

Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images

YASUNARI MIYAGI¹⁻³, KAZUHIRO TAKEHARA⁴ and TAKAHITO MIYAKE⁵

¹Medical Data Labo, Okayama 703-8267; ²Department of Gynecologic Oncology, Saitama Medical University International Medical Center, Hidaka, Saitama 350-1298; ³Department of Gynecology, Miyake Ofuku Clinic, Okayama 701-0204; ⁴Department of Gynecologic Oncology, National Hospital Organization, Shikoku Cancer Center, Matsuyama, Ehime 791-0208; ⁵Department of Obstetrics and Gynecology, Miyake Clinic, Okayama 701-0204, Japan

Received April 26, 2019; Accepted September 9, 2019

DOI: 10.3892/mco.2019.1932

Abstract. The aim of the present study was to explore the feasibility of using deep learning as artificial intelligence (AI) to classify cervical squamous epithelial lesions (SIL) from colposcopy images. A total of 330 patients who underwent colposcopy and biopsy by gynecologic oncologists were enrolled in the current study. A total of 97 patients received a pathological diagnosis of low-grade SIL (LSIL) and 213 of high-grade SIL (HSIL). An original AI-classifier with 11 layers of the convolutional neural network was developed and trained. The accuracy, sensitivity, specificity and Youden's J index of the AI-classifier and oncologists for diagnosing HSIL were 0.823 and 0.797, 0.800 and 0.831, 0.882 and 0.773, and 0.682 and 0.604, respectively. The area under the receiver-operating characteristic curve was 0.826 ± 0.052 (mean \pm standard error), and the 95% confidence interval 0.721-0.928. The optimal cut-off point was 0.692. Cohen's Kappa coefficient for AI and colposcopy was 0.437 ($P < 0.0005$). The AI-classifier performed better than oncologists, although not significantly. Although further study is required, the clinical use of AI for the classification of HSIL/LSIL from by colposcopy may be feasible.

Introduction

With current advancements in computer science, artificial intelligence (AI) has made remarkable progress recently. The hypothetical moment in time when AI becomes so advanced that humanity undergoes a dramatic and irreversible change (1) is likely to arrive in this century. AI has already exceeded human experts in the field of games with perfect information, such as Go (2), showing us novel tactics. Therefore, since AI can recognize some information that conventional procedures

cannot detect, it may provide a more precise diagnosis in practical medicine. AI may also be able to support clinicians in practical medicine, reducing the time and effort necessary. The aim of the present study was to investigate the feasibility of applying deep learning, a type of AI, for gynecological clinical practice.

Uterine cervical cancer continues to be a major public health problem. Cervical cancer is the third most common female cancer and the leading cause of cancer-related mortality among women in Eastern, Western and Middle Africa, Central America, South-Central Asia and Melanesia. New methodologies of cervical cancer prevention should be made available and accessible to women of all countries (3).

Colposcopy is a well-established tool for examining the cervix under magnification (4-6). When lesions are treated with acetic acid diluted to 3-5%, colposcopy can detect and recognize cervical intraepithelial neoplasia (CIN). Classification systems, such as the Bethesda System in 2002 are used to categorize lesions as high-grade squamous intraepithelial lesions (HSIL) or low-grade SIL (LSIL) (7,8). HSIL and LSIL were previously referred to as CIN2/CIN3 and CIN1, respectively. In clinical practice, it is important for clinicians to distinguish HSIL from LSIL in biopsy specimens, since further examination or treatment, such as conization, may be required for HSIL. Expert gynecologic oncologists spend much time and effort to provide precise colposcopy findings.

For these reasons, we explored whether AI can evaluate colposcopy findings as well as a gynecologic oncologist. In the present study, we applied deep learning with a convolutional neural network to the realm of AI to develop an original classifier for predicting HSIL or LSIL from colposcopy images. Deep learning is becoming very popular among machine learning methods, such as logistic regression (9), naive Bayes (10), nearest neighbor (11), random forest (12) and neural network (13). The classifier program was developed using supervised deep learning with a convolutional neural network (14) that tried to mimic the visual cortex of the mammal brain (15-23), in order to categorize colposcopy images as either HSIL or LSIL. The present study demonstrated the effective use of the AI colposcopy image classifier in predicting HSIL or LSIL by comparing the colposcopic diagnosis to that of gynecologic oncologists.

