HIR
Healthcare Informatics Research

Application of Convolutional Neural Network in the Diagnosis of Jaw Tumors

Wiwiek Poedjiastoeti, DDS[1], Siriwan Suebnukarn, DDS, PhD[2]

[1]Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, Trisakti University, Jakarta, Indonesia; [2]Faculty of Dentistry, Thammasat University, Pathumthani, Thailand

Objectives: Ameloblastomas and keratocystic odontogenic tumors (KCOTs) are important odontogenic tumors of the jaw. While their radiological findings are similar, the behaviors of these two types of tumors are different. Precise preoperative diagnosis of these tumors can help oral and maxillofacial surgeons plan appropriate treatment. In this study, we created a convolutional neural network (CNN) for the detection of ameloblastomas and KCOTs. **Methods:** Five hundred digital panoramic images of ameloblastomas and KCOTs were retrospectively collected from a hospital information system, whose patient information could not be identified, and preprocessed by inverse logarithm and histogram equalization. To overcome the imbalance of data entry, we focused our study on 2 tumors with equal distributions of input data. We implemented a transfer learning strategy to overcome the problem of limited patient data. Transfer learning used a 16-layer CNN (VGG-16) of the large sample dataset and was refined with our secondary training dataset comprising 400 images. A separate test dataset comprising 100 images was evaluated to compare the performance of CNN with diagnosis results produced by oral and maxillofacial specialists. **Results:** The sensitivity, specificity, accuracy, and diagnostic time were 81.8%, 83.3%, 83.0%, and 38 seconds, respectively, for the CNN. These values for the oral and maxillofacial specialist were 81.1%, 83.2%, 82.9%, and 23.1 minutes, respectively. **Conclusions:** Ameloblastomas and KCOTs could be detected based on digital panoramic radiographic images using CNN with accuracy comparable to that of manual diagnosis by oral maxillofacial specialists. These results demonstrate that CNN may aid in screening for ameloblastomas and KCOTs in a substantially shorter time.

Keywords: Artificial Intelligence, Ameloblastoma, Odontogenic Tumors, Panoramic Radiography, Oral and Maxillofacial Surgeons

I. Introduction

Ameloblastomas are classified as benign, locally aggressive tumors in the oral and maxillofacial area. The origin is the odontogenic epithelium. The signs and symptoms of ameloblastomas are usually painless and usually identified in the oral panoramic radiography. Asymptomatic swelling of the jawbone is the classic clinical finding [1]. Odontogenic keratocyst (OKC) is classified as a benign cystic lesion. Its origin is also the odontogenic epithelium. The histological findings present a parakeratinized stratified squamous epithelium lining with specific infiltrative and aggressive behavior. The radiographic findings show a mixture of multicystic and

unicystic lesions. Since 2005, the World Health Organization (WHO) has labeled OKCs as keratocystic odontogenic tumors (KCOT) and has classified OKCs as tumors according to their behavior. The histological and radiographic findings of OKCs are more similar to those of neoplasms than to those of odontogenic cysts [2]. The treatment modalities for these two types of tumors are different. KCOTs, like other cystic lesions, are usually enucleated without radical jaw segmentation. Ameloblastomas, on the other hand, require more radical surgical removal than KCOTs. Segmental or en bloc resection of the jaw is performed in solid or multicystic ameloblastomas. Reconstruction is required after surgery using cancellous bone augmentation with a fixation plate. A free vascularized bone flap is used in a large block surgical resection of the jaw. Precise preoperative diagnosis between these two types of tumors can help oral and maxillofacial surgeons plan appropriate treatment [3].

With routine two-dimensional (2D) oral panoramic radiography, differential diagnosis between ameloblastomas and KCOTs is not easy. Ameloblastomas and other cystic lesions in the orofacial areas have similar radiological characteristics in conventional panoramic radiographic imaging [1]. The typical features of ameloblastoma radiographic findings are multilocular or unilocular radiolucent lesions with expanded dilution and extension of the overlying cortical bone. The embedded teeth are found within radiolucent lesions in some cases. These findings do not indicate ameloblastomas and may also indicate other odontogenic tumors, such as KCOTs or other cystic lesions. The evaluation criteria for distinguishing KCOT from ameloblastoma are solitary unilocular lesions extending longitudinally in the posterior region of the mandible. They typically grow along the bone in the anteroposterior dimension. They can sometimes appear compartmentalized, which makes the distinction of ameloblastoma difficult. [3].

