

RESEARCH ARTICLE

Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network

¹Odeuk Kwon, ²Tae-Hoon Yong, ²Se-Ryong Kang, ³Jo-Eun Kim, ⁴Kyung-Hoe Huh, ⁴Min-Suk Heo, ⁴Sam-Sun Lee, ⁴Soon-Chul Choi and ^{2,4}Won-Jin Yi

¹Department of Oral and Maxillofacial Radiology, School of Dentistry, Seoul National University, Seoul, South Korea;

²Department of Biomedical Radiation Sciences, Graduate School of Convergence Science and Technology, Seoul National

University, Seoul, South Korea; ³Department of Oral and Maxillofacial Radiology, Seoul National University Dental Hospital,

Seoul, South Korea; ⁴Department of Oral and Maxillofacial Radiology, School of Dentistry and Dental Research Institute, BK21, Seoul National University, Seoul, South Korea

Objectives: The purpose of this study was to automatically diagnose odontogenic cysts and tumors of both jaws on panoramic radiographs using deep learning. We proposed a novel framework of deep convolution neural network (CNN) with data augmentation for detection and classification of the multiple diseases.

Methods: We developed a deep CNN modified from YOLOv3 for detecting and classifying odontogenic cysts and tumors of both jaws. Our data set of 1282 panoramic radiographs comprised 350 dentigerous cysts (DCs), 302 periapical cysts (PCs), 300 odontogenic keratocysts (OKCs), 230 ameloblastomas (ABs), and 100 normal jaws with no disease. In addition, the number of radiographs was augmented 12-fold by flip, rotation, and intensity changes. We evaluated the classification performance of the developed CNN by calculating sensitivity, specificity, accuracy, and area under the curve (AUC) for diseases of both jaws.

Results: The overall classification performance for the diseases improved from 78.2% sensitivity, 93.9% specificity, 91.3% accuracy, and 0.86 AUC using the CNN with unaugmented data set to 88.9% sensitivity, 97.2% specificity, 95.6% accuracy, and 0.94 AUC using the CNN with augmented data set. CNN using augmented data set had the following sensitivities, specificities, accuracies, and AUCs: 91.4%, 99.2%, 97.8%, and 0.96 for DCs, 82.8%, 99.2%, 96.2%, and 0.92 for PCs, 98.4%, 92.3%, 94.0%, and 0.97 for OKCs, 71.7%, 100%, 94.3%, and 0.86 for ABs, and 100.0%, 95.1%, 96.0%, and 0.97 for normal jaws, respectively.

Conclusion: The CNN method we developed for automatically diagnosing odontogenic cysts and tumors of both jaws on panoramic radiographs using data augmentation showed high sensitivity, specificity, accuracy, and AUC despite the limited number of panoramic images involved.

Dentomaxillofacial Radiology (2020) 49, 20200185. doi: [10.1259/dmfr.20200185](https://doi.org/10.1259/dmfr.20200185)

Cite this article as: Kwon O, Yong T-H, Kang S-R, Kim J-E, Huh K-H, Heo M-S, et al. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. *Dentomaxillofac Radiol* 2020; 49: 20200185.

Keywords: Automatic diagnosis; odontogenic cysts and tumors; convolution neural network (CNN); panoramic radiograph; data augmentation

Introduction

Computer-aided diagnosis (CAD) has been utilized to identify cavities and periodontitis lesions as well as maxillary sinusitis, osteoporosis, and other pathologies in the oral and maxillofacial field.¹ It can provide dental professionals with a valuable second opinion by detecting and classifying pathological changes automatically. Conventional CAD systems require extraction of the most significant features before training to successfully recognize or classify images, but feature extraction is a difficult and time-consuming task. A recent method based on deep learning, a subset of machine learning, can overcome this limitation by automatically extracting relevant features during training, and uses the whole image directly without best-feature representation.²⁻⁴ Deep learning-based methods have been used extensively to solve complex problems in radiology.⁵ A deep convolutional neural network (CNN), a type of deep learning, is the most commonly used method for organ segmentation^{6,7} as well as classification^{8,9} and detection^{10,11} of organs and related diseases in medical imaging. Various attempts have been made to determine specific characteristics of target regions intended for detection and classification.¹² Research into the applications of CAD using deep CNNs has been expanding rapidly, and is expected to produce more accurate diagnoses at faster rates.¹