Correspondence to: Dr Yasunari Miyagi, Medical Data Labo, 289-48 Yamasaki, Naka Ward, Okayama 703-8267, Japan
E-mail: ymiyagi@mac.com

Key words: colposcopy, cervical cancer, cervical intraepithelial neoplasia, deep learning, artificial intelligence

Patients and methods

Patients. This retrospective study used fully deidentified patient data and was approved by the Institutional Review Board of Shikoku Cancer Center (approval no. 2017-81). This study was explained to the patients, who underwent cervical biopsy by gynecologic oncologists at Shikoku Cancer Center from January 1, 2012 to December 31, 2017. Patients were also directed to a website with additional information, including an opt-out option for the study. The Institutional Review Board of Shikoku Cancer Center approved the opt-out option for patients to choose to withdraw from this study. Gynecologic oncologists at the Shikoku Cancer Center determined the biopsy in routine conventional practice when necessary. A total of 330 patients were enrolled in this study.

Images. Colposcopy images of lesions processed with acetic acid prior to biopsy were captured, cropped to a square and saved in JPEG format. The images were used retrospectively as the input data for deep learning. Gynecologists biopsied the most advanced lesion, the pathological results of which were revealed later.

Preparation for AI. All deidentified images stored offline were transferred to our AI-based system. Each image was cropped to a square and then saved. Twenty percent of the images were randomly selected as the test dataset, and the rest were used as the training dataset. Next, 20% of the training dataset was used as the validation dataset, and the rest was used to train the AI-classifier. Thus, the training, validation and test datasets did not overlap. That way, the AI classifier was trained by a training dataset and simultaneously validated and then tested for the test dataset. The number of training datasets was augmented, as is often done in computer science, in a process known as data augmentation. The training dataset was augmented in this study because the image processing of the arbitrary degrees of rotation can lead to images being included in the same category of different vector data.

AI-classifier. We developed classifier programs using supervised deep learning with a convolutional neural network (14,19). We tested a lot of convolutional neural networks by varying L2 regularization (24,25) and the architectures consisted of a combination of convolution layers with kernels (26-28), pooling layers (29-32), flattened layers (33), linear layers (34,35), rectified linear unit layers (36,37) and a softmax layer (38,39) that demonstrated the probability of LSIL or HSIL from an image (Table I). We also tested the ResNet-50 network (40), which performed very well in the ImageNet Large Scale Visual Recognition Challenge (41) and Microsoft common objects in context MS-COCO (42) competition. We modified the ResNet-50, the first layer of which was replaced with the convolutional layer with a kernel size of 4, stride size of 2, padding size of 2, and input image size of 111x111 pixels, which is the minimum size for the ResNet-50. The last layer of the ResNet-50 was also replaced with a linear layer, followed by a softmax layer with an output with a vector size of 2.

Cross-validation (43-45), a powerful method for model selection, was applied to identify the optimal method of

machine learning. The suitable number of images for the training data was investigated by evaluating the accuracy and variances using the 5-fold cross-validation method. This calculation procedure reveals the optimal number of training data and can be used to avoid overfitting (46-51), which is a modeling error that occurs when a classifier is too closely fit to a limited set of data points. After the optimal number of training data was obtained, the classifier that showed the best accuracy was selected, as is standard practice in computer science. Conventional colposcopy diagnosis and AI colposcopy diagnosis for test dataset were compared.

Development environment. The following development environment was used in the present study: A Mac running OS X 10.14.3 (Apple, Inc.) and Mathematica 11.3.0.0 (Wolfram Research).

Statistical analysis. The laboratory and AI-classifier data were compared. The two proportion between gynecologic oncologists and the classifier using deep learning was compared by two-proportion z-test. The agreements among the conventional colposcopy, the AI classifier and pathological results were evaluated by Cohen's Kappa (52) coefficients. The formula to calculate Cohen's kappa for two raters is as follows:

$$(A_{\text{observed}} - A_{\text{expected by chance}}) / (1 - A_{\text{expected by chance}})$$

where:

A_{observed} = the relative observed agreement among raters,

$A_{\text{expected by chance}}$ = the hypothetical probability of chance agreement. Mathematica 11.3.0.0 (Wolfram Research) was used for all statistical analyses.

