Diagnostic Classification of Cystoscopic Images Using Deep Convolutional Neural Networks

abstract

Purpose The recognition of cystoscopic findings remains challenging for young colleagues and depends on the examiner's skills. Computer-aided diagnosis tools using feature extraction and deep learning show promise as instruments to perform diagnostic classification.

Materials and Methods Our study considered 479 patient cases that represented 44 urologic findings. Image color was linearly normalized and was equalized by applying contrast-limited adaptive histogram equalization. Because these findings can be viewed via cystoscopy from every possible angle and side, we ultimately generated images rotated in 10-degree grades and flipped them vertically or horizontally, which resulted in 18,681 images. After image preprocessing, we developed deep convolutional neural network (CNN) models (ResNet50, VGG-19, VGG-16, InceptionV3, and Xception) and evaluated these models using F1 scores. Furthermore, we proposed two CNN concepts: 90%-previous-layer filter size and harmonic-series filter size. A training set (60%), a validation set (10%), and a test set (30%) were randomly generated from the study data set. All models were trained on the training set, validated on the validation set, and evaluated on the test set.

Results The Xception-based model achieved the highest F1 score (99.52%), followed by models that were based on ResNet50 (99.48%) and the harmonic-series concept (99.45%). All images with cancer lesions were correctly determined by these models. When the focus was on the images misclassified by the model with the best performance, 7.86% of images that showed bladder stones with indwelling catheter and 1.43% of images that showed bladder diverticulum were falsely classified.

Conclusion The results of this study show the potential of deep learning for the diagnostic classification of cystoscopic images. Future work will focus on integration of artificial intelligence–aided cystoscopy into clinical routines and possibly expansion to other clinical endoscopy applications.

Clin Cancer Inform. © 2018 by American Society of Clinical Oncology

Okyaz Eminaga

Nurettin Eminaga

Axel Semjonow

Bernhard Breil

Author affiliations and support information (if applicable) appear at the end of this article.

Corresponding author: Okyaz Eminaga, MD, Department of Urology, Department of Urology, University Hospital of Cologne, Department of Urology, Stanford University School of Medicine, 300 Pasteur Dr, Stanford, CA 94305-5118; e-mail: okyaz.eminaga@gmail.com.

INTRODUCTION

Cystoscopic examination is one of the key semi-invasive diagnostic procedures for urologic and gynecologic diseases. Cystoscopic findings are diverse and range from normal results to urothelial carcinoma of the bladder. However, the recognition of cystoscopic findings remains challenging, and additional invasive procedures, such as biopsy of suspect lesions, are often required. Currently, the outcome of these procedures depends on the examiner's skills, differences of which lead to variation in the interpretation of cystoscopic findings. In such cases, computer-aided diagnosis tools that use feature extraction and deep learning show promise as instruments to perform diagnostic classification and to achieve reproducible cystoscopic findings, which thereby reduces the number of unnecessary biopsies performed. Importantly, computer-aided diagnosis tools comprise a non-invasive and convenient approach for urologists that can be implemented using available infrastructures and current technologies. Recent advances in deep-learning algorithms, imaging techniques, and computational capabilities have facilitated robust pattern recognition and data-structure determination with different data sets.[1-4] However, the applicability of deep learning in diagnostic imaging for cystoscopy has not yet been evaluated. To bridge this gap, this paper introduces models that apply a deep-learning approach to classify cystoscopic images along with strong evidence of their validity.

MATERIALS AND METHODS

We used the digital atlas for cystoscopy, which covers 44 cystoscopic findings that can frequently be seen during the clinical routine and consists of 479 images.[5] The author of this image atlas permitted the use of the open image data in compliance with data privacy regulations.[5] This study followed German data regulations and the Declaration of Helsinki. Table 1 lists all the cystoscopic findings considered by this study. Because these results can be viewed from every possible angle and side by cystoscopy, we generated images rotated in 10-degree grades and flipped them vertically or horizontally, which resulted in 18,681 images.