However, application of deep learning in the field of oral and maxillofacial imaging has been limited to detection of landmarks in cephalograms,¹³ detection and classification of teeth,¹⁴⁻¹⁶ diagnosis of cavities,¹⁷⁻²¹ diagnosis of periodontitis,²² and detection of maxillary sinusitis.²³ One study aimed to achieve automatic segmentation of all teeth,¹⁴ while other research used deep learning to classify the root morphology of the mandibular first molar,²¹ convert a two-dimensional panoramic image to a three-dimensional one,²⁴ and diagnose osteoporosis in panoramic radiographs.^{25,26}

To date, few studies have used deep learning to detect and classify radiolucent lesions in the jaw. A deep learning object detection technique was used for automatic detection and classification of radiolucent lesions in the mandible,²⁷ and for differential diagnosis between ameloblastomas and odontogenic keratocysts of the jaw on panoramic radiographs.²⁸ Precise pre-operative diagnosis of these tumors and cysts of the jaw can help oral and maxillofacial surgeons plan appropriate treatment, but this is more difficult in the maxilla than in the mandible because of superimposition of the normal structures in the maxilla.²⁹ Nonetheless, no study to date has examined the functionality of deep CNN for automatic diagnosis of odontogenic cysts and tumors occurring in both jaws using panoramic radiographs. Therefore, the purpose of this study was to automatically diagnose odontogenic cysts and tumors of both jaws with the highest rate of occurrence in the oral and maxillofacial regions. We proposed a novel framework

of deep CNN with data augmentation for detection and classification of the multiple diseases.

Methods and materials

Data preparation and augmentation of panoramic radiographs

We prepared a total of 1282 panoramic radiographs of patients who visited Seoul National University Dental Hospital from 1999 to 2017. These radiographs included 350 dentigerous cysts (273 in the mandible and 77 in the maxilla), 302 periapical cysts (123 in the mandible and 179 in the maxilla), 300 odontogenic keratocysts (266 in the mandible and 34 in the maxilla), and 230 ameloblastomas (222 in the mandible and 8 in the maxilla). As a control group, we also prepared 100 normal panoramic radiographs. Panoramic radiographs were obtained from adult patients without mixed dentition, and only one radiograph was used per patient. The study was approved by the Institutional Review Board (IRB) of Seoul National University Dental Hospital (ERI18001).

The cysts and tumors were classified finally based on the histopathological diagnosis from biopsies in addition to clinical diagnosis, and we selected panoramic radiographs with the same radiologic diagnosis by two radiologists with more than 15 years of experience. In other words, we used only the panoramic radiographs with identical radiological and histopathological diagnoses. Panoramic radiographs had a resolution of 1976×976 pixels, and each radiograph was labeled manually by drawing rectangular bounding boxes around the lesions using ImageJ³⁰ for training. The bounding boxes included radiographic characteristics of each disease such as cortical margin and internal radiolucent lesions. Average sizes of the annotated lesions were 178×196 pixels for dentigerous cysts, 186×170 pixels for periapical cysts, 297×304 pixels for odontogenic keratocysts, and 386×351 pixels for ameloblastomas. We applied a contrast-limited adaptive histogram equalization (CLAHE) technique to the images to expand the high values and compress the values in the dark layer more effectively.²⁸

Because we had a limited number of panoramic images, we performed data augmentation to increase the amount of training data for deep learning. Using MATLAB (MathWorks, Natick, MA), images were flipped horizontally, rotated by 1° and -1° , and their grayscale was transformed by γ correction ($\gamma = 0.8, 1, 1.2$). As a result, the number of images was increased 12-fold. Images were randomly separated into a training data set (80%) and a test data set (20%) before data augmentation. The training data set for dentigerous cyst, periapical cyst, odontogenic keratocyst, and ameloblastoma comprised 280, 242, 240 and 184 panoramic images, respectively, and test data set, 70, 60, 60, and 46,