Results

Profiles of pathological diagnosis and colposcopy. The pathological diagnoses and corresponding number of patients were as follows: LSIL, 97; HSIL, 213; squamous cell carcinoma, 12; microinvasive squamous cell carcinoma, 1; adenocarcinoma, 5; adenocarcinoma in situ, 2. A total of 310 images of both pathological LSIL and HSIL were used, due of the limited number of available images. Among the 213 pathological HSIL cases, 177, 32, 3 and 1 received a conventional colposcopy diagnosis by gynecologists of HSIL, LSIL, invasive cancer and cervicitis, respectively. Among the 97 pathological LSIL cases, 22, 70 and 5 received a conventional colposcopy diagnosis by gynecologists of HSIL, LSIL and cervicitis, respectively (Table II). The accuracy, sensitivity, specificity, positive predictive value, negative predictive value and Youden's J index (53) for HSIL, as determined by gynecologists were 0.797 (247/310), 0.831 (177/213), 0.773 (75/97), 0.889 (117/199), 0.686 (70/102) and 0.604, respectively (Table III).

AI-classifier results. The best accuracy for HSIL was 0.823 (51/62), when the number of the augmented training data set was 1,488, the value of L2 regularization 0.175, the number of layers of the architecture 11 (Table II) and the image size 70x70 pixels. The sensitivity, specificity, positive predictive

Table I. Architectures of the top classifier that exhibited the highest accuracy.

Layers	Supplementations
Convolution layer	Output channels; 64, Kernel size; 3x3
ReLU	N/A
Pooling layer	Kernel size; 2x2
Convolution layer	Output channels; 64, Kernel size; 3x3
ReLU	N/A
Pooling layer	Kernel size; 2x2
Flatten layer	N/A
Linear layer	Size; 2 ⁹
ReLU	N/A
Linear layer	2
Softmax layer	N/A

The convolutional neural network structures, which consisted of 11 layers of convolutional deep learning, were obtained. ReLU, rectified linear units.

value, negative predictive value and Youden's J index were 0.800 (36/45), 0.882 (15/17), 0.947 (36/38), 0.625 (15/24) and 0.682, respectively (Table III). The accuracy, sensitivity, specificity, positive predictive value, negative predictive value and Youden's J index of the best modified ResNet-50 were 0.790 (49/62), 0.847 (39/46), 0.625 (10/16), 0.867 (39/45), 0.588 (10/17) and 0.472, respectively. The original conventional neural network was better than the modified ResNet-50, although not significantly. There were no differences between the gynecologic oncologists and the best AI in accuracy, sensitivity, specificity, positive or negative predictive value, as determined by a proportional test.

Using confidence score, the area under the receiver-operating characteristic curve (ROC) of the best classifier for predicting HSIL was 0.824±0.052 (mean ± SE), and the 95% confidence interval 0.721-0.928. The ROC curve is shown in Fig. 1. The optimal cut-off point was 0.692.

Comparison of AI-classifier with conventional colposcopy. The association among conventional colposcopy, the AI classifier and pathological results for the test dataset that was 20% of patients of both pathological HSIL and LSIL diagnosed by punch biopsy are shown in Tables IV-VI. Cohen's Kappa (52) coefficients of the conventional colposcopy and pathological results, the AI classifier and pathological results, and the conventional colposcopy and the AI classifier were 0.691, 0.561 and 0.437 (all P<0.0001), respectively. All were more than moderate agreements (54). Conventional colposcopy showed better agreement than the AI, although the difference was not significant. Classification took less than 0.2 sec per image.

Discussion

We developed a classifier using deep learning with convolutional neural networks using images of cervical SILs to predict the pathological diagnosis. In the present study, the

accuracy values achieved by the classifier and by gynecologic oncologists were 0.823 and 0.797, respectively (Table III). The sensitivity values of the classifier and gynecologic oncologists were 0.800 and 0.831, respectively. The specificity values of the classifier and gynecologic oncologists were 0.882 and 0.773, respectively. The accuracy and specificity of the classifier were superior to those of gynecologic oncologists, although the difference was not significant. Only moderate agreement was obtained between conventional colposcopy diagnosis and AI colposcopy diagnosis with the Kappa value of 0.437. McHugh reported that Cohen suggested 0.41 might be acceptable and the Kappa result be interpreted as follows: Values ≤0 as indicating no agreement and 0.01-0.20 as none to slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial and 0.81-1.00 (54). Thus, the Kappa value of 0.437 was acceptable. But, AI colposcopy that might have a potential would not be an alternative to conventional colposcopy without further studies.

