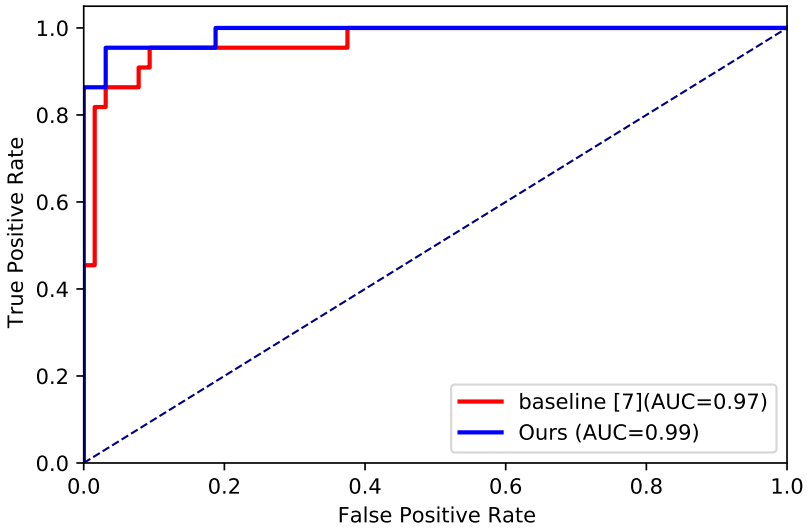

Figure 4: ROC curves for the comparison methods.

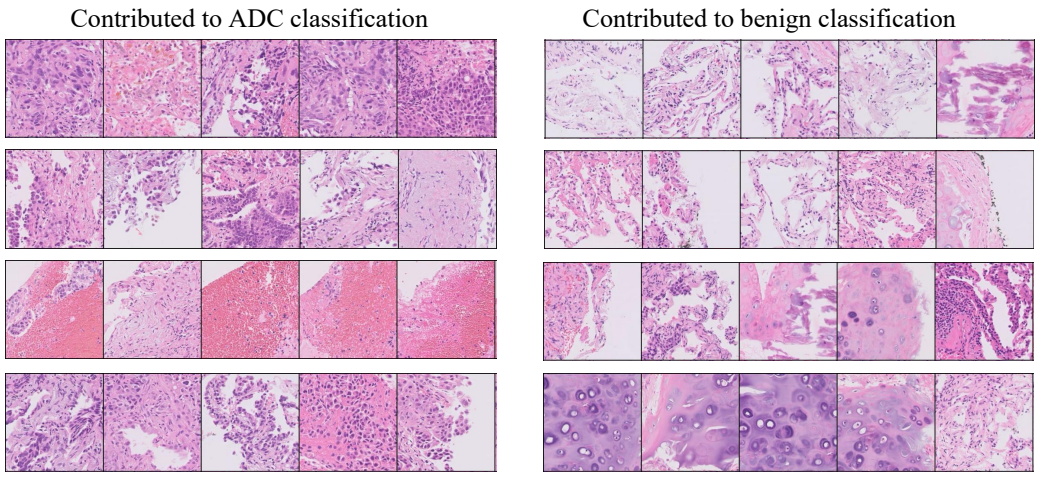

Figure 5: Illustration of the dictionary items constructed by our proposed network.

the brighter the color, the brighter the area that strongly influences the decision. Similarly, in the ground truth, the white area indicates tissues with cancer. Both the brighter part in the influence map and ground truth are similar, which means the proposed method recognized cancer region as notable.

Figure 8 shows the dictionary item that was most similar to the strong activated area in the influence map. A similar item was constructed as a cluster of images that contained cancer cells.

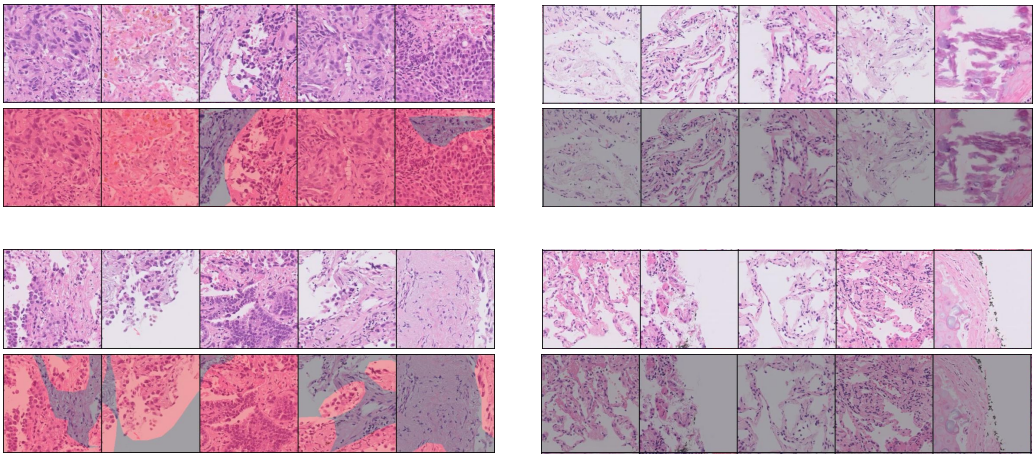

Figure 6: Cancer area in the dictionary items.

