Virchows Archiv (2019) 475:131–138
https://doi.org/10.1007/s00428-019-02594-w

REVIEW AND PERSPECTIVES

Machine learning approaches for pathologic diagnosis

Daisuke Komura[1] · Shumpei Ishikawa[1]

Received: 21 August 2018 / Revised: 7 April 2019 / Accepted: 3 June 2019 / Published online: 20 June 2019

Abstract

Machine learning techniques, especially deep learning techniques such as convolutional neural networks, have been successfully applied to general image recognitions since their overwhelming performance at the 2012 ImageNet Large Scale Visual Recognition Challenge. Recently, such techniques have also been applied to various medical, including histopathological, images to assist the process of medical diagnosis. In some cases, deep learning–based algorithms have already outperformed experienced pathologists for recognition of histopathological images. However, pathological images differ from general images in some aspects, and thus, machine learning of histopathological images requires specialized learning methods. Moreover, many pathologists are skeptical about the ability of deep learning technology to accurately recognize histopathological images because what the learned neural network recognizes is often indecipherable to humans. In this review, we first introduce various applications incorporating machine learning developed to assist the process of pathologic diagnosis, and then describe machine learning problems related to histopathological image analysis, and review potential ways to solve these problems.

Keywords Machine learning · Deep learning · Digital pathology · WSI (whole slide image)

Introduction

Neural network–based deep learning (DL) technologies have evolved rapidly since their overwhelming performance at the 2012 ImageNet Large Scale Visual Recognition Challenge and now exceed human recognition accuracy, at least for general image recognition [1]. In some cases, DL performs better than experienced pathologists in recognizing histopathological images, as exemplified by DL detection of lymph node metastases of breast cancer, first reported in 2017 [2]. The technology progress of digital pathological image analysis is extremely rapid, and many related technologies are also being developed. One such technology merges DL machinery within a microscope and is able to display tumor cells from a microscopic image in real time as augmented reality [3].

Histopathological images differ from general images. For frequently used hematoxylin and eosin (H&E)-stained images, variations in the stainability of individual cases are acknowledged by the image analysis community [4–8], posing a problem that remains unsolved. One issue is that most pathological image data annotations are at a slide level. For example, an entire pathology slide might be labeled as “breast adenocarcinoma,” even though, in most cases, cancer cells are only found in a small part of the slide. This recognition challenge differs substantially from general image tasks such as DL recognition of human faces; consequently, the DL network must be able to learn weakly annotated data.

A major reason why many pathologists are skeptical about the ability of DL technology to accurately recognize histopathological images is because it is difficult for humans to know what exactly the learned neural network recognizes. This problem has also inspired the development of techniques to determine specific features of pathological images that deep neural networks can recognize [9–11]. By comparing this information to the judgments of pathological experts, one can confirm whether the DL network can recognize both trained and novel cases and whether it captures features that are

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00428-019-02594-w) contains supplementary material, which is available to authorized users.

✉ Shumpei Ishikawa
ishum-prm@m.u-tokyo.ac.jp

[1] Department of Preventive Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

histopathologically and biologically rational. It should be very effective to incorporate such steps at each process of parameter optimization.

In this review, we describe some machine learning applications and problems related to histopathological image analysis and review potential solutions for the problems.

Machine learning applications in pathologic diagnosis

Detection, segmentation, or quantification of specific objects

Various applications incorporating machine learning are being developed to assist the process of pathologic diagnosis. Major applications that have been studied so far include detection of specific objects such as cancer cells [12], cell nuclei [13–15], cell divisions [16–18], ducts [19, 20], and blood vessels [21], classification and grading of tumors [2, 22–24], and quantitative evaluation of immunostaining [25]. Supervised learning techniques, which depend on training data with object or diagnostic labels, are usually applied to these applications. It is effective to pick up candidates for tasks that take a lot of time, such as detection of lymph node, cancer micrometastasis, or cytology examination. Some companies such as Indica Labs are developing or have developed such techniques in their products to help pathologists' diagnosis [26–30].

Content-based image retrieval

The process of searching for images that are similar to query images from histopathological databases is referred to as content-based image retrieval (CBIR). CBIR is extremely useful for pathologic diagnosis process [31–35]. Inexperienced (and even experienced) pathologists encounter novel cases with a certain probability. Usually, after discussing with other senior pathologists, they will search for a similar tissue image, turning over page after page within the histopathology atlas. This method of pathology diagnosis is time-consuming, and it is difficult to do a systematic search. CBIR can greatly improve the probability of reaching a correct diagnosis for difficult cases by quickly retrieving similar cases from accessible pathological image databases.