Several reports have used AI (55) for deep learning with convolutional neural networks in medicine (56). The accuracy values of this method with deep learning have been published and include 0.997 for the histopathological diagnosis of breast cancer (57), 0.90-0.83 for the early diagnosis of Alzheimer's disease (58), 0.83 for urological dysfunctions (59), 0.72 (60) and 0.50 (61) for colposcopy, 0.68-0.70 for localization of rectal cancer (62), 0.83 for the diagnostic imaging of orthopedic trauma (63), 0.98 for the morphological quality of blastocysts and evaluation by an embryologist (64), 0.65 for predicting live birth without aneuploidy from a blastocyst image (65) and 0.64-0.88 for predicting live birth from a blastocyst image of patients by age (66).

Several studies have reported a limitation of conventional colposcopy. An investigation of the accuracy of colposcopically-directed biopsy reported a total biopsy failure rate, comprising both non-biopsy and incorrect selection of biopsy site, of 0.20 in CIN1, 0.11 in CIN2 and 0.09 in CIN3 (67). The colposcopic impression of high-grade CIN had a sensitivity of 0.54 and a specificity of 0.88, as determined by 9 expert colposcopists in 100 cervigrams (68). The sensitivity of an online colposcopic assessment of high-grade disease (CIN2 and CIN3) by 20 colposcopists was 0.39 (69). Thus, conventional colposcopy does not provide good sensitivity, even by colposcopists. By contrast, the accuracy and sensitivity reported in this study for predicting HSIL from colposcopy images using deep learning was 0.823 and 0.800, respectively, which appears favorable. Since the classifier was not trained in colposcopy findings such as acetowhite epithelium, mosaic, punctuation, it may recognize some features of cervical SILs by itself in high-dimensional space. It is possible that the AI-classifier may recognize features that colposcopists do not, such as complexity of the shape of the lesion, relative or absolute brightness of acetowhite, distribution of punctuation density, and quantitative marginal evaluation of borders. The pathological results were obtained and defined by punch biopsy in this study, as it was not ethically recommended for patients with LSIL (CIN1) diagnosed by colposcopy to undergo conization or hysterectomy. If the pathological results were defined by conization or hysterectomy, the advanced lesion would have been revealed, and thus both conventional colposcopy and the AI classifier may have demonstrated different results. In the present study, we only tried to compare the effectiveness

Table II. Characteristics of the 330 patients that underwent colposcopy and biopsy by gynecologic oncologists.

Patient characteristics	Pathological HSIL (n=213)	Pathological LSIL (n=97)
Age (years)		
Mean \pm SD	31.66 \pm 5.01	33.75 \pm 8.94
Median	32	33
Range	19-46	19-62
HPV		
Type 16 positive	75	2
Type 18 positive	5	2
Type 16 and 18 positive	1	0
Positive, but not type 16 or 18	123	33
Negative	6	6
Not available	3	54
Colposcopic diagnosis		
HSIL	177	22
LSIL	32	70
Cervicitis	1	5
Invasive cancer	3	0

HSIL, high-grade squamous intraepithelial lesions; LSIL, low-grade squamous intraepithelial lesions; SD, standard deviation.

Table III. Comparison between gynecologic oncologists and the top classifier using deep learning.

Variable	Gynecologic oncologists	AI
Accuracy	0.797 (247/310)	0.823 (51/62)
Sensitivity	0.831 (177/213)	0.800 (36/45)
Specificity	0.773 (75/97)	0.882 (15/17)
Positive predictive value	0.889 (177/199)	0.947 (36/38)
Negative predictive value	0.686 (70/102)	0.625 (15/24)
Youden's J index	0.604	0.682

Bracketed data indicates the number of corresponding selected cases/the number of relevant cases. AI, artificial intelligence.

of AI with that of conventional colposcopy for SIL. When the AI will be used to predict more advanced diseases, such as squamous cell carcinoma, adenocarcinoma and AIS, the pathological diagnosis should be provided by not punch biopsy but conization or hysterectomy.