This implies that the proposed method made decisions based on highly influential instances, which are similar to the dictionary items that contributed to the classification of ADC in the training dataset. The classification process makes humans easily ascertain the basis of the decision. From the pathologist's perspective, they can judge whether the AI's judgments are appropriate or not based on the presented evidence. Furthermore, if those evidences are not appropriate, it may be possible to improve the quality of the AI model by editing the items in the dictionary.

5. Conclusion

This paper proposed a dictionary-based explainable CNN trained with a multiple instance learning framework. The proposed method overcomes a challenging problem, which is explainability of the basis behind the decisions made by CNN. We verified our method by using two types of pathological images, namely colorectal cancer dataset and the biopsied lung tissue. The results showed that the method yielded high classification accuracy while providing its explanations regardless of the differences in organs. Our method has significant advantages compared with the use of conventional methods. We believe that the proposed method will be helpful in confirming trustworthiness of the diagnosis performed by CNNs. We plan to confirm the dictionary items constructed by our method from a pathology viewpoint.

Acknowledgments

This paper is based on results obtained from a project, JPNP20006, commissioned by the New Energy and Industrial Technology Development Organization (NEDO). Computational resource of AI Bridging Cloud Infrastructure (ABCI) provided by AIST was used.

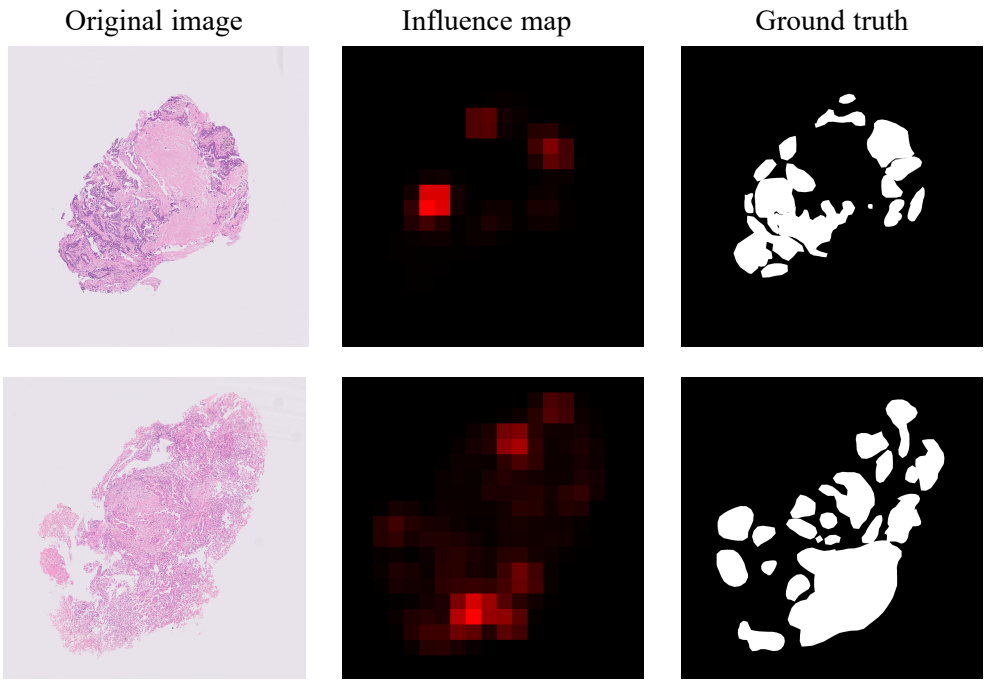

Figure 7: Comparison between high influential area in a WSI and annotated area (Ground truth) by the pathologist.

References

[1] Y. Xu, J. Zhang, E. I.-C. Chang, M. Lai, Z. Tu, Context-constrained multiple instance learning for histopathology image segmentation, in: N. Ayache, H. Delingette, P. Golland, K. Mori (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 623–630.

