RESEARCH

Endometrium Cancer Detection with ConvNet

Dilek Erdem[1], Mete Yildirim[2], Ali Buhur[3]

[1]Department of Gynecology and Obstetrics, Alanya Education and Research Hospital, Alanya Alaaddin Keykubat University, Antalya, Turkey
[2]Department of Electrical and Electronic Engineering, Faculty of Engineering, Ege University, Izmir, Turkey ,Ege University, Izmir, Turkey
[3]Department of Gynecology and Obstetrics ,Kanuni Sultan Süleyman Education and Research Hospital, Istanbul, Turkey

ORCID; 0000-0001-6199-0333[1], 0000-0001-6335-4752[2] 0000-0003-1228-0962[3]

Abstract: Endometrial cancer is diagnosed using histopathologic images. The endometrium histopathologic images are commonly described, such as normal endometrium, endometrial polyp, endometrial hyperplasia, and endometrial adenocarcinoma by pathologists. In this study, we developed a deep learning model to help medical practitioners to classify endometrium histopathologic images and diagnose endometrial cancer. A convolutional neural network is developed and used to classify histopathologic images. The proposed model for classification was adapted to handle both binary and multiclass classification. In the binary classification phase, the model was trained for various scenarios, such as classification for normal endometrium against abnormal endometrium. For the multiclass classification, the model was trained to classify for various classes. Approaches and results for these classifications were discussed in detail with tables. Accuracy, the area under the curve, precision, recall scores, and F1 scores were calculated and used as evaluation metrics. The result showed that the proposed model achieves better classification than the previous studies, specifically in binary classification, and could be used as a helpful diagnostic tool.

Received : 31/01/2022
Received in revised form : 03/04/2022
Accepted : 04/10/2022
Available online : 06/21/2022

Keywords:

Convolutional neural network
Deep Learning
Endometrial cancer
Uterus

Corresponding Author:
Dilek Erdem
E-mail; dilek.erdem@alanya.edu.tr
http://dx.doi.org/10.47009/jamp.2022.4.2.4

Int J Acad Med Pharm,
2022; 4 (2); 16-20

INTRODUCTION

The uterus is an organ of the female reproductive system and has two structurally and functionally different parts, the cervix and the corpus. The corpus is the muscular part at the upper end, and the cervix is the fibrous part at the lower end. In addition, the transition zone between these two structures is called the isthmus. The inner surface of the corpus is covered with endometrium. The superficial part of the endometrium undergoes structural and functional changes in the proliferation, secretion, and menstrual phases during the menstrual cycle. Pathologies such as polyps, hyperplasia, and cancer may develop from the endometrium [1] %83 of corpus uteri cancers are endometrial cancers [2] According to Global Cancer, the 5 year prevalence of corpus uteri cancers prevalence among women of all ages worldwide was approximately one million and the number of deaths was over three hundred thousand. Although many different methods are used to diagnose endometrial cancer, the definitive diagnosis is made by histopathological examination [3]

The workload of pathologists is overwhelming due to the insuficient number of pathologists. Evaluation of the specimens takes a long time, and when the workload increases, the rate of incorrect assessment increases [4] The advancements in modern computing technology have brought new trends to a wide range of practical applications in numerous different industries. Artificial intelligence (AI) is one of the major breakthroughs in the digitalized world in terms of healthcare widely applied for half a century [5] The concept and definition of artificial intelligence have changed from past to present. However, the origin of the idea is to build intelligent machines that imitate human level thinking and behavior to perform specific tasks. Common misconceptions tend to limit the capabilities of AI to humanlike robots or self driving cars. However, AI encompasses way beyond this narrow approach, including processing large amounts of data generated daily. For example, an AI model "AplpaGo Zero" beat the "Go" champion in 2016 and the more advanced model beat the old one a hundred times in a row. The importance of artificial intelligence in digital medicine comes at this point. AI in digitized medicine promises alterations in practical health care interventions using computer algorithms and software to gain information, process it, and give satisfactory results to the end user. The initial step here is to empower the machines to learn certain tasks using machine learning algorithms [6]

Machine learning is considered a subdiscipline of artificial intelligence that performs various complex tasks such as pattern recognition, predictive high-dimensional data analysis, and image interpretations[7] The rise of machine learning and artificial neural networks trace back to the early 1960s[8] However, more advanced forms were applied in the late 1990s with a shifted focus toward methods in statistics and probabilistic approaches, specifically in medical diagnosis [9]. Just as machine learning is a subset of AI, deep learning is a subset of machine learning and a special type of very large multilayered neural networks[10]. Deep learning differs from traditional neural networks due to its self-learning ability through the filtration of information through multiple hidden layers and its large number of hidden layers [11]. The neural nets of the 1960s were not suffcient to meet medical demands. Today's much faster computers, bigger information datasets, and technical refinements allow researchers and healthcare providers to overcome several limitations [12].