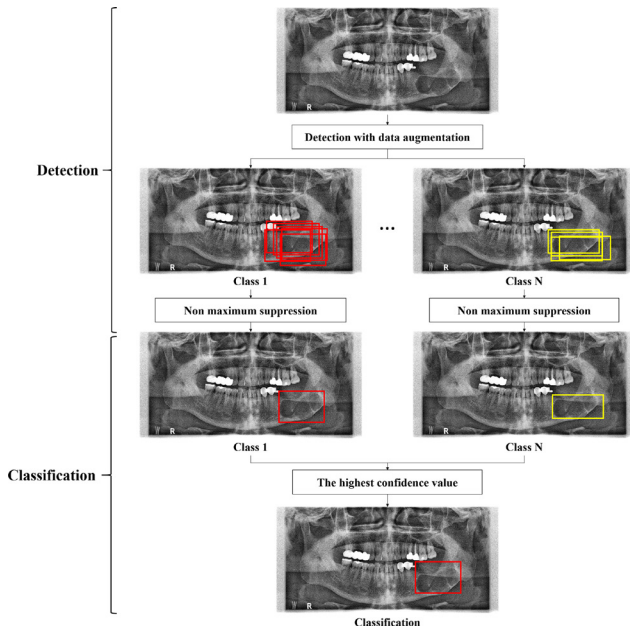


Figure 1 A framework of the deep CNN for detection and classification of multiple diseases. CNN, convolutional neural network.

respectively. The training data set was used to train the CNN, and the test data set was used to evaluate the final trained model.

A deep CNN model for detection and classification of multiple diseases

We developed a modified CNN from the YOLOv3 based on the Darknet-53 network for detecting and classifying multiple diseases on panoramic radiographs (Figure 1). YOLOv3 predicted bounding boxes at three different scales using features from scales extracted using a similar concept to feature pyramid networks

(FPNs).³¹ In the last of these layers, the CNN replaced the softmax function with independent logistic classifiers to calculate the objectness of the input belonging to a specific class.³² The CNN predicted the bounding box, confidence (objectness), and class predictions (Figure 1). Each bounding box had a confidence value calculated from the logistic regression. We modified the input resolution from 320×320 pixels to 608×608 pixels, which made it possible to predict lesions of smaller pixel sizes. The location of skip connection was also changed from the convolution layers of $76 \times 76 \times 256$ to $76 \times 76 \times 256$ to those of $152 \times 152 \times 128$ to $152 \times 152 \times 128$ in order to use four times upsampling instead of two times for reducing the losses of spatial information (Figure 2). The network was trained on a total of 8000 epochs with a 64-batch size and one- or two-stride size by using pre-prepared augmented images. An adaptive moment estimation solver was used to optimize the network with a learning rate of 0.001 and momentum of 0.9.

After training, the network outputted bounding boxes and the confidence that the bounding box enclosed a lesion for input panoramic images. Intersection over union (IoU) score, ratio of the area of intersection and area of union of the predicted bounding box and ground truth bounding box, was computed in the final set of outputs. We used 0.5 as the IoU threshold value to obtain the average precision (AP) score for disease detection (Figure 1). The network could detect one or more bounding boxes of possible classes for one lesion at the input panoramic image. If multiple bounding boxes of the same class were generated for one lesion, only one bounding box among the boxes was predicted for the class by non-maximum suppression. Finally, one bounding box of the highest confidence among the multiple predictions was adopted for disease classification (Figure 1).