[2] E. Cosatto, P.-F. Laquerre, C. Malon, H.-P. Graf, A. Saito, T. Kiyuna, A. Marugame, K. Kamijo, Automated gastric cancer diagnosis on H & E-stained sections; ltraining a classifier on a large scale with multiple instance machine learning, in: M. N. Gurcan, A. Madabhushi (Eds.), Medical Imaging 2013: Digital Pathology, volume 8676, International Society for Optics and Photonics, SPIE, 2013, pp. 51 – 59. URL: https://doi.org/10.1117/12.2007047. doi:10.1117/12.2007047.

[3] Y. Xu, J.-Y. Zhu, E. I.-C. Chang, M. Lai, Z. Tu, Weakly supervised histopathology cancer image segmentation and classification, Medical Image Analysis 18 (2014) 591–604. URL: https://www.sciencedirect.com/science/article/pii/S1361841514000188. doi:https://doi.org/10.1016/j.media.2014.01.010.

[4] K. Simonyan, A. Zisserman, Very deep convolutional networks for large scale image recognition, in: International Conference on Learning Representations (ICLR), 2015.

[5] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (2015) 436–444.

[6] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE

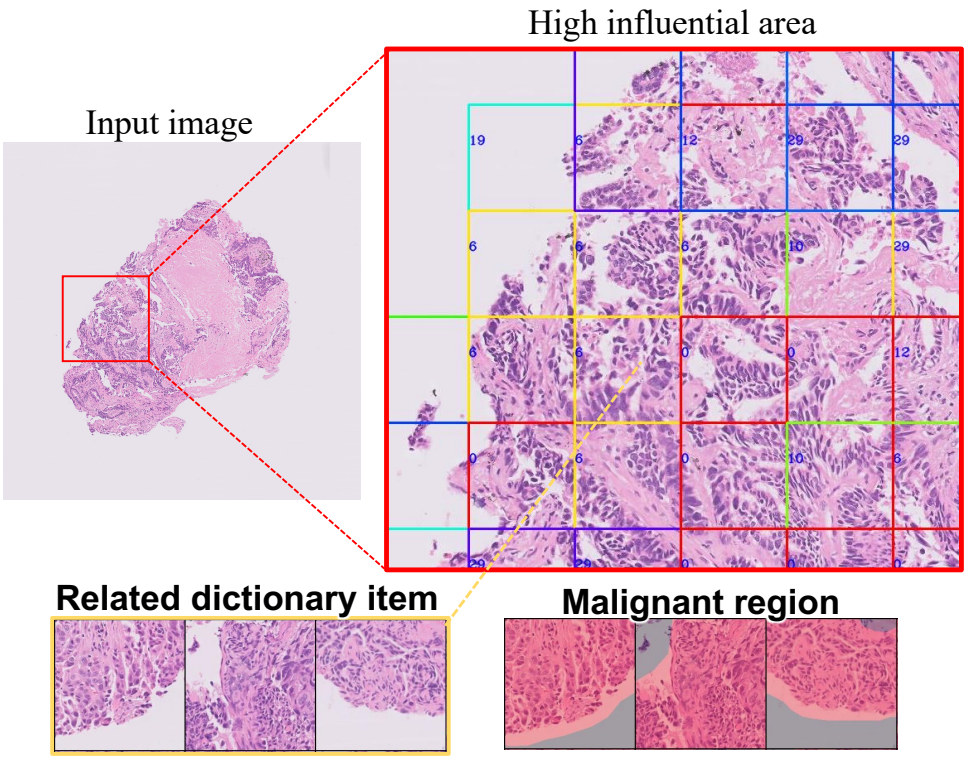

Figure 8: Most similar dictionary item in the high influential area.

Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[7] M. Ilse, J. M. Tomczak, M. Welling, in: Proceedings of the 35th International Conference on Machine Learning (ICML), 2018, pp. 3376–3391.

[8] P. Chikontwe, M. Kim, S. J. Nam, H. Go, S. H. Park, Multiple instance learning with center embeddings for histopathology classification, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Springer International Publishing, Cham, 2020, pp. 519–528.

[9] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. W. K. Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, T. J. Fuchs, Clinical-grade computeational pathology using weakly supervised deep learning on whole slide images, nature medicine 25 (2019) 1301–1309.

[10] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber, L. Jansen, C. C. Reyes-Aldasoro, I. Zörnig, D. Jäger, H. Brenner, J. Chang-Claude, M. Hoffmeister, N. Halama, Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study, PLOS Medicine 16 (2019) 1–22. URL: https://doi.org/10.1371/journal.pmed.1002730. doi:10.1371/journal.pmed.1002730.

[11] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: Visualising

image classification models and saliency maps, CoRR abs/1312.6034 (2013).

[12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: International Conference on Computer Vision (ICCV), IEEE, 2017.