Recent applications of deep learning in medicine involve the omics, the study of the genome (genomics) and proteins (transcriptomics, proteomics, and metabolomics), bioimaging (concerning biological cells and tissues), medical imaging (visual representations of body parts), BBMI (brain/body machine interfaces) and public and medical health management (PmHM) [13]. Successful consequences of the use of deep learning in computer vision paved the way for clinical image processing, data analysis, and diagnostics, particularly on the Magnetic Resonance Imaging (MR) scans of lesions in various tissue structures and tumors [14]. Deep learning is also used to classify pathological images [15]. Diagnosing from pathologic images is crucial and can take a long time. However, classifying them with machine learning can help to shorten the diagnosis process. Hence the diagnosis of pathology images with machine learning will facilitate the work of experts and save resources. However, the progress is slower than expected regarding the use of deep learning in medicine due to the shortage of staff who can establish a relationship between these two disciplines [16]. There are very few studies about the detection of endometrial cancer with deep learning. One study offered a computer-aided diagnosis 2 method to classify histopathological endometrial cancer lesions into four groups [17], and another study used deep learning method to determine myometrial invasion depth by using MR images [18]. Another study classiffied histopathological images using a Convolutional Neural Network [19].

MATERIAL and METHODS

Ethical approval

Hao S. et al enable other researchers such as us to download and use histopathlogical images freely, therefore it was not required to get ethical approval.

Histopathological Images

We used the data set of 3302 histopathological images classified into four classes from a study conducted by Hao Sun et al. available for authors for further researches [17]. Hao S. et al enable other researchers such as us to download and use them, freely. (https://doi.org/10.6084/m9.figshare.7306361.v2) The images where collected from samples which underwent a standart protocol of histopathological reporting, examined and chosen by three skilled pathologist unanimously. Then these images were processed into dijital images. Diagnoses of these images were classified in four groups which are endometrial polyp (EP), endometrial hyperplasia (EH), normal endometrium (NE), and endometrial adenocarcinoma (EA). Furthermore, EH is classified into Complex" (short for complex hyperplasia without atypia) and Simple (short for simple hyperplasia without atypia) classes while NE is classified into Follicular, Luteal, Menstrual. Table 1 shows the dataset in summary. We also defined two general classes to differ EA ("Malignant") from NE, EP, and EH ("Benign"). The subclasses are hidden in the table, and the dataset has imbalanced distribution. The class with the least number of images (EA) is like %40 of the class with the greatest number of images (NE). Imbalance among classes in the dataset makes the training process harder. Hence, we set up class weights for training. Instead of classifying among only the given for classes, we redesign the classes that have di_erent imbalance ratios. Table 2 shows all classification scenarios where imbalanced distribution is improved relatively. For example, in Task-7 classification is done between NE and (EH, EP), hence number of images in each one come close to each other. If this is not possible, for example Task-6, we show the model the images from a group with lesser number of images more frequently to introduce balance for the model training.

Deep Learning Basics

Deep learning is a sub-branch of machine learning and can be considered the biggest step towards real artificial intelligence. Deep refers to the number of layers used in the model. Structurally, it consists of artificial neural networks consisting of many layers, each of which generally has many neurons. However, it has models like Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) that are different from common neural networks. The most important feature of deep learning is that it shows a performance that is directly proportional to the increasing amount of information, unlike other learning models. Also, deep learning models work better with larger models that are known as scalability.

Table 1: The distribution of original histopathological images dataset

	NE	EA	EP	EH	Total
Number of images	1333	535	798	636	3302

EP: Endometrial polyp, EH: endometrial hyperplasia, NE: normal endometrium, EA: endometrial adenocarcinoma

Table 2: The model used for classification

	Total parameters	Trainable parameters	Non-trainable parameters
Model	53,593	52,409	1,184

Deep learning can process raw data, known as feature extraction. One of the pioneers in the field defines it as ‘’Deep learning methods aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower-level features. Automatically learning features at multiple levels of abstraction allow a system to learn complex functions mapping the input to the output directly from data, without depending completely on human-crafted features" [20]. Deep learning model training owes its success to the intelligent implementation of backpropagation. Rumelhart [11] states that backpropagation is just a clever application of the chain rule. The use of backpropagation is quite old, but because of some wrong approaches, it did not have the good performance it has today in the past. These are: using a small number of datasets, wrong initializing weights, and the wrong type of non-linearity[10].