CBIR requires quantification of similarities among images from both histopathological and general image viewpoints. Although this may be an extremely complicated evaluation, texture information extracted by the deep neural network (deep texture) accurately depicts pathological image characteristics in cancer cases and has been used to search for similar images and to achieve accurate results [31].

Machine learning problems related to histopathological image analysis

Normalizing images from different pathology laboratories

H&E staining is used worldwide for pathologic diagnosis, but how the tissue is stained varies case by case, depending on staining and washing conditions. Images generated from such stained slides can vary further depending on the scanner. Except for very rare cases where staining variations are evenly present across compared groups (e.g., comparing cancer and healthy tissue or different forms of cancer), these variations systematically affect the entire learning process because of between-group differences in preparation timing and other factors. Methods for addressing this challenge include color normalization, which adjusts the color of pixels so as to match the color distribution of the source image to that of a target image [4–8] (Fig. 1a), and data augmentation, which creates images of various color patterns by adding random hues and adjusting sharpness, brightness, and contrast [36, 37] (Fig. 1b). Also, since H&E staining solution is a mixture of hematoxylin and eosin dye, reported algorithms can separate the color tones of the image into individual components, which are then normalized, respectively [38].

However, the color tone of the tissue image depends on what the object of the image is. For example, lymph nodes, cartilage, lungs, and brain tissue have different color tones, so it is impossible to forcibly achieve a uniform color tone simply by image-wide color normalization, which compromises pathological information. Such color tone transformation could be effective in cases where the constituent components (epithelium, mesenchyme, blood vessels, etc.) are quite similar. Appropriate reference images for each dataset should be selected while taking this point into account, which is usually difficult. In recent years, generative adversarial networks (GAN) [39], the techniques for learning and generating color tones using "generative model" technology, have also been applied. GAN is composed of two different neural networks called generator and discriminator, and these two networks are trained simultaneously to generate images. The generator tries to produce artificial images which are indistinguishable to the real image for discriminator, while the discriminator, which is generally a convolutional neural network, tries to differentiate real images from artificial images produced by the generator. After the iterative training, the generator becomes able to produce real-looking images. Although it was originally developed to generate real-looking images, other applications including color normalization [40, 41] and data augmentations [42] are emerging recently. These methods for color normalization enable the generation of H&E image dataset with high color consistency, without selecting a reference image.