[13] B. Korbar, A. M. Olofson, A. P. Miraflor, C. M. Nicka, M. A. Suriawinata, L. Torresani, A. A. Suriawinata, S. Hassanpour, Looking under the hood: Deep neural network visualization to interpret whole-slide image analysis outcomes for colorectal polyps, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 821–827. doi:10.1109/CVPRW.2017.114.

[14] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, B. Kim, Sanity checks for saliency maps, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 31, Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper/2018/file/294a8ed24b1ad22ec2e7efea049b8737-Paper.pdf.

[15] O. Li, H. L. C. Chen, C. Rudin, Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions, in: Proceedings of AAAI, 2018, pp. 123–456.

[16] C. Chen, O. Li, D. Tao, A. Barnett, C. Rudin, J. K. Su, This looks like that: Deep learning for interpretable image recognition, in: Advances in Neural Information Processing Systems, volume 32, Curran Associates, Inc., 2019.

[17] K. Uehara, M. Murakawa, H. Nosato, H. Sakanashi, Multi-scale explainable feature learning for pathological image analysis using convolutional neural networks, in: 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 1931–1935. doi:10.1109/ICIP40778.2020.9190693.

[18] K. Uehara, M. Murakawa, H. Nosato, H. Sakanashi, Prototype-based interpretation of pathological image analysis by convolutional neural networks, in: Pattern Recognition. ACPR2019. Lecture Notes in Computer Science, volume 12047, 2019, pp. 640–652.

[19] L. Hou, D. Samaras, T. M. Kurç, Y. Gao, J. E. Davis, J. H. Saltz, Patch-based convolutional neural network for whole slide tissue image classification, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society, 2016, pp. 2424–2433. URL: https://doi.org/10.1109/CVPR.2016.266. doi:10.1109/CVPR.2016.266.

[20] A. Cruz-Roa, A. Basavanhally, F. González, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, A. Madabhushi, Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks, in: M. N. Gurcan, A. Madabhushi (Eds.), Medical Imaging 2014: Digital Pathology, volume 9041, International Society for Optics and Photonics, SPIE, 2014, pp. 1 – 15. URL: https://doi.org/10.1117/12.2043872. doi:10.1117/12.2043872.

[21] Y. Wang, T. Peng, J. Duan, C. Zhu, J. Liu, J. Ye, M. Jin, Pathological image classification based on hard example guided cnn, IEEE Access 8 (2020) 114249–114258. doi:10.1109/ACCESS.2020.3003070.

[22] S. Wang, D. M. Yang, R. Rong, X. Zhan, G. Xiao, Pathology image analysis using segmentation deep learning algorithms, The American Journal of Pathology 189 (2019) 1686–1698. URL: https://www.sciencedirect.com/science/article/pii/S0002944018311210. doi:https://doi.org/10.1016/j.ajpath.2019.05.007.

[23] L. Chan, M. Hosseini, C. Rowsell, K. Plataniotis, S. Damaskinos, Histosegnet: Semantic segmentation of histological tissue type in whole slide images, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 10661–10670. doi:10.1109/ICCV.2019.01076.

[24] T. G. Dietterich, R. H. Lathrop, T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles, Artificial Intelligence 89 (1997) 31–71. URL: https://www.sciencedirect.com/science/article/pii/S0004370296000343. doi:https://doi.org/10.1016/S0004-3702(96)00034-3.

[25] R. C. Bunescu, R. J. Mooney, Multiple instance learning for sparse positive bags, in: Proceedings of the 24th International Conference on Machine Learning, ICML '07, Association for Computing Machinery, 2007, p. 105–112. URL: https://doi.org/10.1145/1273496.1273510. doi:10.1145/1273496.1273510.

[26] J. Amores, Multiple instance classification: Review, taxonomy and comparative study, Artificial Intelligence 201 (2013) 81–105.

[27] J. Gang, F. Yuan, Z. Bing, Medical image semantic annotation based on mil, in: 2013 ICME International Conference on Complex Medical Engineering, 2013, pp. 85–90. doi:10.1109/ICCME.2013.6548217.

[28] G. Quellec, G. Cazuguel, B. Cochener, M. Lamard, Multiple-instance learning for medical image and video analysis, IEEE Reviews in Biomedical Engineering 10 (2017) 213–234. doi:10.1109/RBME.2017.2651164.

[29] J. Yao, X. Zhu, J. Jonnagaddala, N. Hawkins, J. Huang, Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks, Medical Image Analysis 65 (2020) 101789. URL: https://www.sciencedirect.com/science/article/pii/S1361841520301535. doi:https://doi.org/10.1016/j.media.2020.101789.

[30] Z. Lin, M. Feng, C. N. dos Santos, M. Yu, B. Xiang, B. Zhou, Y. Bengio, A structured self-attentive sentence embedding, in: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, 2017.