Deep learning can process many forms of data, including images, text, video, and audio, which differs from other neural networks (NN). The type of deep learning architecture used at object recognition in image data is named the Convolutional Neural Network (CNN). CNN can be considered as one of the modern supervised learning algorithms like Multilayer Perceptron Networks (MPN) and Long Short-Term Memory Recurrent Neural Networks (LSTM- RNN) that use the backpropagation in training. Since this study is related to image classiffication, it is more appropriate to dive into a little detail of the main blocks of CNN.

ConvNets

ConvNets, also known as CNNs, is used for image classification. Data can be structured or unstructured, and an image is classiffied as unstructured data. It can contain thousands of numbers when converted to digital form since every pixel on the image is represented with a number. For example, a grayscale image with a pixel size 32 by 32 consists of 32 x 32 numbers. In the case of a color image, the number multiplied by 3. One can think of it as a future vector applied fully connected layers with 3076 features that require a big weighting matrix plus a bias for a single neuron. The addition of more and more layers and using a bigger size image introduces inefficiency in computing, at least in model training. Besides, converting pixel values into a long vector
may cause loss of spatial information between pixels. CNN eliminates these problems intelligently

The building blocks of CNN are known to be convolution, pooling, and fully connected layers. A CNN model can have multiple of these layers. The convolutional name comes from the convolution operation in math but works differently. It carries out filtering, feature mapping, and zero paddings. The filters are essentially neurons of CNN architecture. The neurons take the pixel values as numbers and produce output. The output of one filter is called a future map. The zero-padding ensures the edge of input is not read off by creating artificial inputs around the real input. The pooling layers downsamples the feature map. The max-pool is generally used for pooling, and it is not a must to use it after every convolutional layer. The pooling is simple and implemented by taking the maximum or the average value of the input in the defined field to create a future map at the output. It generally combines or generalizes features of the previous layer; hence, it helps reduce overfitting in the training process. The real benefit of pooling is often observed by experiment. The experience shows that max-pool averaging often produces better results [21]. It is general practice to have fully connected layers through the end of the model. As the name implies, all the neurons are connected to all the input in this type of layer. This layer combines all learned features in a non- linear manner, and at the very last layer, it is used with the soft-max layer for the prediction of classes.

RESULTS

The developed model is fully provided in supplementary documents. The model is easily adapted to a different type of classiffication by changing the last layer according to the desired number of classes.

Table.2 shows the total parameters, trainable parameters, and non-trainable parameters of the proposed model. Generally, the numbers of parameters are very moderate in the proposed model compared to Resnet50 (23 million) and VGG16 (138 million). Actually most of the models used for image classification generally have million of parameters unlike the proposed model. Hence, the proposed model requires less training time. Small models do not require as much data as large models (so many layers and residual connections). It is not possible to train large models with a small dataset to have good results. It can have overfitting that causes the model to lose its generality.

Since the dataset used in this study is small, we use image augmentation such as random flip, random rotation, random contrast, random translation, and random width (Figure 2). %90 of the available data is used for training. Furthermore, we set up class weights in training progress for the given classiffication task.

We consider area under curve(AUC), accuracy, precision, recall, and F1 score values to evaluate all models (Figure 3). AUC is calculated from the receiver operating characteristic curve (ROC), and it ranges in value from 0 to1. If predictions are 100% correct, the model will have an AUC of 1:0. AUC is desirable because it is scale invariant and classification-threshold-invariant. However, these two may not always be desirable. It may be the case when the classes are imbalanced, which is true for this study. Accuracy shows how often a classiffier is correct. Precision tells us how often it is correct when it predicts yes. Recall or Sensitivity shows how often it predicts ‘’yes" when it is really ‘’yes". F1 Score is the choice as a better measure to use if we need a balance between Precision and Recall, and when there is an uneven class distribution.