For image processing, image color was equalized by applying contrast-limited adaptive histogram equalization (the clip limit was set to 12, and a tile grid size of 3×3 was defined). During the cystoscopic examination, these findings are usually captured in the middle of the image, which has a dark area that surrounds the cystoscopic findings; lighting disbalance; and blurring around the edges, because the plastic part of the camera adapter is attached to the edge of the cystoscopic objective, which causes a shadow effect. Therefore, we created a mask from each image that is applied to remove the blurred edge and the noise from the dark area of the normalized image (Fig 1); for mask generation, the color gradient of the original image was reduced to 64 colors using the posterization effect. After that, a binary mask was generated after the image was converted to grayscale and the threshold was set as 10. To remove the remaining noise from the binary mask, we applied erosion first and then dilation (opening procedure) to each mask; then, we repeated in reverse (ie, dilation followed by erosion; closing procedure). Finally, the normalized image was clipped by the binary mask to get an image that included only the focused finding in the middle of the original image.

We used the neural network models ResNet50,[6] VGG-19,[7] VGG-16,[7] InceptionV3,[8] and Xception[9] to classify the cystoscopic images and compared their classification performances. Also, we developed two deep convolutional neural networks (CNN) architectures: in the first architecture, the filter size of the next convolutional layer was defined as 90% the filter size of the previous layer; in the second, a deep CNN model, the harmonic series of the initial filter size n, $(1/2)n$, $(1/3)n...$, $(1/k)n$ (where n is the original filter size and k is the serial number of the convolutional layer) was applied to define the filter size of each convolutional layer. Scaled exponential linear units were applied to the activation function in these two models.[10]

We trained these models for the diagnostic classification of all cystoscopic images. Additional information about the CNN architecture of each model can be obtained from the Data Supplement. The dropout regularization technique was used to reduce the risk of overfitting by randomly selecting nodes to be dropped at a given probability for each weight update cycle.[11] Regularization mechanisms, such as Dropout and L1/L2-weight regularization, were deactivated at testing time. The solver algorithm Stochastic Gradient Descent was used to compute adaptive learning rates for each parameter, which thereby optimized CNN. We repeated the training phase of each model using the optimization algorithm Adaptive Moment Estimation (ie, ADAM) instead of Stochastic Gradient Descent.[12] The maximum number of training epochs was set to 40, and an early stopping algorithm was used to stop training after five consecutive epochs did not improve the mean squared error for the output error estimation in each model. Batch size was defined as 18 because of memory limitations. Furthermore, a training set (60%), validation set (10%), and test set (30%) were randomly generated from the study data set. All models were trained on the training set, validated on the validation set, and evaluated on the test set. Classification performance of the final test set was evaluated once using classification accuracy, precision, recall, F-measure (F1 score) and a confusion matrix. We generated an activation map for each class to identify the structures associated with the activation of the function and the determination of the diagnostic class. Our analyses were based on Python 2.7 (Python Software Foundation, Wilmington, DE) and applied the Keras framework and DIGITS (NVIDIA, Santa Clara, CA), which is built on the Caffe framework, to develop the models. All analyses were performed on a server with an Intel i7 processor with 8 GB RAM (Intel, Santa Clara, CA) and an NVIDIA GeForce GTX 1080 Ti GPU with 11 GB VRAM.

Table 1. Cystoscopic Findings Considered for the Diagnostic Classification of Cystoscopic Images