In clinical practice, it is important for clinicians to distinguish HSIL from LSIL in biopsy specimens. Further examination or treatment, such as conization, may be required for HSIL. When a reliable classifier indicates HSIL from colposcopy images in clinical practice, the clinician should consider biopsy. The accuracy values of the classifier and gynecologists for detecting HSIL were 0.823 and 0.797, respectively. The classifier might help untrained clinicians to avoid or reduce the risks of overlooking HSIL. When the AI-classifier can perform higher in terms of accuracy, sensitivity and specificity for classifying HSIL/LSIL, clinicians

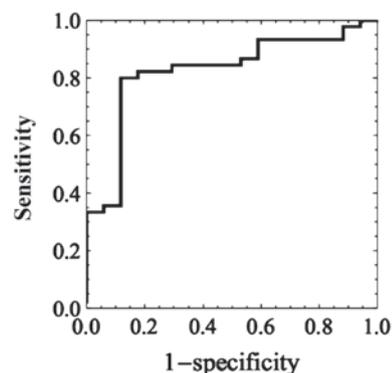


Figure 1. The receiver-operating characteristic curve of the best classifier for predicting high-grade squamous intraepithelial lesions. The value of the area under the curve is 0.824 ± 0.052 (mean \pm standard error) and the 95% confidence interval ranged between 0.721-0.928.

will be able to perform more precise practice, referencing AI. Furthermore, a gynecologist could reduce the time and effort it takes to become a colposcopy expert and, as a result, improve in other skills, training and activities.

The architecture of the neural network has progressed. The LeNet study published in 1998 (70) consisted of 5 layers. AlexNet, published in 2012 (38), consisted of 14 and Google Net, published in 2014 (35), was constructed from a combination of micro networks. ResNet-50, published in 2015 (41), consisted of modules with a shortcut process. The Squeeze-and-Excitation Networks were published in 2017 (71). AI used for image recognition is still being developed. Progress in AI will allow us to achieve better results. Image information is one of the parameters that need to be investigated. Only 15x15 pixels are used to detect cervical cancer (72). In a colposcopy study (61), it was reported that the accuracy for images of 150x150 pixels was better than

Table IV. Conventional colposcopy diagnosis and pathological results of the test data set.

Lesion type	Conventional colposcopy diagnosis	
	HSIL	LSIL
Pathological HSIL	39	6
Pathological LSIL	2	15

Cohen's Kappa coefficient was 0.691, $P < 0.0001$. HSIL, high-grade squamous intraepithelial lesions; LSIL, low-grade squamous intraepithelial lesions; AI, artificial intelligence.

Table V. AI colposcopy diagnosis and pathological result for test data set.

Lesion type	AI colposcopy diagnosis	
	HSIL	LSIL
Pathological HSIL	36	9
Pathological LSIL	3	14

Cohen's Kappa coefficient was 0.561, $P < 0.0001$. HSIL, high-grade squamous intraepithelial lesions; LSIL, low-grade squamous intraepithelial lesions; AI, artificial intelligence.

Table VI. Conventional colposcopy diagnosis and AI colposcopy diagnosis for test data set.

	AI colposcopy diagnosis	
	HSIL	LSIL
Conventional colposcopy HSIL	32	9
Conventional colposcopy LSIL	7	14

Cohen's Kappa coefficient was 0.437, $P < 0.0005$. HSIL, high-grade squamous intraepithelial lesions; LSIL, low-grade squamous intraepithelial lesions; AI, artificial intelligence.

those for 32x32 or 300x300 pixels. Hence, image size remains an issue. We used 70x70 and 111x111 pixels for our images, in order to use the original neural networks and the modified ResNet-50, respectively. The original conventional neural network was better than the modified ResNet-50, although not significantly. We believe that a pixel size of 70x70 falls within the acceptable range. Regularization values are also important parameters for constructing a good classifier that avoids overfitting. If the regularization value is too low, overfitting occurs. If the value is too large, the classifier will not be trained well. Choosing the appropriate number of training datasets is also very important. If the number of training datasets is too high, the accuracy will be lower

and more variances will be observed. The validation dataset, as well as L2 regularization, also prevent overfitting. The appropriate number of training datasets must be achieved to obtain a good classifier. More varied patterns of images may be needed for datasets. Ordinarily, 500-1,000 images are prepared for each class during image classification with deep learning (61,73). Such a large number of datasets will improve the accuracy and specificity of the classifier with deep learning.

In the present study, a classifier was developed based on deep learning, which used images of uterine cervical SILs to predict pathological HSIL/LSIL. Its accuracy was 0.823. Although further study may be required to validate the classifier, we demonstrated that AI may have a clinical use in colposcopic examinations and may provide benefits to both patients and medical personnel.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets generated and/or analyzed during the present study are not publicly available, since data sharing is not approved by the Institutional Review Board of Shikoku Cancer Center (approval no. 2017-81).