We set up the predefined tasks, as given in Table.3. Task-1 trains the model on the original dataset to have a trained model that can predict an image class among EA, EP, EH, NE. Table 3 shows that Task-1 has an accuracy of %76 on the validation dataset that shows that the proposed model can achieve the same accuracy as in Sun H (2019). However, we still consider that this performance is poor and looked at different classiffication strategies. Hence, we set up a second task to train the model to classify between malignant and other classes. The classification metrics suggest that the model shows very good promising achievement for this task with0:98 area under the curve, % 96 accuracy, and 0:975 F1 Score (Table 4). It means the model has very high sensitivity and specificity that it is far better than Sun, H., et al. [17]

The negative side of Task-2 is that it tells us if an image belongs to EA class but does not tell if it is not EA what it is. Hence, we set up Task-3 to determine if an image in another class what class it is. Unfortunately, this model does not perform well on Task-3 and offers a very similar result to Task-1. Still, Task-3 makes us understand that the model has difficulties in classifying among NE, EP and EH. Hence, we set up Task-4 to Task-9 to examine and determine what classes are hard to classify. A close examination of Task-4 to Task-9 shows that the model is less successful, separating NE and EP classes. We also consider a case in Task-10 where the model is trained to classify an image, either Normal or Abnormal. The Normal class only has NE images, whereas the Abnormal classes have images of EA, EP, and EH. Since Abnormal includes EA, EP, EH, we consider that it can create alerts for close patient monitoring. However, the result is not very promising since the F1 score remains around %77. The reason for that comes from the fact that the model is not very successful in separating an image of EP from an image of NE. When the tasks are evaluated all together, it is seen that following the decision path from Task-2 and then Task-8, an EA case could be determined with the F1 score of 0:975 (precision:0:98 and recall: 0:97) and an EH case could be determined with the F1 score of 0:925.

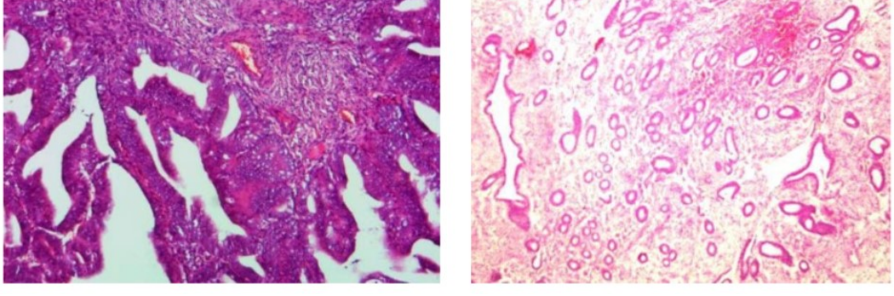

Figure 1. Endometrial adenocarcinoma (EA-left) and Normal endometrium (NE- right).

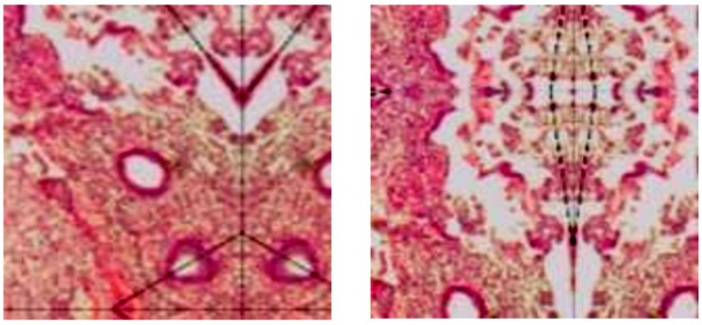

Figure 2. The augmented images with random zooming and flipping.

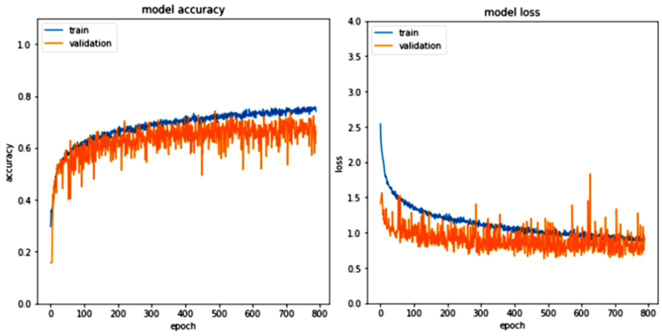

Figure 3. The accuracy and loss curve.