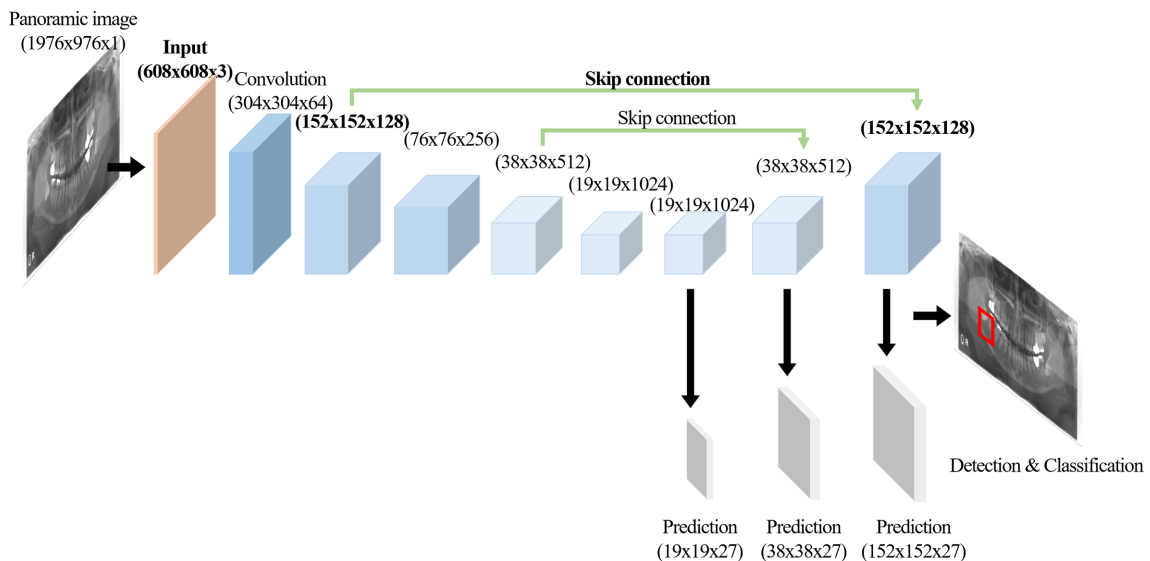


Figure 2 A CNN architecture modified from YOLOv3 with the modified layers in Bold. CNN, convolutional neural network.

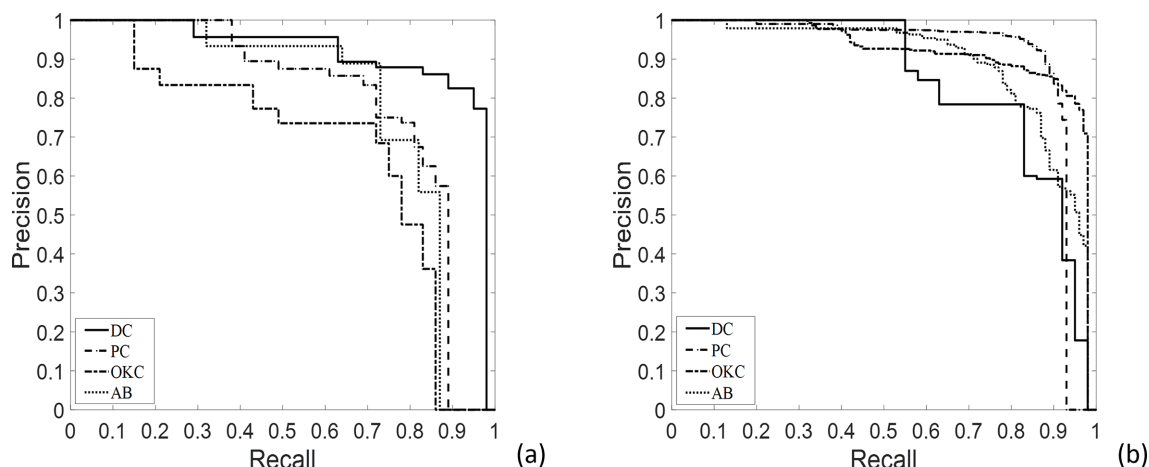


Figure 3 PR curves from automatic detection of dentigerous cysts, periapical cysts, odontogenic keratocysts, and ameloblastomas without (a) and with (b) data set augmentation. AB, ameloblastoma; DCs, dentigerous cysts; OKCs, odontogenic keratocysts; PCs, periapical cysts.