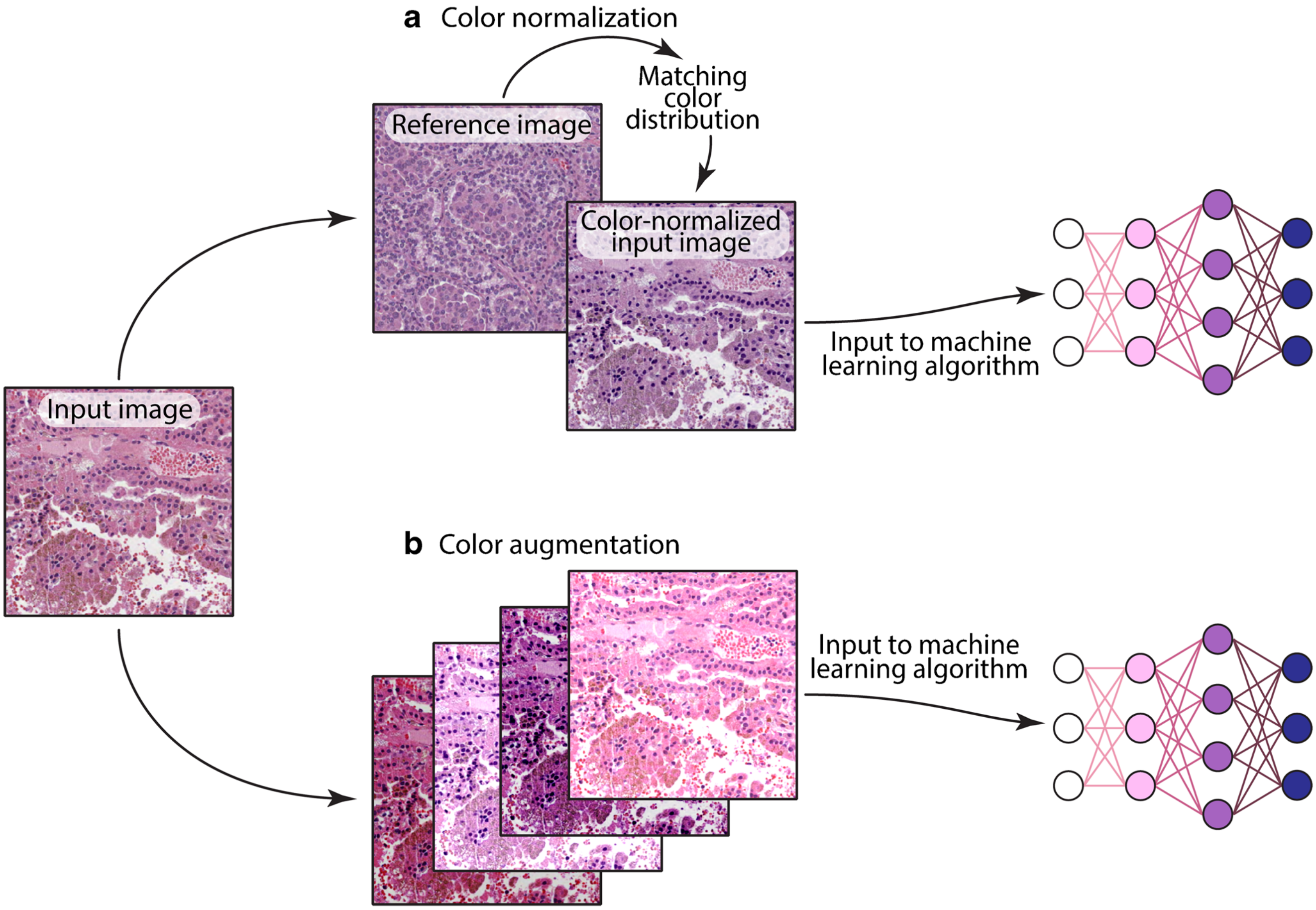

Fig. 1 (**a**) Color normalization and (**b**) color augmentation for color variation problems

Learning from limited numbers of images or images with weak annotations

Transfer learning

Training large neural network models from scratch requires a huge amount of training data to achieve good performance. When only a small number of training samples are available, there are two options to train neural network; training shallower neural network models from scratch or using pre-trained deep neural network models as a starting point. The latter is called transfer learning. Currently, due to the lack of pre-trained neural network models specifically designed for histopathological image analysis, the ones trained with general images such as ImageNet are often used for the purpose. Thus, a pre-trained network trained with a large number of histopathological images is highly demanded. Although there are useful image resources available to researchers, educators, and students, most of the dataset for humans is relatively small and the data is not ready-to-use for machine learning algorithm. Recently, there is a growing number of large-scale datasets available (see reviews [43] for other datasets). Some of the dataset are listed and updated in grand challenges [44]. Although these datasets will be useful for developing pre-trained models for histopathological image analysis, they still have problems. For example, most of the currently available digital pathological images are annotated with the disease name for each case/slide (case-level annotation). In cases of cancer, for example, the cancer may cover only a small portion of the slide; most of the whole slide image (WSI) may therefore be actually non-cancer tissue. Also, the relative position of the cancerous area within the WSI also varies considerably, case by case. This is very different from the process of learning facial images or radiology images, where the region of interest (ROI) is usually present within specific positional and size ranges and the entire ROI is easy to be aligned. Within this context of "weak annotation," DL performance is generally low, and to solve this problem, additional annotations must be added to the image data. In many cases, it is necessary to surround the ROI such as the cancerous region within the WSI, which is a labor-intensive process that requires histopathological expertise and can only be performed on relatively small datasets. Although there are several published pathology image databases, only a few databases with region-level annotations are available [43], and usually, the region-level annotations are performed de novo for a particular purpose.