Table 3: The classification scenarios

Mission	Classes	# of images	Explanation
Task-1	NE	1333	Four-class classification among NE, EA, EP, EH
	EA	535	
	EP	798	
	EH	636	
Task-2	Malignant	535	Two-class classification, Malignant has EA images, Other has NE, EP, EH images
	Other	2767	
Task-3	NE	1333	Three-class classification among NE, EP, EH
	EP	798	
	EH	636	
Task-4	EH	636	Two-class classification between EH and NE
	NE	1333	
Task-5	EP	798	Two-class classification between EP and EH
	EH	636	
Task-6	EP	798	Two-class classification between EP and NE
	NE	1333	
Task-7	NE	1333	Two-class classification between NE and (EH, EP)
	EP&EH	1434	
Task-8	EH	636	Two-class classification between EH and (EP&NE)
	EP&NE	2131	
Task-9	EP	798	Two-class classification between EP and (NE&EH)
	NE&EH	1969	
Task-10	Normal	1333	Two-class classification, Abnormal has EA, EH, EP images, Normal has NE images
	Abnormal	1969	

EP: Endometrial polyp, EH: endometrial hyperplasia, NE: normal endometrium, EA: endometrial adenocarcinoma

Table 4: The model evaluation metrics.

Mission	AUC	Accuracy	Precision	Recall	F1 Score
Task-1	0.93	0.76	0.78	0.73	0.75
Task-2	0.98	0.96	0.98	0.97	0.975
Task-3	0.91	0.76	0.78	0.72	0.745
Task-4	0.92	0.86	0.87	0.91	0.89
Task-5	0.96	0.91	0.87	0.91	0.89
Task-6	0.89	0.83	0.88	0.87	0.875
Task-7	0.88	0.81	0.85	0.76	0.80
Task-8	0.92	0.89	0.91	0.94	0.925
Task-9	0.90	0.85	0.70	0.67	0.684
Task-10	0.89	0.83	0.80	0.74	0.77

DISCUSSION

In this study, we develop a CNN model to assist medical practitioners in the diagnosis of endometrium cancer from histopathological images. We aim to develop a new model with better classification performance. The proposed model is trained for both binary and multiclass classification. The performance of the model has been analyzed by creating many different binary and multiclass tasks. Unlike the previous study,[17] the imbalance between classes is addressed in model design, and many data augmentation techniques are used to reduce the problems that may be caused by the scarcity of data. In addition, the effect of the imbalance in the data distribution is tried to be reduced with the classiffications targeting different tasks. To evaluate the performance of the model, AUC, accuracy, precision, and recall scores are given with the table 4. F1 score is used as a core evaluation metric since it is the most regarded one when the imbalance is an issue among the given classes. Considering this score, it was observed that the model was very successful in a binary classification where malignant versus benign classification is objected.

To find out why this success could not be achieved in multiclass classification; many binary classiffcation cases were examined. It was observed that the difficulty comes from classiffication between NE and EP classes, and it reduces training and validation performance of the model where these two are considered in separate classes. The lower part of the uterine segment and deep endometrial basalis can closely resemble and can be confused with polyps [22], and explain the difficulty of classiffication between NE and EP classes. To solve this problem, it is advised to obtain more images of these two classes, and further studies are needed.

To work this patern in the clinical environment we need to consider few things. First, collecting more image data in volume and type will help to enrich this dataset and it will be useful for more accurate identification of more types of endometrial diseases. For example, this dataset does not contain any image of atypical endometrial hyperplasia, which is premalignant. Second, we advise researcher to design an interactive framework of human in-the loop for ConvNet. [23] ConvNet needs to learn how to attach human expert's feedback to correct its wrong decisions. Third, applying maching learning techniques like deep reinforcement learning [24] and transfer learning [25], will amplify the self-learning ability that can convert human expert's knowledge and skill into machine-readable representation and teach itself.

As far as we know, our study is the first one using ConvNet with histopathological images of endometrium. However, there is another study conducted by Dong[26] et. al to use AI technology to evaluate the depth of myometrial invasion using a deep learning technique on magnetic resonance images in early stage endometrial cancers. Masazaku S. et al.'s study[27] on application deep learning to the classification of images from colposcopy has shown us, that clinicians and researchers, who are not specialists in artificial intelligence or machine learning, can utilize deep learning as decision support systems. The strong point of our study is that it shows us that we can use ConvNet for differing malign endometrial lesions from non-malignant specimens. Thus it can be a decision spport system to either propose a decision to the pathologists to ease their decisions or it can eliminate a part of data that should be processed.

Conflict of interest

The authors declare that there are no conflict of interests.

Financial disclosure

The authors declared that this study has received no financial support.