Diagnoses	No. of Occurrences	No. of Images
Abnormal vessels: bleeding	17	663
Bilharzia	6	234
Bladder amyloidosis	4	156
Bladder diverticulum	12	468
Bladder neck sclerosis and barrier	7	273
Bladder septum or membrane	8	312
Bladder stones and indwelling catheter	12	468
Bladder stones: ureteral and urethral stones	35	1,365
Bladder tuberculosis	10	390
Bladder wall edema - decubitus - bleeding	11	429
Bricker conduit: coecocystoplasty	8	312
Catheter balloon encrustments	23	897
Cystitis cystica + glandularis	19	741
Diverticulum of urethra in man	6	234
Diverticulum of urethra in woman	3	117
False passage	7	273
Hemangioma B + microscopic examination	2	78
Interstitial (Hunner): cystitis	6	234
Intrusion from outside the bladder	7	273
Large bladder bezoar	6	234
Large prostate + third lobe	11	429
Moderate enlargement of the prostate	6	234
Nonspecific cystitis: inflammation	21	819
Normal bladder	34	1,326
Normal prostate	15	585
Normal urethra: female	8	312
Normal urethra: male	16	624
Primary adenocarcinoma	5	195
Prostatic cancer	9	351
Signet ring cell cancer	4	156
Slight enlargement of the prostate	4	156
Squamous carcinoma	5	195
Trabeculation	7	273
Urethral stricture: fistulation, diverticulum	10	390
Urethritis caused by indwelling catheter	8	312
Urothelial cancer grade 1	7	273
Urothelial cancer grade 2	31	1,209
Urothelial cancer grade 3	19	741
Urothelial cancer in bladder diverticulum	15	585
Urothelial dysplasia: carcinoma in situ	14	546
Vesico-intestinal fistula	10	390
Vesico-uterine fistula caused by sectio cesarea	2	78
Vesico-vaginal fistula	6	234
Züdlerprosthesis penetrating bladder wall	3	117
Total	479	18,681
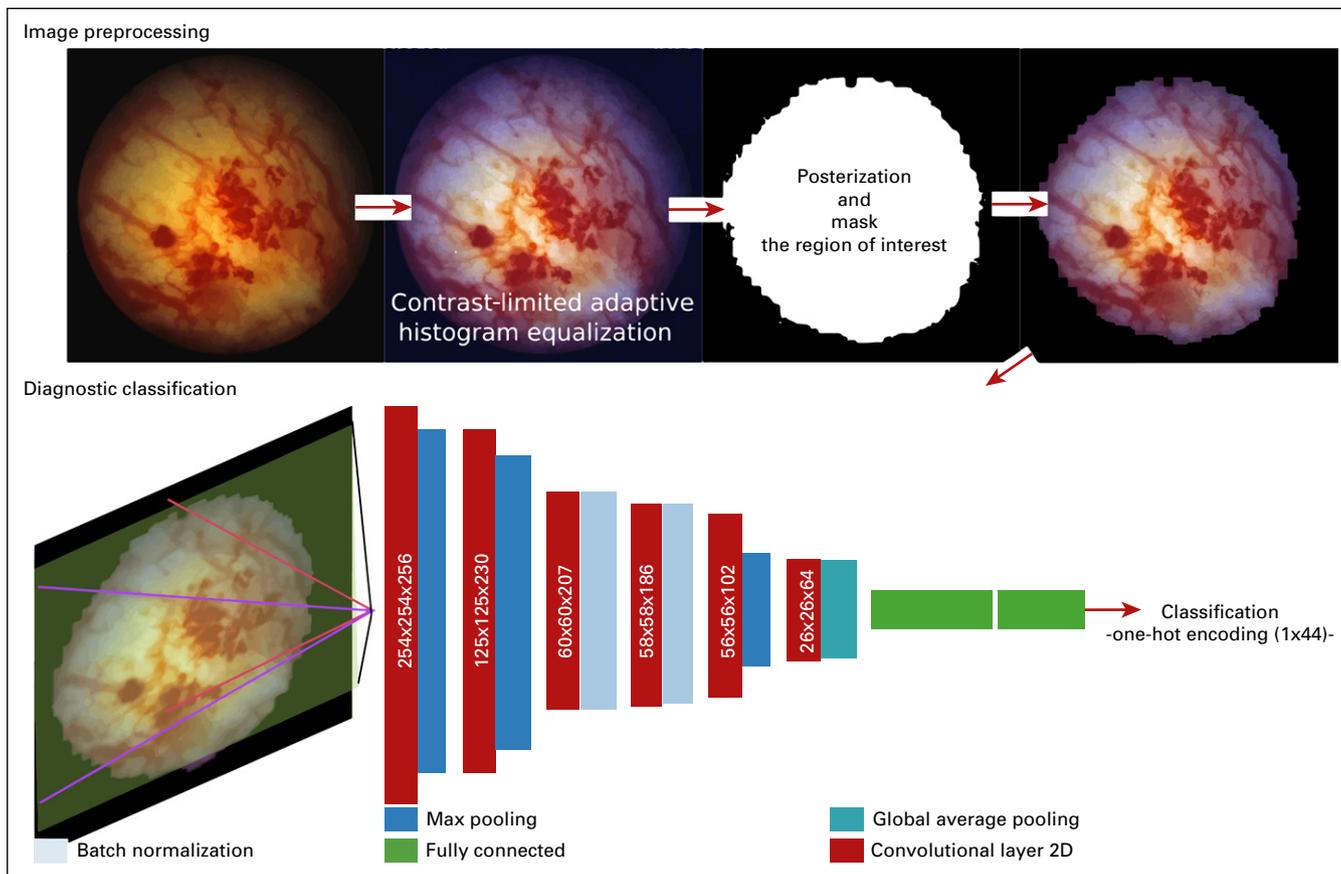
Fig 1. The algorithm for the image processing and the diagnostic classification of cystoscopic images. Here, we provide the architecture of the deep convolutional neural network with the harmonic series of the initial filter size (the dimensions of each convolutional layer are given inside the rectangle). 2D, two dimensional.