Authors' contributions

YM designed the current study, performed AI programming, produced classifiers by AI, performed statistical analysis and wrote the manuscript. KT performed clinical intervention, data entry and collection, designed the current study, and critically revised the manuscript. TM designed the current study and critically revised the manuscript.

Ethics approval and consent to participate

The protocol for the present retrospective study used fully deidentified patient data and was approved by the Institutional Review Board of Shikoku Cancer Center (approval no. 2017-81). The study protocol was explained to the patients who underwent cervical biopsy at the Shikoku Cancer Center from January 1, 2012 to December 31, 2017. Patients were also directed to a website with additional information, including an opt-out option, allowing them to not participate. Written informed consent for was not required, according to the guidance of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Patient consent for publication

The current study was explained to the patients who underwent cervical biopsy at the Shikoku Cancer Center from January 1, 2012 to December 31, 2017. Patients were also directed to a website with additional information, including an

opt-out option that let them know they had the right to refuse publication.

Competing interests

YM and TM declare that they have no competing interests. KT reports personal fees from Taiho Pharmaceuticals, Chugai Pharma, AstraZeneca, Nippon Kayaku, Eisai, Ono Pharmaceutical, Terumo Corporation and Daiichi Sankyo, outside of the submitted work.

References

- Müller VC and Bostrom N: Future progress in artificial intelligence: A survey of expert opinion. In: *Fundamental Issues of Artificial Intelligence*. Springer, Berlin, pp555-572, 2016.
- Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, *et al*: Mastering the game of Go without human knowledge. *Nature* 550: 354-359, 2017.
- Arbyn M, Castellsagué X, de Sanjosé S, Bruni L, Saraiya M, Bray F and Ferlay J: Worldwide burden of cervical cancer in 2008. *Ann Oncol* 22: 2675-2686, 2011.
- Garcia-Arteaga JD, Kybic J and Li W: Automatic colposcopy video tissue classification using higher order entropy-based image registration. *Comput Biol Med* 41: 960-970, 2011.
- Kyrgiou M, Tsoumpou I, Vrekoussis T, Martin-Hirsch P, Arbyn M, Prendiville W, Mitrou S, Koliopoulos G, Dalkalitis N, Stamatopoulos P and Paraskevaides E: The up-to-date evidence on colposcopy practice and treatment of cervical intraepithelial neoplasia: The Cochrane colposcopy and cervical cytopathology collaborative group (C5 group) approach. *Cancer Treat Rev* 32: 516-523, 2006.
- O'Neill E, Reeves MF and Creinin MD: Baseline colposcopic findings in women entering studies on female vaginal products. *Contraception* 78: 162-166, 2008.
- Waxman AG, Chelmow D, Darragh TM, Lawson H and Moscicki AB: Revised terminology for cervical histopathology and its implications for management of high-grade squamous intraepithelial lesions of the cervix. *Obstet Gynecol* 120: 1465-1471, 2012.
- Darragh TM, Colgan TJ, Cox JT, Heller DS, Henry MR, Luff RD, McCalmont T, Nayar R, Palefsky JM, Stoler MH, *et al*: The lower anogenital squamous terminology standardization project for HPV-associated lesions: Background and consensus recommendations from the college of American pathologists and the American society for colposcopy and cervical pathology. *J Low Genit Tract Dis* 16: 205-242, 2012.
- Dreiseitl S and Ohno-Machado L: Logistic regression and artificial neural network classification models: A methodology review. *J Biomed Inform* 35: 352-359, 2002.
- Ben-Bassat M, Klove KL and Weil MH: Sensitivity analysis in Bayesian classification models: Multiplicative deviations. *IEEE Trans Pattern Anal Mach Intell* 2: 261-266, 1980.
- Friedman JH, Baskett F and Shustek LJ: An algorithm for finding nearest neighbors. *IEEE Trans Comput* 24: 1000-1006, 1975.
- Breiman L: Random forests. *Mach Learn* 45: 5-32, 2001.
- Rumelhart D, Hinton G and Williams R: Learning representations by back-propagating errors. *Nature* 323: 533-536, 1986.
- Bengio Y, Courville A and Vincent P: Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35: 1798-828, 2013.
- Fukushima K: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36: 193-202, 1980.