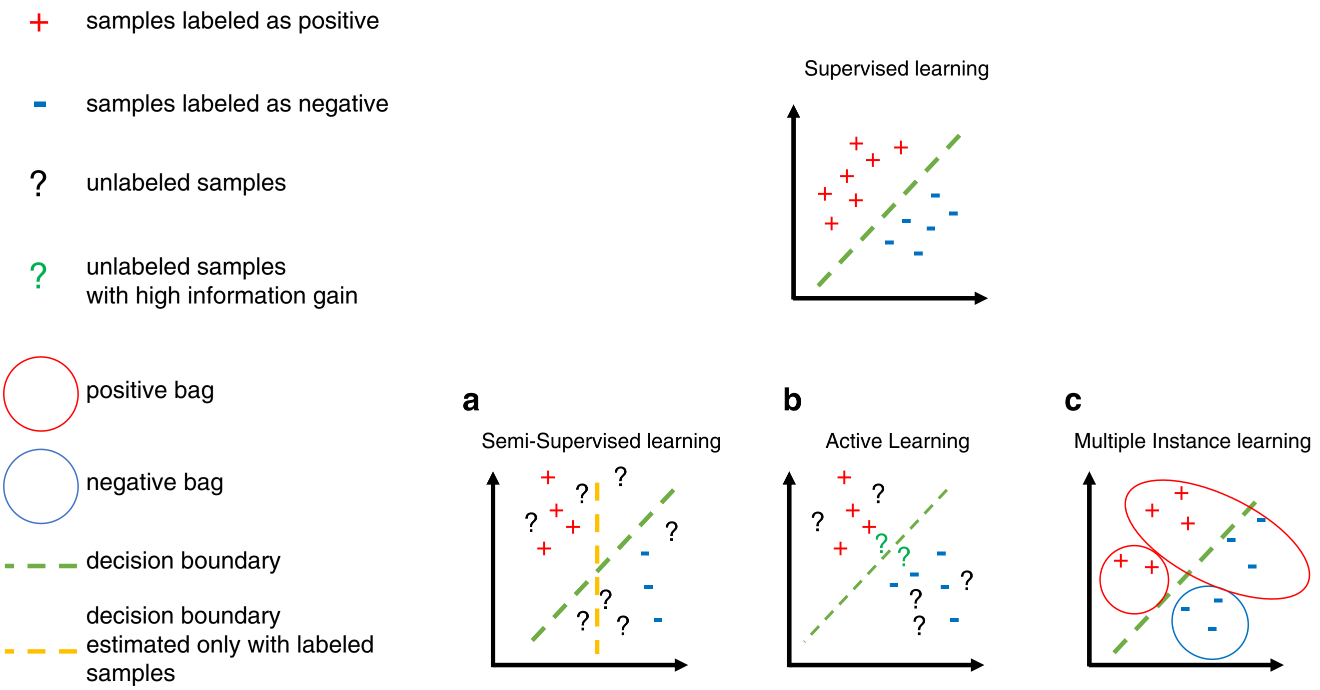

Fig. 2 Learning from limited number of images or images with weak annotations. (**a**) Semi-supervised learning. The yellow line indicates a decision boundary estimated only with labeled samples. In semi-supervised learning, unlabeled samples improve the accuracy of the decision boundary (indicated by the green line). (**b**) Active learning. Unlabeled samples close to the decision boundary can have high information gain. In active learning, such samples are presented to human annotators to be labeled. (**c**) Multiple-instance learning. Instance or bag-level classifiers are developed using bag labels only

Semi-supervised learning and active learning

Under such circumstances, learning technologies for weakly annotated data and small datasets are extremely important. Semi-supervised learning is a machine learning method to improve classification performance based on a small number of annotated images and a relatively large number of non-annotated images [45] (Fig. 2a). This is a valid method for tasks like classifying cancerous vs. normal tissues, where the two groups are separated within the data space. Active learning is an efficient method for identifying and presenting images that are the most difficult to judge during initial training. This method improves the discrimination performance by adding labeling information, facilitating network training by limited labeling costs [46, 47]. Active learning selects one or more unlabeled images that maximize performance gain when the annotation of the images is given (Fig. 2b). Since active learning can be used when there are many unlabeled images, semi-supervised learning techniques are often used together.

Multiple-instance learning

In addition, there are several learning algorithms for weakly annotated images. Multiple-instance learning (MIL) can be used in situations where the cancerous region is included in a part of the image [34, 35], fitting well with much currently available histopathological image information. In MIL settings, a set of instances called “bag” is labeled. In WSIs, instances corresponded to patch or subregion in a WSI and bags correspond to WSIs. In traditional MIL settings, bags are labeled “positive” when at least one instance is positive and “negative” when all the instances are negative (Fig. 2c). Cancer diagnosis fits well to the settings.

Campanella et al. [48] created a cancer tissue classification machine from 8000 WSI of prostate cancer using MIL methods. In addition, a learning method for a mixture of radiation images of a ROI (strong annotation) and weakly annotated images has also been proposed [49]. Such a method is useful for pathological image analysis.