RESULTS

Table 2 lists the classification accuracies for the diagnostic classification of the cystoscopic images using the deep-learning models. We observed that the XCeption-based model achieved the highest F1 score (99.52%) followed by the ResNet-based model (99.48%) and the harmonic-series–based model (99.45%). The deep CNN architecture is in the Data Supplement. The classification accuracy of each diagnostic class for XCeption also is in the Data Supplement. When we focused on the images misclassified by the Xception-based model, 7.86% of images that showed bladder stones with indwelling catheter and 1.43% of images that showed bladder diverticulum were falsely classified; when we evaluated these misclassified images, we found that they showed stones adhered to the indwelling catheter. One image that showed a large median lobe of the prostate was falsely classified as a healthy prostate, and 1.43% of the images that showed bladder diverticulum were classified as urothelial cancer in bladder diverticulum. We found that the introduced models accurately identified malignant urologic findings (ie, bladder urothelial carcinoma). Also, these models could distinguish among interstitial cystitis, other types of cystitis, and carcinoma in situ.

Although the harmonic-series–based model has a simplified CNN architecture, its classification performance is comparable to those of ResNet and Xception. Figure 2 shows the activation maps from the harmonic-series–based model, which indicate the patterns associated with the corresponding findings.

DISCUSSION

This study shows that a deep CNN can be used for diagnostic classification of cystoscopic images. Currently, deep learning is state of the art for image recognition and segmentation because of its robust and accurate classification performance.[1-4] The application of deep learning for image segmentation of computer or magnetic resonance imaging has been reported for prostate, bladder, lung, breast, colon, and eye diseases.[13-20] However, the use of deep learning for the diagnostic classification of cystoscopic

Table 2. Classification Accuracies of Different Models for the Diagnostic Classification of Cystoscopic Images

Model	Classification Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Difference* (%)
XCeption	99.52	99.54	99.52	99.52	0
ResNet50	99.48	99.48	99.48	99.48	0.04
Harmonic series concept	99.45	99.45	99.45	99.45	0.07
90%-layer concept	99.11	99.11	99.11	99.11	0.41
InceptionV3	98.73	98.86	98.73	98.74	0.78
VGG-16	97.42	97.82	97.35	97.59	1.93
VGG-19	95.47	95.65	95.47	95.47	4.05