Several researchers have attempted to differentiate between these two tumors using computed tomography and magnetic resonance imaging [4,5]. It has been reported that the diversity of computed tomography density in KCOTs is higher than that in ameloblastomas. According to the findings of these studies, we have tried to obtain a more effective classification with an artificial intelligence method, especially with a system that can screen medical images. Deep learning has powerful machine learning algorithms that can be employed to classify diseases using large numbers of retrospective medical images as input data [6]. Deep learning enables calculation models that consist of multiple layers of processing to learn data representations with several abstraction levels.

Convolutional neural networks (CNNs) are powerful machine learning algorithms for image learning. CNNs are constructed to learn various types of data, especially 2D images. The learning mechanisms mimic the function of the visual cortex of the brain, where there is a hierarchy of simple and complex cells. The ultimate goal of CNN training algorithms is to optimize the weighting parameters in each layer of the architecture. A learning algorithm combines simpler features into complex features, leading to the ultimate hierarchical representations derived from image data [7].

Biomedical imaging is another vigorously researched domain in which deep learning is extensively used. Considering the requirement to optimize an enormous number of weighting parameters in CNNs, most deep learning algorithms depend upon a large number of data samples for training. Unfortunately, data collection in medical informatics is considered a complex and expensive process. In the clinical context, it is not easy to obtain a large number of training datasets, leading to overfitting of the model under the constraints of the small datasets [6]. As a result, a modified deep CNN has been investigated, which is based on transfer learning of an unsupervised pre-training from a large number of datasets to solve the problem of the small datasets [8].

In this study, we developed a CNN for the detection of ameloblastomas and KCOTs in digital panoramic images. We used the transfer learning from the large sample dataset on our radiographic dataset with known biopsy results for the secondary training. To assess the ability of the developed CNN to classify ameloblastomas and KCOTs using panoramic radiographs, we tested the system by comparing its accuracy in diagnosing ameloblastomas and KCOTs with that of oral and maxillofacial surgeons.

II. Methods

The data were obtained retrospectively from a university hospital in the context of a protocol approved by the Ethical Committee of Thammasat University (No. 021/2559). The dataset comprised 250 ameloblastomas and 250 KCOT lesions on panoramic digital X-ray images with known biopsy results. To overcome the imbalance of data entry, we focused our study on 2 tumors with equal distributions of input data. The training data comprised 200 ameloblastoma images and 200 KCOT images. The test data comprised 50 ameloblastoma images and 50 KCOT images. To overcome the limitation of the small training dataset, we applied the data augmentation method to increase the number of train-

ing datasets. Second, a VGG-16 (16-layer CNN) [9] was pre-trained in ImageNet, and refined with our secondary dataset training. The gradient weighted class activation maps (Grad-CAM) were also developed to identify the discrimination areas on the panoramic digital X-ray images, along with the prediction probability. A separate test dataset with known biopsy results was evaluated to compare the performance of the developed CNN with that of board certified oral and maxillofacial specialists.

1. Data Preparation

The original panoramic X-ray images tended to have high brightness with a lower contrast value. We applied inverse logarithm transformation to expand the high value (white pixel) and compress the values for the dark layers to achieve appropriate opacity for image viewing. The image quality was improved by histogram equalization to overcome the low contrast in the images. In this study, we used Horos, an open-source software for free medical image viewers with a package of DICOM images as well as preprocessing and management tools. The result of the preprocessing method is shown in Figure 1. To reduce the over-fitting as a result of learning from the small datasets, this work used data augmentation to expand the size of the image samples through offline horizontal flipping before the data was fed to the machine learning model. By flipping all the images horizontally, the size of the training and test datasets was doubled.
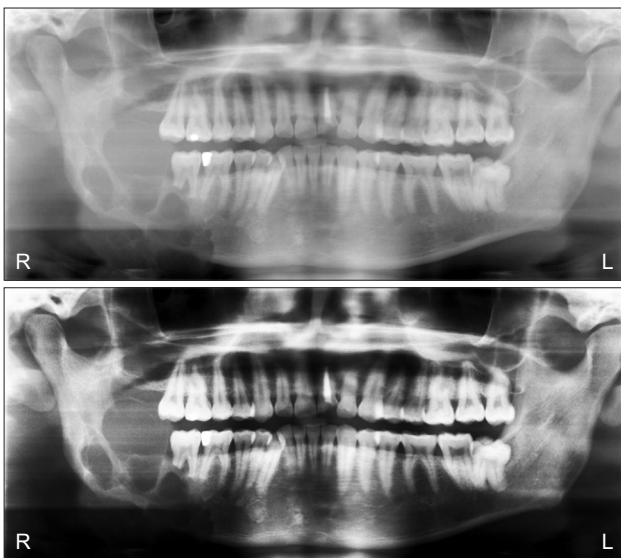

Figure 1. Original input image (upper) was preprocessed using inverse logarithm transformation and histogram equalization (lower).