Evaluation of detection and classification performance of the deep CNN model

We evaluated the performance of the developed model using a test data set not used for training. For detection performance, we calculated recall (sensitivity) ($TP/(FN + TP)$), precision (positive predictive value) ($TP/(FP + TP)$), F1-score ($2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$), and average precision (AP) values, which are common parameters in object detection.^{33,34} The precision-recall curve (PRC) was also computed from the model's detection output by varying the confidence threshold that determined what was counted as a model-predicted positive detection.³³ AP was calculated as the average value of precision across all recall values. We calculated a confusion matrix to evaluate the classification performance of the CNN for diseases.³⁵ For classification performance, we calculated sensitivity ($TP/(FN + TP)$), specificity ($TN/(TN + FP)$), and accuracy ($(TN + TP)/(TN + TP + FN + FP)$) from the confusion matrix. The receiver operating characteristic curve (ROC) was also computed from the model's classification output by varying the confidence threshold for each disease.³⁶ We calculated the value of the area under the ROC curve (AUC).

Results

We measured AP, precision, recall, and F1-score values to evaluate model performance for lesion detection with

or without data augmentation (Table 1). AP values for dentigerous cysts (DCs), periapical cysts (PCs), odontogenic keratocysts (OKCs), and ameloblastoma (ABs) using non-augmented dataset were 0.91, 0.79, 0.67, and 0.78, respectively, compared to 0.84, 0.89, 0.91, and 0.88, for augmented data set, respectively (Table 1). Mean AP, precision, recall, and F1-score for DCs, PCs, OKCs, and ABs were 0.79, 0.78, 0.74, and 0.76 for non-augmented data set and 0.88, 0.87, 0.83, and 0.85 for augmented data set, respectively (Table 1). The precision-recall (PR) curves for detecting lesions are shown in Figure 3.

Table 2 shows the confusion matrix for classifying DCs, PCs, OKCs, ABs, and normal jaws using the developed CNN with or without data augmentation. To assess the classification performance of the CNN for absence of diseases, panoramic radiographs with no diseases were designated normal. Table 3 summarizes the classification results of sensitivity, sensitivity, accuracy, and AUC for each disease. Figure 4 shows the correctly and falsely classified lesions with and without data set augmentation. The receiver operating characteristic (ROC) curves for classifying multiple diseases are shown in Figure 5. For DCs, lesions correctly classified by the CNN had a crown, radiolucent lesion, and surrounding cortical bony margin both without (Figure 4 (a)) and with data augmentation (Figure 4(e)). Sensitivity, specificity, accuracy, and AUC for DCs without augmentation were 87.1%, 94.7%, 94.1% and 0.91, respectively, compared

Table 1 AP, mean AP, precision (positive predictive value), recall (sensitivity), and F1-score for detecting DC, PC, OKC, and AB of the developed CNN with or without data set augmentation

	AP				Mean AP	Precision	Recall	F1-score
	DC	PC	OKC	AB				
With augmentation	0.84	0.89	0.91	0.88	0.88 ± 0.04	0.87	0.83	0.85
Without augmentation	0.91	0.79	0.67	0.78	0.79 ± 0.12	0.78	0.74	0.76

AB, ameloblastoma; AP, average precision; DC, dentigerous cyst; OKC, odontogenic keratocyst; PC, periapical cyst.

Table 2 Confusion matrix for classifying DC, PC, OKC, AB, and normal jaws (Normal) by the developed CNN, with or without data set augmentation

Input	Without augmentation					With augmentation				
	DC	PC	OKC	AB	Normal	DC	PC	OKC	AB	Normal
DC	0.87	0.05	0.03	0.00	0.05	0.92	0.03	0.03	0.00	0.03
PC	0.06	0.72	0.08	0.00	0.14	0.03	0.84	0.04	0.00	0.10
OKC	0.06	0.06	0.78	0.03	0.06	0.00	0.00	0.98	0.00	0.02
AB	0.05	0.05	0.27	0.55	0.09	0.00	0.00	0.22	0.72	0.06
Normal	0.00	0.00	0.01	0.00	0.99	0.00	0.00	0.00	0.00	1.00