*Reference: XCeption.

images remains unclear. We found that models based on deep learning are capable of predicting cystoscopic findings from cystoscopic images with high accuracy. Moreover, these models can estimate the likelihood of different diagnoses from cystoscopic images and list the most likely results, which can be used to adjust diagnoses. We demonstrated that deep-learning models are capable of distinguishing carcinoma in situ from cystitis or interstitial cystitis using cystoscopic images; the diagnosis of carcinoma in situ is challenging and requires photodynamic diagnosis cystoscopy.[21] However, future work is necessary to compare the diagnostic performance of artificial intelligence (AI)–aided cystoscopy with photodynamic diagnosis cystoscopy.

Fig 2. Activation maps for different diagnostic classes as determined by the deep convolutional neural network model, in which filter size is defined by the harmonic series number.
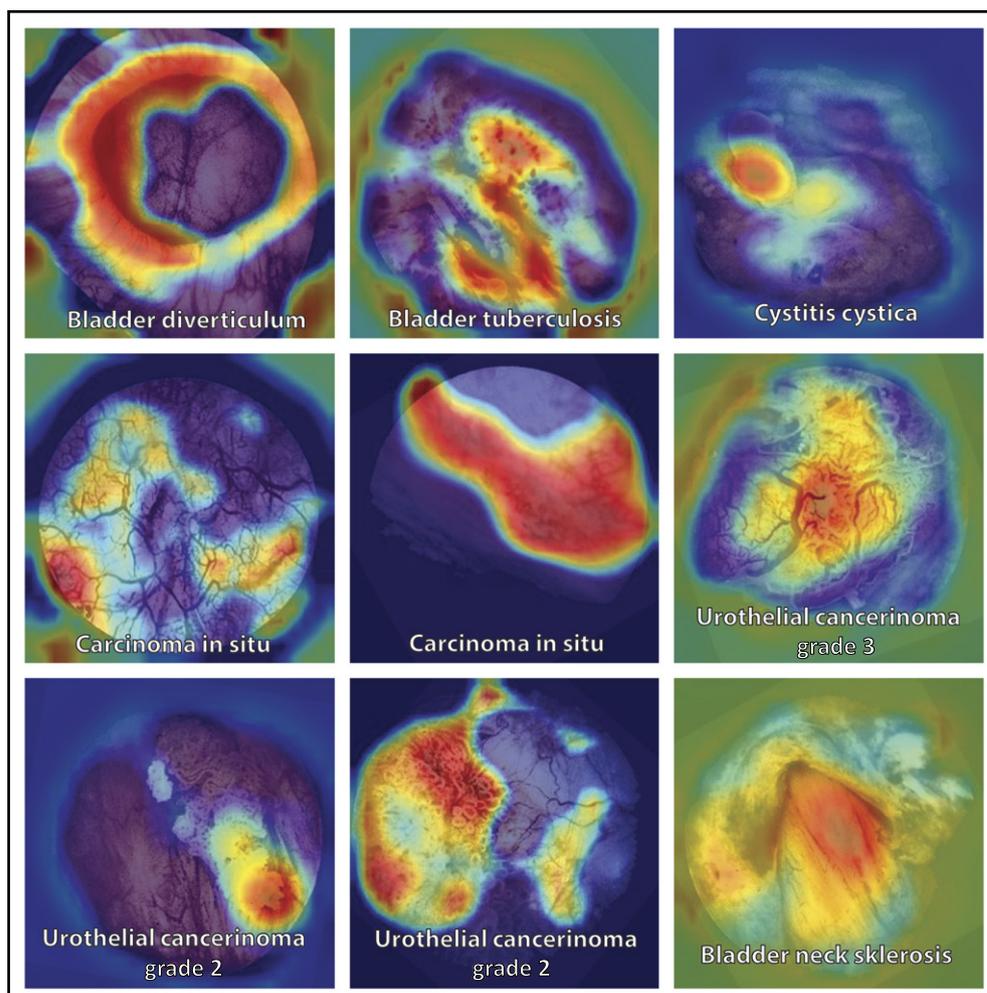
The high quality of the input data is essential to develop robust models for diagnostic classification. Therefore, we included ideal cystoscopic images that showed different findings from resources for medical education and residency training.