Understanding what features DL neural networks recognize from histology images

To pathologists who are skeptical about the application DL for histopathological images, the main issue is that they cannot know what kind of pathological features are used for classification. Typical DL processes for histopathological image class identification (such as cancer versus normal tissue) only provides the result of the classification and its score; there is no way for the pathologists to know the rationale for classification and the features captured by the DL network.

for dog
(German short-haired pointer)

for laptop

Fig. 3 Visual explanations from deep neural network using Grad-CAM. In the two pictures below, the visible area is important for predicting the corresponding class (left: dog, right: laptop). Grad-CAM images were generated using CLOUDCV (http://gradcam.cloudcv.org)

Black-box visualization techniques

Gradient Class Activation Mapping (Grad-CAM) [9] reveals factors that contribute to the class judgment of an image by randomly perturbing the numerical values of the image and mapping the degree of fluctuation in the class judgment score (Fig. 3). Using this algorithm, we can confirm the basis of class determination and determine whether or not the trained network captured biologically reasonable features. This method is also useful for examining the learning processes of a DL neural network, which are complicated and often undergo local optimization specific to the training dataset. It may be reasonable to integrate Grad-CAM technology during the trial and error processes of hyper-parameter optimization.

Although not yet applied to histopathological images, influence function can provide different but important aspects of the classifiers. Influence functions approximately quantify the effect of any training samples on the value of loss function for any test samples. Although the computational cost for exact computation is huge, efficient approximation algorithm is recently developed [10]. Influence function can be used to identify harmful or helpful training data in classifying test data.

Bayesian deep learning

Bayesian deep learning is an extension of conventional deep learning technology which enables the model to represent their uncertainty based on the Bayesian theory. This technology is for yet to be applied to histopathological images, but have drawn attention in various domains including medical image analysis [50–53]. In Bayesian convolutional neural networks, all (or possibly some) of the weights are probabilistic distribution, instead of a fixed value. Since it can deal with uncertainty in nature, the uncertainties of the model and its prediction can be estimated. Another advantage of Bayesian methods over a non-Bayesian model is that they are more robust against overfitting. As DL models have extremely large number of parameters, a sophisticated method for approximate learning or inference is required. Although it seems a promising technology, development of computationally tractable model while keeping the advantage of Bayesian model is still in its infancy.

Class imbalance

Typically, obtaining samples of different types of tumors or rare diseases are very hard compared with normal/benign samples, and this poses a so-called class imbalance problem in machine learning algorithms. Since classifiers generally try to maximize the overall classification accuracy, misclassification of minority class tends to be undervalued. The simplest approach to the problem in neural network is to give larger weights to the loss for minority classes. Another popular method is to under-sample the majority class or over-sample the minority class, or both (see [54, 55] for detail). Another approach proposed more recently utilizes specialized GAN which generates new minority-class images to restore the balance in imbalanced datasets [42], although its effectiveness in histopathological image analysis should be assessed.

Incorporating external knowledge into the learning framework

Clinicians make decisions based not only on the image data itself but also on external knowledge such as molecular data. Incorporating such knowledge into deep learning framework could improve the diagnostic accuracy. This is called multimodal deep learning. Since dimension of the image data is generally much larger than other types of clinical data, the model might ignore important clinical information during the training if the clinical data is simply added to the input. Currently, there are two strategies to tackle this problem; feature concatenation and ensemble classifier. In feature concatenation, extracted features from the image in the convolutional layers are merged in the last fully connected layers. Mobadersany et al. adopted this strategy to predict patient's survival using pathological image and genomic data [56]. Another possible approach is to combine the final output of DL models based on images with those of other classifiers based on clinical information using ensemble classifiers such as random forest.

Uncertainty in reference standard

Even if we successfully obtain an annotation, some of the annotations can still be wrong. It is very important to get a reliable label since mislabeling could have an adverse impact on the classification accuracy. The molecular profile or other test such as immunohistochemistry [2] partially solves this problem, but not completely. Additionally, it is sometimes impossible to make clear biological delineations in subtle cases that even experts have difficulties in distinguishing between categories like cancer versus normal. It seems that we have no clear answer to this problem yet.

Currently, the reference standard is often provided by one expert pathologist. However, since the pathologic diagnosis includes subjective interpretation, it is more desirable to make an effort to collect diagnoses from multiple pathologists and to solve the disagreement to improve the model. This is demonstrated in a recently published papers dealing with diabetic retinopathy diagnosis based on retinal images [57]. The report has shown that thorough debate by a group of retinal specialists on disagreed cases until everyone reaches agreement on the final grade lead to more accurate diagnosis and identification of subtle findings than simply taking a majority vote. Another possible solution is to incorporate the annotator's uncertainty into the learning framework such as designing loss function which gives weights to the annotation uncertainty.