- Hubel DH and Wiesel TN: Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195: 215-243, 1968.
- Hubel DH and Wiesel TN: Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148: 574-591, 1959.
- Schmidhuber J: Deep learning in neural networks: An overview. *Neural Netw* 61: 85-117, 2015.
- LeCun Y, Bottou L, Orr GB and Müller KR: Efficient BackProp. In: *Neural Networks: Tricks of the Trade*. Springer, Berlin, 1998.
- LeCun Y, Bottou L, Bengio Y and Haffner P: Gradient-based learning applied to document recognition. *Proc IEEE* 86: 2278-2324, 1998.
- LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W and Jackel LD: Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1: 541-551, 1989.
- Serre T, Wolf L, Bileschi S, Riesenhuber M and Poggio T: Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell* 29: 411-426, 2007.
- Wiatowski T and Bölcskei H: A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Trans Inf Theory* 64: 1845-1866, 2018.
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R: Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15: 1929-1958, 2014.
- Nowlan SJ and Hinton GE: Simplifying neural networks by soft weight-sharing. *Neural Comput* 4: 473-493, 1992.
- Bengio Y: Learning deep architectures for AI. *Found Trends Mach Learn* 2: 1-127, 2009.
- Mutch J and Lowe DG: Object class recognition and localization using sparse features with limited receptive fields. *Int J Comput Vision* 80: 45-57, 2008.
- Neal RM: Connectionist learning of belief networks. *Artificial Intell* 56: 71-113, 1992.
- Ciresan D, Meier U, Masci J, Gambardella LM and Schmidhuber J: Flexible, high performance convolutional neural networks for image classification. *IJCAI Proc Int Joint Conf Artificial Intell* 22: 1237-1242, 2011.
- Scherer D, Müller A and Behnke S: Evaluation of pooling operations in convolutional architectures for object recognition. In: *Artificial Neural Networks (ICANN) 2010. Lecture Notes in Computer Science*. Diamantaras K, Duch W, Iliadis LS (eds). Springer, Berlin, pp92-101, 2010.
- Huang FJ and LeCun Y: Large-scale learning with SVM and convolutional for generic object categorization. *Computer vision and pattern recognition. IEEE Comput Soc Conf* 1: 284-291, 2006.
- Jarrett K, Kavukcuoglu K, Ranzato MA and LeCun Y: What is the best multi-stage architecture for object recognition? In: *2009 IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan*, pp2146-2153, 2009.
- Zheng Y, Liu Q, Chen E, Ge Y and Zhao JL: Time series classification using multi-channels deep convolutional neural networks. Li F, Li G, Hwang S, Yao B, Zhang Z (eds). *Web-Age Information Management, WAIM 2014*. Springer, Cham, pp298-310, 2014.
- Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, *et al*: Human-level control through deep reinforcement learning. *Nature* 518: 529-533, 2015.
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, *et al*: Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1-9, 2015.
- Glorot X, Bordes A and Bengio Y: Deep sparse rectifier neural networks. *Proc Fourteenth Int Conf Artificial Intell Stat*: 315-323, 2011.
- Nair V and Hinton G: Rectified linear units improve restricted Boltzmann machines. *Proc Int Conf Mach Learn*: 807-814, 2010.
- Krizhevsky A, Sutskever I and Hinton GE: Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Proc Syst*: 1097-1105, 2012.
- Bridle JS: Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: *Neurocomputing*. Soulié FF, Héroult J (eds). Springer, Berlin, 1990.
- He K, Zhang X, Ren S and Sun J: Deep residual learning for image recognition. *arXiv*: 1512.03385, 2015.
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M and Berg AC: Imagenet large scale visual recognition challenge. *Int J Comput Vision* 115: 211-252, 2015.
- Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár and C. Lawrence Zitnick: 'Microsoft coco: Common objects in context.' In *European conference on computer vision*, pp. 740-755. Springer, Cham, 2014.
- Kohavi R: A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proc Int Joint Conf Artificial Intell* 2: 1137-1143, 1995.
- Schaffer C: Selecting a classification method by cross-validation. *Mach Learn* 13: 135-143, 1993.
- Rezaei-zadeh P, Tang L and Liu H: Cross-validation. In: *Encyclopedia of Database Systems*. Liu L, Özsu MT (eds). Springer, New York, 2009.