2. CNN Training Algorithm

Generally, machine learning depends upon the training data and test data having similar characteristics. Unfortunately, in the clinical context, it is not easy to acquire training data that have characteristics similar to those of the test data. To overcome this difficulty, the transfer learning method has been applied to neural convolution networks [8]. To acquire an ultimate transfer model, our neural network is first trained on a large and similar example set of source domain data for the transmission of small pattern knowledge.

We used the VGG-16 [9] as our network architecture. The VGG-16 consists of 16 layers of convolutional and fully-connected layers. Thirteen convolutional layers form 5 groups, and each group is followed by a max-pooling layer that down-samples the images to reduce computational cost and control over-fitting. We pre-trained the VGG-16 CNN with the ImageNet [10], a large image dataset of more than 1,000 object classes. Providing an input test image, the pre-trained VGG-16 CNN will generate the test image's probabilities associated with each object category. Various hyper-parameters were explored, including the learning rate and the number of epochs. We tried various learning rates from 1e-3 to 1e-5 and discovered that the standard learning rate at 1e-3 in many deep learning applications was not appropriate for our data and caused significant over-fitting. The lowest learning rate with 1e-5 with reasonable epochs for 300 was able to prevent the problem from being exceeded in the experiment.

3. Model Interpretation

To interpret the network predictions, we used class activation mappings (CAMs) to generate heatmap visualizations of the areas of each digital panoramic radiographic image that mostly indicate the presence of a tumor [11]. To produce the CAMs, each image was entered into the final trained CNN. The output feature maps were extracted at the end of each convolutional layer. A 'class activation' heatmap is a 2D grid of scores associated with a specific output class, computed for every location in any input image, indicating how important each location is with respect to the class considered.

4. Model Evaluation

The final trained CNN computed the probability of an ameloblastoma or KCOT occurring in a panoramic radiographic image. The separated set of test panoramic radiographs predicted by the final trained CNN was compared to the results produced by 5 board certified oral and maxillofacial surgeons. The sensitivity, specificity, and accuracy of the

ameloblastoma or KCOT classifications by the CNN and those by the oral and maxillofacial surgeons were measured and compared. The receiver operating curves (ROC) were plotted and the area under the ROC (AUC) was evaluated. All statistical analyses were performed using Stata/MP version 14.2, and a *p*-value of < 0.05 was considered statistically significant. The details of patient information were not accessible to the CNN training process, and the oral and maxillofacial surgeons who were involved in the evaluation were not identified. This study was approved by the Ethical Committee.

III. Results

The final CNN developed in this study generated a binary classification output of the probability of ameloblastoma or KCOT per panoramic radiographic image. The AUC was 0.88, and the sensitivity, specificity, and accuracy were 81.8%, 83.3%, and 83.0%, respectively. The total calculation time taken by the CNN to analyze all the images was 38 seconds. The overall sensitivity, specificity, and accuracy for the classification of ameloblastoma or KCOT by the 5 oral and maxillofacial surgeons were 81.1% (SD = 11.3), 83.2% (SD = 9.7%), and 82.9% (SD = 8.1%), respectively. The average time to evaluate all the images of the test datasets was 23.1 (SD = 3.0) minutes. The AUC was 0.88 at the cutoff point of the operating threshold of 0.43. The performance of the CNN pre-trained with ImageNet was not statistically different from that of the 5 oral and maxillofacial surgeons in terms of its sensitivity, specificity, and accuracy. However, the CNN provided the classification in a significantly shorter time in comparison to the oral and maxillofacial surgeons.