Image preprocessing or image augmentation is necessary to develop robust models that are less affected by noise in the cystoscopic images. We could achieve a rotation/shear-independent classification by applying image augmentation that mimics the conditions of the cystoscopic examination, in which the cystoscope can take different rotation positions. The contrast of the cystoscopic images is affected by the imbalanced distribution of the cystoscopic light caused by strong lighting at the center of the examination view. We could overcome the problem of the imbalanced illumination of the cystoscopic images by using contrast-limited adaptive histogram equalization.

We proposed a novel and simple CNN architecture for diagnostic classification that was based on harmonic series numbers that achieved a classification performance comparable to those of well-known CNN models applied for image classification. The trained diagnostic classification models can screen 64 images per second on average and thus can be fitted for video streaming from the cystoscopic camera, which does not require the interruption of the cystoscopic examination to capture cystoscopic images. As a consequence, future work will be directed to the implementation of one of these models in clinical routines to evaluate the clinical impact of our model.

In summary, this work proves the usability of deep learning to predict and classify cystoscopic findings with high accuracy. We included images of cystoscopic findings seen in clinical routine. The deep-learning model can be integrated into the AI-aided cystoscopic imaging diagnostic tool (AI cystoscopy) that supports urologists and gynecologists during the cystoscopic examination. AI cystoscopy will be helpful for training and medical education, during which resident physicians or medical students can learn to differentiate diagnoses using visual evaluation. Our introduced model may be used to differentiate benign lesions from malignant lesions to reduce the number of unnecessary biopsies. The deep-learning approach also facilitates

the weighing of features found in cystoscopic images, reduces the effects of noise and eliminates background noise (eg, the surrounding black frame, shadow effect).

This work has some limitations. First, we used still images and not cystoscopic videos, which may limit the diagnostic capabilities of the algorithm and possibly introduce error. Also, this study is limited by image quality, because the images came from patients who underwent cystoscopic examination before 1985, and image quality has improved since then. Nevertheless, this work is a proof-of-concept study that shows the application of deep convolutional networks to determine pathologic findings from cystoscopic images. Moreover, we did not have any notable changes in cystoscopic diagnosis since 1985, because cystoscopic findings have already been defined for more than two centuries.[22] Copyright concerns prevented us from using newer books and resources with cystoscopic images.

Another limitation is the low image quality captured by an analog camera; the quality of endoscopic equipment has dramatically improved since the images were captured because of the use of flexible fiberoptic endoscopes and high-definition digital cameras. However, image recognition could detect the findings in the cystoscopic images from 1985 with an F1 score of 0.99, which suggests that our approach can be applied to high-quality images. We did not consider the area under the receiver-operating characteristic curve, because there is a criticism of the application of the curve for the evaluation of the classification performance of machine-learning models.[23]

Although we have tried to consider all possible variations of the findings locations, we could not consider all possible angles for the cystoscopic findings. Our image collection includes all essential findings that can be seen during the clinical routine. However, this data set is limited by the sample size of certain subgroups, which may not represent a variety of findings. Another limitation is that our data sets from the cystoscopic atlas are more homogeneous than the data sets generated from consecutive patient cases. CNN models can identify cystoscopic findings more easily from rotated images than from new images. Therefore, the recognition results should be verified on additional data sets from real cases.

This study has proven the potential of CNN for the diagnostic classification of cystoscopic images. Although some subsets were falsely recognized, the neural network architecture can be improved to increase the accuracy performance for these subgroups.

Prospectively, we aim to verify this model on real clinical cases and compare its classification performance with that of human visual assessment. Future work also will focus on development of a model that facilitates the evaluation of the cystoscopic images in real time and its implementation in clinical routine.