Perspective and conclusions

As mentioned above, the major obstacle facing machine learning of pathological images is inadequate image dataset annotation. The current development stage differs considerably from that of the radiation image where a lot of image data at the same angle and position have already been accumulated. This may be confusing to some who thinks that all types of medical image analysis in general are making progress uniformly.

Deep learning technologies are rapidly progressing, and it may overcome a number of current problems in a few years. Related to this, AlphaGo, developed by Google, trained itself to become a strong network by creating its own training data. GAN technology is used for pathological data analysis to automatically prepare image datasets necessary for subsequent DL [58], and we eagerly await future uses of this technology. As the observed rapid developments are accompanied by frequent updates, it is unlikely that any one technology will emerge as gold standard for machine learning technology for image analysis.

It is needless to say that developing new technology is important, but in order to make the technology available in clinical practice, it is also important to perform appropriate realistic evaluation of the developed models. Such evaluation includes observer studies, diagnosis under realistic time constraints, and diagnosis under the access to other regular diagnostic information such as immunohistochemistry.

Many pathologists may be concerned about whether AI will deprive work from human workers. However, it seems too pessimistic for several reasons. First, even though deep learning could surpass the performance of human pathologists in terms of accuracy in simple classification task with sufficient number of training data, the current result showing the superiority of AI is evaluated in limited settings, and the potential utility of machine learning algorithms for pathologic diagnosis in a clinical setting requires further validation as described above. Additionally, there are some pieces of evidence showing that cooperation of human and AI will boost the diagnostic performance as deep learning and human fail in different ways [59]. Therefore, we expect that rather than competing against AI, human pathologists may choose to cooperate with AI and machine learning techniques to improve the efficiency of pathologic diagnosis.

Acknowledgments This study was supported by the Practical Research for Innovative Cancer Control from the Japan Agency for Medical Research and Development (AMED) (S.I.).

Contributions Ishikawa S and Komura D wrote and reviewed the manuscript.

Funding This research was supported by AMED under the Practical Research for Innovative Cancer Control, grant number JP19ck0106400 (S.I.).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. Int J Comput Vis 115:211–252. https://doi.org/10.1007/s11263-015-0816-y
2. Bejnordi BE, Veta M, van DPJ et al (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318:2199–2210. https://doi.org/10.1001/jama.2017.14585
3. An augmented reality microscope for cancer detection. In: Google AI Blog. http://ai.googleblog.com/2018/04/an-augmented-reality-microscope.html. Accessed 13 Aug 2018
4. Ciompi F, Geessink O, Bejnordi BE, et al (2017) The importance of stain normalization in colorectal tissue classification with convolutional networks. arXiv:170205931 [cs]
5. Bejnordi BE, Litjens G, Timofeeva N et al (2016) Stain specific standardization of whole-slide histopathological images. IEEE Trans Med Imaging 35:404–415. https://doi.org/10.1109/TMI.2015.2476509
6. Khan AM, Rajpoot N, Treanor D, Magee D (2014) A nonlinear mapping approach to stain normalization in digital histopathology

images using image-specific color deconvolution. IEEE Trans Biomed Eng 61:1729–1738. https://doi.org/10.1109/TBME.2014.2303294