We identified the most important features used by the model in its tumor classification using the class activation mappings to the dimensions of the image and overlaying the image. Figures 2–5 show several examples of heatmaps on the correct ameloblastoma and KCOT prediction task. Figure 2 is an ameloblastoma showing a well-defined multilocular lesion on the right angle-ramus region. The septa inside the lesion are thin and straight and located perpendicular to the periphery, giving the appearance of soap bubbles. Resorption of the first molar roots is seen. Figure 3 is an ameloblastoma showing an extensive unilocular lesion on the mental region of the mandible. The boundary is well delineated and there is no matrix calcification. Figures 4 and 5 are KCOTs showing a unilocular radiolucent lesion in the posterior mandible. These cases were associated with an embedded tooth. The CNN localized tumors it identified with the heatmap areas of the panoramic radiographic image that were most important for making a tumor classification. The captions for each image were provided by one oral and maxillofacial surgeon who did not participate in the model evaluation.
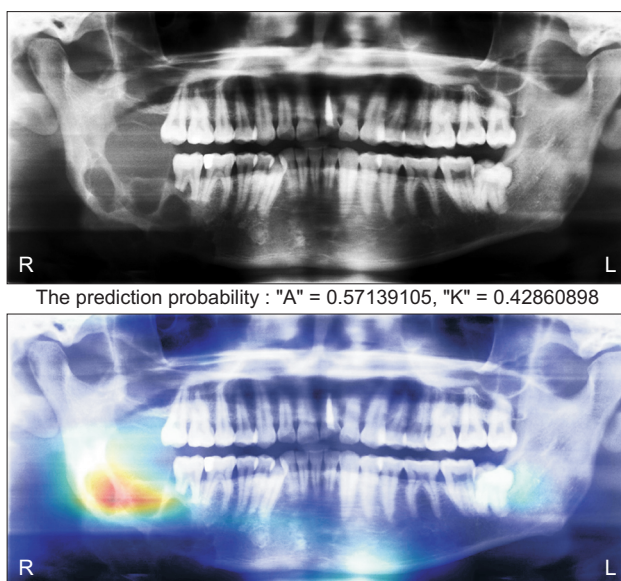

The prediction probability : "A" = 0.57139105, "K" = 0.42860898

Figure 2. Patient with multilocular cystic radiolucency at the right angle of the mandible. The model correctly classifies the ameloblastoma (probability = 0.57) and labels the correct location.
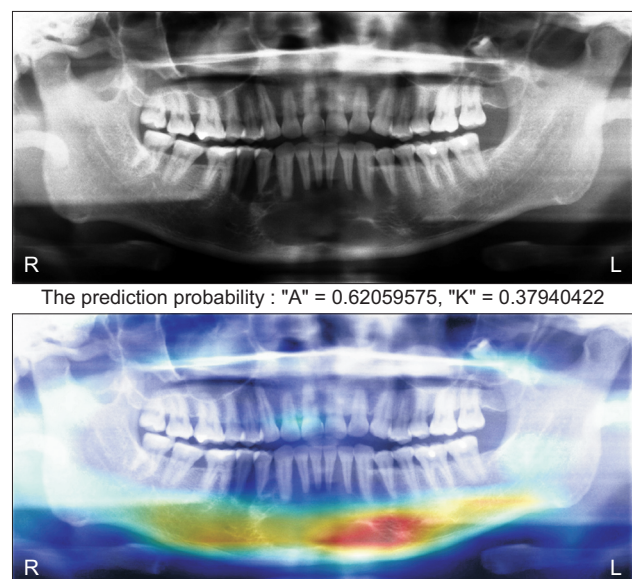

The prediction probability : "A" = 0.62059575, "K" = 0.37940422

Figure 3. Patient with a unilocular cystic radiolucency at the anterior part of the mandible. The model correctly classifies the ameloblastoma (probability = 0.62) and labels the correct location.
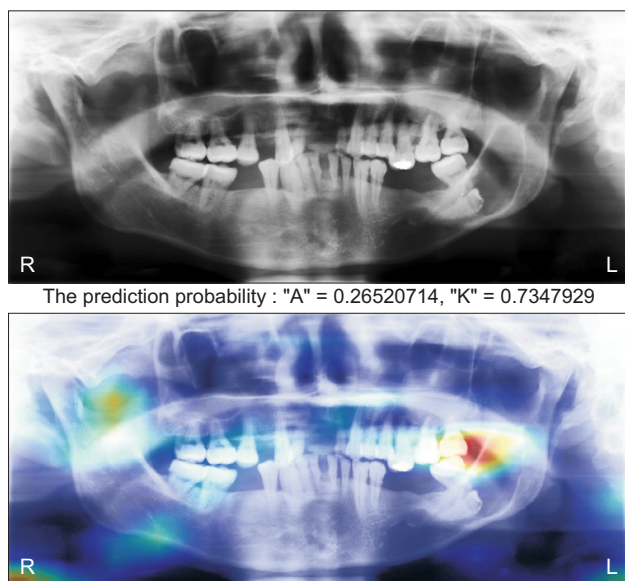
The prediction probability : "A" = 0.26520714, "K" = 0.7347929

Figure 4. Patient with a unilocular cystic radiolucency at the left angle of the mandible. The model correctly classifies the keratocystic odontogenic tumors (probability = 0.73) and labels the correct location.