REFERENCES

1. Hoffman B, Schorge J, Schaffer J, Halvorson L, Bradshaw K et al. *Williams Gynecology, Second Edition*. McGraw Hill Professional; 2012:702–713.
2. Mahdy H, Casey MJ, Crotzer D. Endometrial Cancer. 2021 Aug 25. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 3025223
3. Revel A, Shushan A. Investigation of the infertile couple: hysteroscopy with endometrial biopsy is the gold standard investigation for abnormal uterine bleeding. *Human Reproduct on* . 2002;17(8):1947–1949.
4. Litjens G, S_Anchez CI, Timofeeva N, Hermsen M, Nagtegaal I. et al. Deep learning as a tool for increased accuracy and e_ciency of histopathological diagnosis. *Sc ent f c Reports*. 2016;6(1):1–11.
5. Hinton G. Deep learning:a technology with the potential to transform health care. *Jama*. 2018;320(11):1101–1102.
6. Christopher B. *Pattern recogn t on and mach ne learn ng*. Cambridge, U.K. Springer; 2006.
7. Naylor CD. On the prospects for a (deep) learning health care system. *JAMA: the journal of the American Medical Association*. 2018.
8. Samuel AL. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*. 2000;44(1.2):206–226.
9. Langley P. The changing science of machine learning. *Mach ne Learning*. 2011;82(3):275–279.
10. Bengio Y, Yann L. Scaling learning algorithms towards AI. *Large-scale Kernel Machines*. 2007;34:1–41.
11. Rumelhart DE, Hinton GE, Williams RJ. Learning Representations by Backpropagating Errors. 1986:533–536.
12. Lee J-G, Jun S, Cho Y-W, Lee H, Kim GB et al. Deep learning in medical imaging: General overview. *Korean journal of rad ology: official journal of the Korean Radiological Society*. 2017;18(4):570.
13. Zemouri R, Zerhouni N, Racoceanu D. Deep learning in the biomedical applications: Recent and future status. *Appl ed Sc ences (Basel, Switzerland)*. 2019;9(8):1526.
14. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Br ef ngs n Bioinformatics*. 2018;19(6):1236–1246.
15. Osamu I, Fahdi K, Kei K, Rambeau M, Koji A et al. Deep Learning Models for Histopathological Classi cation of Gastric and Colonic Epithelial Tumours. *Sc ent f c Reports*. 2020;10(1).
16. Drukker L, Noble JA, Papageorghiou AT. Introduction to arti cial intelligence in ul- trasound imaging in obstetrics and gynecology. *Ultrasound in Obstetrics and Gynecology*. 2020;56:498–505.
17. Sun H, Zeng X, Xu T, Peng G, Ma Y. Computer-Aided Diagnosis in Histopathological Images of the Endometrium Using a Convolutional Neural Network and Attention Mechanisms. *IEEE Journal of B omed cal and Health Informatics*. 2020;24(6):1664–1676.
18. Chen X, Wang Y, Shen M, Yang B, Zhou Q et al. Deep learning for the determination of myometrial invasion depth and automatic lesion identification in endometrial cancer MR imaging: a preliminary study in a single institution. *European Rad ology*. 2020:1–10.
19. Hatipoglu N, Bilgin G. Classification of histopathological images using convolutional neural network.; 2014.
20. Lamblin P, Bengio Y. Important gains from supervised ne-tuning of deep architectures on large labeled sets. Deep Learning and Unsupervised Feature Learning Workshop. 2010:1–8.
21. LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems. IEEE; 2010.
22. Robboy SJ. Robboy's Pathology of the Female Reproductive Tract. Elsevier Health Sciences; 2009.
23. Zanzotto FM. Viewpoint: Human-in-the-loop artificial intelligence. *The Journal of Artificial Intelligence Research*. 2019;64:243–252.
24. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529–533.
25. Pan SJ, Yang Q. A survey on transfer learning. *IEEE transact ons on knowledge and data engineering*. 2010;22(10):1345–1359.
26. Dong HC, Dong HK, Yu MH, Lin YH, Chang CC. Using Deep Learning with Convolutional Neural Network Approach to Identify the Invasion Depth of Endometrial Cancer in Myometrium Using MR Images: A Pilot Study. *Int J Env ron Res Publ c Health*. 2020;17(16):5993.
27. Sato, M., Horie, K., Hara, A., Miyamoto, Y., Kurihara, K., et al. "Application of deep learning to the classification of images from colposcopy". *Oncology Letters* 15.3 (2018): 3518-3523