7. Li X, Plataniotis KN (2015) A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics. IEEE Trans Biomed Eng 62:1862–1873. https://doi.org/10.1109/TBME.2015.2405791
8. Sethi A, Sha L, Vahadane AR, Deaton RJ, Kumar N, Macias V, Gann PH (2016) Empirical comparison of color normalization methods for epithelial-stromal classification in H and E images. J Pathol Inform 7:17. https://doi.org/10.4103/2153-3539.179984
9. Selvaraju RR, Cogswell M, Das A, et al (2016) Grad-CAM: visual explanations from deep networks via gradient-based localization. arXiv:161002391 [cs]
10. Koh PW, Liang P (2017) Understanding black-box predictions via influence functions. arXiv:170304730 [cs, stat]
11. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M (2014) Striving for simplicity: the all convolutional net. arXiv:14126806 [cs]
12. Bayramoglu N, Kannala J, Heikkilä J (2016) Deep learning for magnification independent breast cancer histopathology image classification. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp 2440–2445
13. Hou L, Nguyen V, Samaras D, et al (2017) Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. arXiv:170400406 [cs]
14. Transfer learning for cell nuclei classification in histopathology images | SpringerLink. https://link.springer.com/chapter/10.1007/978-3-319-49409-8_46. Accessed 22 Nov 2017
15. Xing F, Yang L (2016) Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. IEEE Rev Biomed Eng 9:234–263. https://doi.org/10.1109/RBME.2016.2515127
16. Roux L, Racoceanu D, Loménie N, Kulikova M, Irshad H, Klossa J, Capron F, Genestie C, Naour GL, Gurcan MN (2013) Mitosis detection in breast cancer histological images an ICPR 2012 contest. J Pathol Inform 4:8. https://doi.org/10.4103/2153-3539.112693
17. Mitosis detection in breast cancer histological images. http://ludo17.free.fr/mitos_2012/index.html. Accessed 29 Nov 2017
18. Veta M, van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A, Gonzalez F, Larsen ABL, Vestergaard JS, Dahl AB, Cireşan DC, Schmidhuber J, Giusti A, Gambardella LM, Tek FB, Walter T, Wang CW, Kondo S, Matuszewski BJ, Precioso F, Snell V, Kittler J, de Campos TE, Khan AM, Rajpoot NM, Arkoumani E, Lacle MM, Viergever MA, Pluim JPW (2015) Assessment of algorithms for mitosis detection in breast cancer histopathology images. Med Image Anal 20:237–248. https://doi.org/10.1016/j.media.2014.11.010
19. Chen H, Qi X, Yu L, Heng PA (2016) DCAN: deep contour-aware networks for accurate gland segmentation. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2487–2496
20. Sirinukunwattana K, Pluim JPW, Chen H, Qi X, Heng PA, Guo YB, Wang LY, Matuszewski BJ, Bruni E, Sanchez U, Böhm A, Ronneberger O, Cheikh BB, Racoceanu D, Kainz P, Pfeiffer M, Urschler M, Snead DRJ, Rajpoot NM (2017) Gland segmentation in colon histology images: the glas challenge contest. Med Image Anal 35:489–502. https://doi.org/10.1016/j.media.2016.08.008
21. Kather JN, Weis C-A (2016) Validation data set for automatic blood vessel segmentation in colorectal cancer histology (IHC)
22. Gupta V, Bhavsar A (2017) Breast cancer histopathological image classification: is magnification important? In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 769–776
23. Araújo T, Aresta G, Castro E, Rouco J, Aguiar P, Eloy C, Polónia A, Campilho A (2017) Classification of breast cancer histology images using convolutional neural networks. PLoS One 12:e0177544. https://doi.org/10.1371/journal.pone.0177544
24. CAMELYON17. https://camelyon17.grand-challenge.org/. Accessed 21 Aug 2017
25. Liu J, Xu B, Zheng C, et al (2018) An end-to-end deep learning histochemical scoring system for breast cancer tissue microarray. arXiv:180106288 [cs]
26. HALO AI – Indica Labs. http://www.indicalab.com/halo-ai/. Accessed 20 Dec 2018
27. PAIGE. https://www.paigeai.com/. Accessed 20 Dec 2018
28. PathAI. https://www.pathai.com/. Accessed 20 Dec 2018
29. Proscia. https://proscia.com/. Accessed 20 Dec 2018
30. Contextvision. http://www.contextvision.com/pathology/. Accessed 20 Dec 2018
31. Luigi: large-scale histopathological image retrieval system using deep texture representations | bioRxiv. https://www.biorxiv.org/content/early/2018/07/19/345785. Accessed 13 Aug 2018
32. Caicedo JC, González FA, Romero E (2011) Content-based histopathology image retrieval using a kernel-based semantic annotation framework. J Biomed Inform 44:519–528. https://doi.org/10.1016/j.jbi.2011.01.011
33. Mehta N, Raja'S A, Chaudhary V (2009) Content based sub-image retrieval system for high resolution pathology images using salient interest points. In: Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE. IEEE, pp 3719–3722
34. Qi X, Wang D, Rodero I, Diaz-Montes J, Gensure RH, Xing F, Zhong H, Goodell L, Parashar M, Foran DJ, Yang L (2014) Content-based histopathology image retrieval using CometCloud. BMC Bioinform 15:287
35. Sridhar A, Doyle S, Madabhushi A (2015) Content-based image retrieval of digitized histopathology in boosted spectrally embedded spaces. J Pathol Inform 6:41. https://doi.org/10.4103/2153-3539.159441
36. Lafarge MW, Pluim JPW, Eppenhof KAJ, et al (2017) Domain-adversarial neural networks to address the appearance variability of histopathology images. arXiv:170706183 [cs]
37. ScanNet: a fast and dense scanning framework for metastatic breast cancer detection from whole-slide images - semantic scholar. /paper/ScanNet-A-Fast-and-Dense-Scanning-Framework-for-Me-Lin-Chen/9484287f4d5d52d10b5d362c462d4d6955655f8e. Accessed 22 Nov 2017
38. Vahadane A, Peng T, Sethi A, Albarqouni S, Wang L, Baust M, Steiger K, Schlitter AM, Esposito I, Navab N (2016) Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans Med Imaging 35:1962–1971. https://doi.org/10.1109/TMI.2016.2529665
39. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al (2014) Generative adversarial networks
40. Shaban MT, Baur C, Navab N, Albarqouni S (2018) StainGAN: stain style transfer for digital histological images. arXiv:180401601 [cs]
41. Zanjani FG, Zinger S, Bejnordi BE, et al (2018) Histopathology stain-color normalization using deep generative models
42. Mariani G, Scheidegger F, Istrate R, et al (2018) BAGAN: data augmentation with balancing GAN
43. Komura D, Ishikawa S (2018) Machine learning methods for histopathological image analysis. Comput Struct Biotechnol J 16:34–42. https://doi.org/10.1016/j.csbj.2018.01.001
44. grand-challenges - Home. https://grand-challenge.org/. Accessed 17 Dec 2018
45. Peikari M, Zubovits J, Clarke G, Martel AL (2015) Clustering analysis for semi-supervised learning improves classification