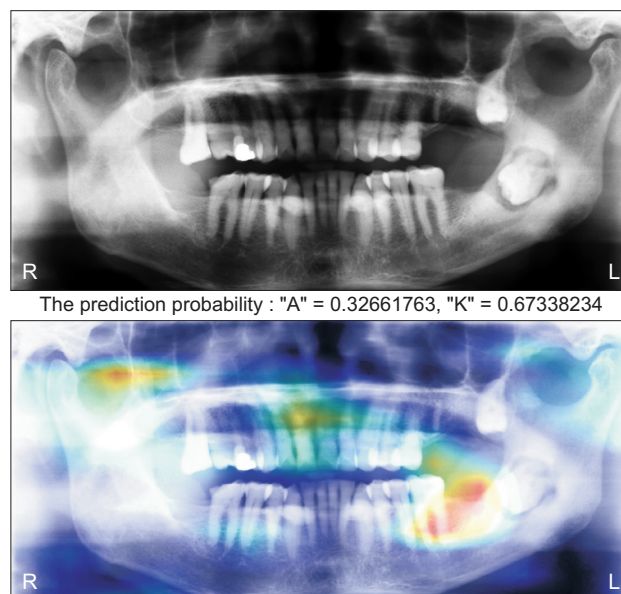The prediction probability : "A" = 0.32661763, "K" = 0.67338234

Figure 5. Patient with a unilocular cystic radiolucency at the left angle of the mandible. The model correctly classifies the keratocystic odontogenic tumors (probability = 0.67) and labels the correct location.

IV. Discussion

In recent years, deep learning algorithms have been used in many areas for classification tasks. CNNs have been used to classify types of diabetics from retinography, tumor types, and malignancies on human skin because CNN has a reputation for good performance in classification tasks on images. Given the need to optimize an enormous number of weighting parameters in CNNs, most deep learning algorithms require balanced and large amounts of data. Unfortunately, this does not apply to problems in medical informatics [12]. Generally, the data collection process in biomedical research is complex and expensive; thus, the size of biomedical datasets is limited. Moreover, the class distributions are unequal in nature, with one instance of a class significantly greater than instances of other classes. In reality, there is much less data available from treatment or experiment groups than from normal or control groups. There is very limited evidence of negative outcomes of drug or intervention studies. Most evidence is difficult to publish due to data protection restrictions and ethical concerns, resulting in further imbalance in the data distribution [13].

Transfer learning is one promising algorithmic modification technique that overcomes the problems of limited and unbalanced data. The technique is simple but effective, using an unsupervised pre-training mechanism of larger datasets. Unsupervised pre-training learns and produces model representation for each diagnostic class and generates a more

specific result [14]. In particular, transfer learning that is pre-trained using a large amount of data from similar but different areas and fine-tuned with a smaller amount of real data has generated promising results [15]. A previous study that used transfer learning from an ImageNet database and fine-tuned with chest X-ray images was highly accurate in identifying chest pathologies [16].

Our deep learning model is a VGG-16 CNN that accepts a panoramic radiograph and generates the probability outputs of ameloblastoma and KCOT. To overcome the imbalance of data entry, we focused our study on 2 jaw tumors with equal distributions of input data. To overcome the inadequacy of data, we use transfer learning by pre-training with the large amount of input data available on the Internet. We chose VGG-16 after considering its powerful results in learning ImageNet; and the algorithms are relatively simple compared to other CNNs [10].

We have identified two limitations of this study. First, the oral and maxillofacial specialists and the deep learning model were presented with only frontal X-ray radiographs. However, it has been suggested that during diagnosis specialists require lateral view X-rays in up to 15% of cases [17]. We therefore expect this setup to provide a conservative assessment of performance. Second, neither the model nor the oral and maxillofacial specialists were allowed to use the medical histories of patients, which has been shown to affect radiological diagnostics in the interpretation of X-rays [18].