46. Yu L, Chen H, Dou Q, Qin J and Heng PA: Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans Med Imaging* 36: 994-1004, 2017.
47. Caruana R, Lawrence S and Giles CL: Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Adv Neural Inf Proc Syst*: 402-408, 2001.
48. Baum EB and Haussler D: What size net gives valid generalization? *Neural Comput* 1: 151-160, 1989.
49. Geman S, Bienenstock E and Doursat R: Neural networks and the bias/variance dilemma. *Neural Comput* 4: 1-58, 1992.
50. Krogh A and Hertz JA: A simple weight decay can improve generalization. *Adv Neural Inf Proc Syst* 4: 950-957, 1992.
51. Moody JE: The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. *Adv Neural Inf Proc Syst* 4: 847-854, 1992.
52. Cohen J: A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20: 37-46, 1960.
53. Youden WJ: Index for rating diagnostic tests. *Cancer* 3: 32-35, 1950.
54. McHugh ML: Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 22: 276-282, 2012.
55. Miyagi Y, Fujiwara K, Oda T, Miyake T and Coleman RL: Development of new method for the prediction of clinical trial results using compressive sensing of artificial intelligence. *J Biostat Biometric* 3: 202, 2018.
56. Abbod MF, Catto JW, Linkens DA and Hamdy FC: Application of artificial intelligence to the management of urological cancer. *J Urol* 178: 1150-1156, 2007.
57. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, Hulsbergen-van de Kaa C, Bult P, van Ginneken B and van der Laak J: Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 6: 26286, 2016.
58. Ortiz A, Munilla J, Górriz JM and Ramírez J: Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. *Int J Neural Syst* 26: 1650025, 2016.
59. Gil D, Johnsson M, Chamizo JMG, Paya AS and Fernandez DR: Application of artificial neural networks in the diagnosis of urological disfunctions. *Expert Syst Appl* 36: 5754-5760, 2009.
60. Simões PW, Izumi NB, Casagrande RS, Venson R, Veronezi CD, Moretti GP, da Rocha EL, Cechinel C, Ceretta LB, Comunello E, *et al*: Classification of images acquired with colposcopy using artificial neural networks. *Cancer Inform* 13: 119-124, 2014.
61. Sato M, Horie K, Hara A, Miyamoto Y, Kurihara K, Tomio K and Yokota H: Application of deep learning to the classification of images from colposcopy. *Oncol Lett* 15: 3518-3523, 2018.
62. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmar C, Bakers FCH, Peters NHGM, Beets-Tan RGH and Aerts HJWL: Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Sci Rep* 7: 5301, 2017.
63. Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, Sköldenberg O and Gordon M: Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop* 88: 581-586, 2017.
64. Khosravi P, Kazemi E, Zhan Q, Toschi M, Makmsten J, Hickman C, Meseguer M, Rosenwaks Z, Elemento O, Zaninovic N and Hajirasouliha I: Robust automated assessment of human blastocyst quality using deep learning. *bioRxiv* 394882, 2018.
65. Miyagi Y, Habara T, Hirata R and Hayashi N: Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image. *Reprod Med Biol* 18: 204-211, 2019.
66. Miyagi Y, Habara T, Hirata R and Hayashi N: Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age. *Reprod Med Biol* 18: 190-203, 2019.
67. Sideri M, Garutti P, Costa S, Cristiani P, Schincaglia P, Sassoli de Bianchi P, Naldoni C and Bucchi L: Accuracy of colposcopically directed biopsy: Results from an online quality assurance programme for colposcopy in a population-based cervical screening setting in Italy. *BioMed Res Int* 2015: 614035, 2015.
68. Sideri M, Spolti N, Spinaci L, Sanvito F, Ribaldone R, Surico N and Bucchi L: Interobserver variability of colposcopic interpretations and consistency with final histologic results. *J Low Genit Tract Dis* 8: 212-216, 2004.
69. Massad LS, Jeronimo J, Katki HA and Schiffman M; National Institutes of Health/American Society for Colposcopy and Cervical Pathology Research Group: The accuracy of colposcopic grading for detection of high grade cervical intraepithelial neoplasia. *J Low Genit Tract Dis* 13: 137-144, 2009.
70. LeCun Y, Haffner P, Bottou L and Bengio Y: Object recognition with gradient-based learning. In *Shape, contour and grouping in computer vision*. Springer, Berlin, Heidelberg: 1999.
71. Hu J, Shen L and Sun G: Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp7132-7141, 2018.
72. Kudva V, Prasad K and Guruvare S: Automation of detection of cervical cancer using convolutional neural networks. *Crit Rev Biomed Eng* 46: 135-145, 2018.
73. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM and Thrun S: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542: 115-118, 2017.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.