performance of digital pathology. In: Machine learning in medical imaging. Springer, Cham, pp 263–270

46. Doyle S, Monaco J, Feldman M, Tomaszewski J, Madabhushi A (2011) An active learning based classification strategy for the minority class problem: application to histopathology annotation. BMC Bioinform 12(424). https://doi.org/10.1186/1471-2105-12-424
47. Padmanabhan RK, Somasundar VH, Griffith SD, Zhu J, Samoyedny D, Tan KS, Hu J, Liao X, Carin L, Yoon SS, Flaherty KT, DiPaola RS, Heitjan DF, Lal P, Feldman MD, Roysam B, Lee WMF (2014) An active learning approach for rapid characterization of endothelial cells in human tumors. PLoS One 9: e90495. https://doi.org/10.1371/journal.pone.0090495
48. [1805.06983] Terabyte-scale deep multiple instance learning for classification and localization in pathology. https://arxiv.org/abs/1805.06983. Accessed 13 Aug 2018
49. Li Z, Wang C, Han M, et al (2017) Thoracic disease identification and localization with limited supervision. arXiv:171106373 [cs, stat]
50. Gal Y, Ghahramani Z (2015) Bayesian convolutional neural networks with Bernoulli approximate variational inference. arXiv: 150602158 [cs, stat]
51. Novak R, Xiao L, Lee J, et al (2018) Bayesian convolutional neural networks with many channels are Gaussian processes. arXiv: 181005148 [cs, stat]
52. Shridhar K, Laumann F, Maurin AL, et al (2018) Bayesian convolutional neural networks with variational inference. arXiv: 180605978 [cs, stat]
53. Zhao G, Liu F, Oler JA, Meyerand ME, Kalin NH, Birn RM (2018) Bayesian convolutional neural network based MRI brain extraction on nonhuman primates. NeuroImage 175:32–44. https://doi.org/10.1016/j.neuroimage.2018.03.065
54. Buda M, Maki A, Mazurowski MA (2017) A systematic study of the class imbalance problem in convolutional neural networks. https://doi.org/10.1016/j.neunet.2018.07.011
55. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G (2017) Learning from class-imbalanced data: review of methods and applications. Expert Syst Appl 73:220–239. https://doi.org/10.1016/j.eswa.2016.12.035
56. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, Brat DJ, Cooper LAD (2018) Predicting cancer outcomes from histology and genomics using convolutional networks. Proc Natl Acad Sci U S A 115:E2970–E2979. https://doi.org/10.1073/pnas.1717139115
57. Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, Peng L, Webster DR (2018) Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology 125:1264–1272. https://doi.org/10.1016/j.ophtha.2018.01.034
58. Hou L, Agarwal A, Samaras D, et al (2017) Unsupervised histopathology image synthesis. arXiv:171205021 [cs]
59. Sayres R, Taly A, Rahimy E, Blumer K, Coz D, Hammel N, Krause J, Narayanaswamy A, Rastegar Z, Wu D, Xu S, Barb S, Joseph A, Shumski M, Smith J, Sood AB, Corrado GS, Peng L, Webster DR (2018) Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. Ophthalmology 0:552–564. https://doi.org/10.1016/j.ophtha.2018.11.016