Currently, our work is a standalone application of deep

learning for digital dental panoramic data. The ultimate goal is to integrate the application with the hospital information system. In this direction, the state-of-the-art deep learning approaches must be improved in terms of data integration, interoperability, and security before they can be applied effectively in the clinical domain.

In summary, the accuracy of the developed CNN was similar to that of oral and maxillofacial specialists in diagnosing ameloblastomas and KCOTs based on digital dental panoramic images. CNN can assist screening for ameloblastomas and KCOTs in a much shorter time and help to reduce the workload of oral and maxillofacial surgeons. Further investigation should be carried out to further validate and improve CNN so that it can be used widely for such screening and diagnostic applications.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. Apajalahti S, Kelppe J, Kontio R, Hagstrom J. Imaging characteristics of ameloblastomas and diagnostic value of computed tomography and magnetic resonance imaging in a series of 26 patients. Oral Surg Oral Med Oral Pathol Oral Radiol 2015;120(2):e118-30.

2. Jaeger F, de Noronha MS, Silva ML, Amaral MB, Grossmann SM, Horta MC, et al. Prevalence profile of odontogenic cysts and tumors on Brazilian sample after the reclassification of odontogenic keratocyst. J Craniomaxillofac Surg 2017;45(2):267-70.

3. Ariji Y, Morita M, Katsumata A, Sugita Y, Naitoh M, Goto M, et al. Imaging features contributing to the diagnosis of ameloblastomas and keratocystic odontogenic tumours: logistic regression analysis. Dentomaxillofac Radiol 2011;40(3):133-40.

4. Hayashi K, Tozaki M, Sugisaki M, Yoshida N, Fukuda K, Tanabe H. Dynamic multislice helical CT of ameloblastoma and odontogenic keratocyst: correlation between contrast enhancement and angiogenesis. J Comput Assist Tomogr 2002;26(6):922-6.

5. Minami M, Kaneda T, Ozawa K, Yamamoto H, Itai Y, Ozawa M, et al. Cystic lesions of the maxillomandibular region: MR imaging distinction of odontogenic keratocysts and ameloblastomas from other cysts. AJR Am J Roentgenol 1996;166(4):943-9.

6. Min S, Lee B, Yoon S. Deep learning in bioinformatics. Brief Bioinform 2017;18(5):851-69.

7. Shichijo S, Nomura S, Aoyama K, Nishikawa Y, Miura M, Shinagawa T, et al. Application of convolutional neural networks in the diagnosis of helicobacter pylori infection based on endoscopic images. EBioMedicine 2017;25:106-11.

8. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35(5):1285-98.

9. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [Internet]. Ithaca (NY): arXiv.org; c2015 [cited at 2018 Jul 15]. Available from: https://arxiv.org/pdf/1409.1556.pdf.

10. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis 2015;115(3):211-52.

11. Sturm I, Lapuschkin S, Samek W, Muller KR. Interpretable deep neural networks for single-trial EEG classification. J Neurosci Methods 2016;274:141-5.

12. Oh S, Lee MS, Zhang BT. Ensemble learning with active example selection for imbalanced biomedical data classification. IEEE/ACM Trans Comput Biol Bioinform 2011;8(2):316-25.

13. Malin BA, Emam KE, O'Keefe CM. Biomedical data privacy: problems, perspectives, and recent advances. J Am Med Inform Assoc 2013;20(1):2-6.

14. Erhan D, Bengio Y, Courville A, Manzagol PA, Vincent P, Bengio S. Why does unsupervised pre-training help deep learning? J Mach Learn Res 2010;11:625-60.

15. Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 2016;35(5):1153-9.

16. Bar Y, Diamant I, Wolf L, Greenspan H. Deep learning with non-medical training used for chest pathology identification. In: Medical imaging 2015: computer-aided diagnosis (Proceedings of SPIE 9414). Bellingham (WA): International Society for Optics and Photonics; 2015.

17. Raoof S, Feigin D, Sung A, Raoof S, Irugulpati L, Rosenow EC 3rd. Interpretation of plain chest roentgenogram. Chest 2012;141(2):545-58.

18. Berbaum K, Franken EA Jr, Smith WL. The effect of comparison films upon resident interpretation of pediatric chest radiographs. Invest Radiol 1985;20(2):124-8.