Affiliations

Okyaz Eminaga, Stanford Medical School, Stanford, CA; University Hospital of Cologne, Cologne, France; **Nurettin Eminaga**, St Mauritius Therapy Clinic, Meerbusch; **Axel Semjonow**, University Hospital Muenster; and **Bernhard Breil**, Niederrhein University of Applied Sciences, Krefeld, Germany.

REFERENCES

1. Araújo T, Aresta G, Castro E, et al: Classification of breast cancer histology images using convolutional neural networks. PLoS One 12:e0177544, 2017

2. Poplin R, Newburger D, Dijamco J, et al: Creating a universal SNP and small indel variant caller with deep neural networks. bioRxiv 092890. https://doi.org/10.1101/092890

3. Huang Y, Zheng H, Liu C, et al: Epithelium-stroma classification via convolutional neural networks and unsupervised domain adaptation in histopathological images. IEEE J Biomed Health Inform 21:1625-1632, 2017

4. He K, Gkioxari G, Dollar P, et al: Mask R-CNN. https://arxiv.org/abs/1703.06870

5. Schönebeck J: Atlas of Cystoscopy. New York, NY, Grune and Stratton, 1985

6. He K, Zhang X, Ren S, et al: Deep residual learning for image recognition: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, June 26-July 1, 2016, pp 770-778

7. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. https://arxiv.org/abs/1409.1556

8. Szegedy C, Vanhoucke V, Ioffe S, et al: Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, June 26-July 1, 2016, pp 2818-2826

9. Chollet F: Xception: Deep learning with depthwise separable convolutions. ArXiv. Presented at the 3rd International Conference for Learning Representations, San Diego, CA, May 7-9, 2015

10. Klambauer G, Unterthiner T, Mayr A, et al: Self-normalizing neural networks. Adv Neural Info Process Syst. https://arxiv.org/abs/1706.02515

11. Srivastava N, Hinton G, Krizhevsky A, et al: Dropout: A simple way to prevent neural networks from overfitting. J Machine Learning Res 15:1929-1958, 2014

12. Kingma DP, Ba J: Adam: A method for stochastic optimization. Presented at the 3rd International Conference for Learning Representations, San Diego, CA, May 7-9, 2015. https://arxiv.org/abs/1412.6980

13. Takahashi H, Tampo H, Arai Y, et al: Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy. PLoS One 12:e0179790, 2017

14. Gulshan V, Peng L, Coram M, et al: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316:2402-2410, 2016

15. Antropova N, Huynh BQ, Giger ML: A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med Phys 44:5162-5171, 2017

16. Rios Velazquez E, Parmar C, Liu Y, et al: Somatic mutations drive distinct imaging phenotypes in lung cancer. Cancer Res 77:3922-3930, 2017

17. Cheng R, Roth HR, Lay N, et al: Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks. J Med Imaging (Bellingham) 4:041302, 2017

18. Cha KH, Hadjiiski L, Samala RK, et al: Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. Med Phys 43:1882-1896, 2016

19. Liu J, Wang D, Lu L, et al: Detection and diagnosis of colitis on computed tomography using deep convolutional neural networks. Med Phys 44:4630-4642, 2017

20. Cha KH, Hadjiiski L, Chan HP, et al: Bladder cancer treatment response assessment in CT using radiomics with deep-learning. Sci Rep 7:8738, 2017

21. Lapini A, Minervini A, Masala A, et al: A comparison of hexaminolevulinate (Hexvix) fluorescence cystoscopy and white-light cystoscopy for detection of bladder cancer: Results of the HeRo observational study. Surg Endosc 26:3634-3641, 2012

22. Samplaski MK, Jones JS: Two centuries of cystoscopy: The development of imaging, instrumentation and synergistic technologies. BJU Int 103:154-158, 2009

23. Hand DJ: Measuring classifier performance: A coherent alternative to the area under the ROC curve. Mach Learn 77:103-123, 2009