to 91.4%, 99.2%, 97.8%, and 0.96, respectively, with augmentation. For PCs, correctly classified lesions had a root apex, radiolucent lesion, and surrounding cortical bony lesion both without (Figure 4 (b)) and with data augmentation (Figure 4 (f)). Sensitivity, specificity, accuracy, and AUC for PCs were 71.9%, 95.0%, 91.2%, and 0.84, respectively, without augmentation, and 82.8%, 99.2%, 96.2%, and 0.92, respectively, with augmentation. For OKCs, correctly classified lesions were unilocular radiolucent cystic lesions associated with the third molar on the mandible without augmentation (Figure 4 (c)) and large radiolucent cystic lesions extending from the border of the maxilla with augmentation (Figure 4(g)). Sensitivity, specificity, accuracy, and AUC for OKCs without augmentation were 78.9%, 89.4%, 87.8%, and 0.83, respectively, compared to 98.4%, 92.3%, 94.0%, and 0.97, respectively, with augmentation. ABs presented as large unilocular lesions on the maxilla's anterior side without augmentation (Figure 4(d)), and radiolucent lesions including the crown of the third molar with teeth root resorption on the mandible with augmentation (Figure 4(h)). Sensitivity, specificity, accuracy, and AUC of the CNN for ABs was 54.3%, 99.1%, 90.3%, and 0.77, respectively, without augmentation, and 71.7%, 100%, 94.3%, and 0.86, respectively, with augmentation. Lastly, sensitivity, specificity, accuracy, and AUC for normal jaws was 99.0%, 91.1%, 92.9% and 0.94, respectively, without augmentation, and 100%, 95.1%, 96.0% and 0.97, respectively, with augmentation. Mean sensitivity, specificity, accuracy,

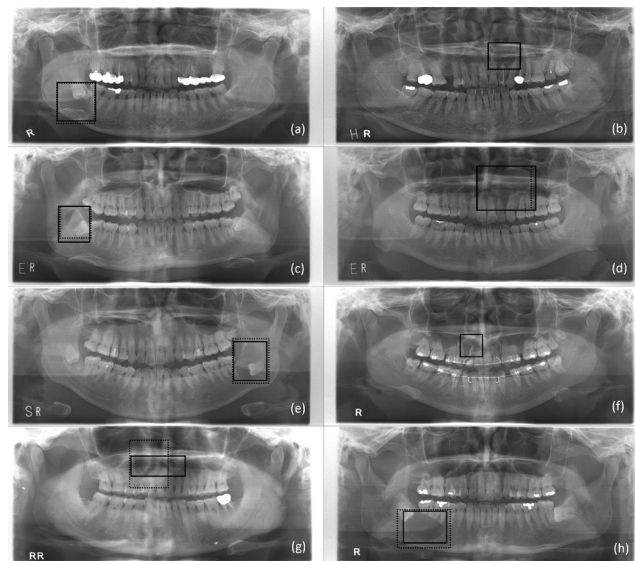


Figure 4 Lesions annotated by the radiologist (solid line) and correctly classified by the developed CNN model (dotted line) as adentigerous cyst (a), periapical cyst (b), odontogenic keratocyst (c), and ameloblastoma (d) without data set augmentation. A correctly classified (solid line) dentigerous cyst (e), periapical cyst (f), odontogenic keratocyst (g), and ameloblastoma (h) with data set augmentation, and a falsely classified (dotted line) odontogenic keratocyst (e), none (f), dentigerous cyst (g), and dentigerous cyst (h) without data set augmentation. CNN, convolutional neural network.

and AUC for all diseases were 78.2%, 93.9%, 91.3%, and 0.86, respectively, without augmentation, and 88.9%, 97.2%, 95.6%, and 0.94, respectively, with augmentation.

Discussion

A deep CNN, a type of deep learning model, automatically and adaptively learns spatial hierarchies of image features by using multiple building blocks of convolution layers, pooling layers, and fully connected layers.²⁻⁴ Various imaging modalities, such as periapical, cephalometric, panoramic radiographs, as well as CBCT

Table 3 Sensitivity, specificity, accuracy, and AUC for classifying DC, PC, OKC, AB, and normal jaws (Normal) by the developed CNN, with or without data set augmentation

	Without augmentation				With augmentation			
	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
DC	87.1	94.7	94.1	0.91	91.4	99.2	97.8	0.96
PC	71.9	95.0	91.2	0.84	82.8	99.2	96.2	0.92
OKC	78.9	89.4	87.8	0.83	98.4	92.3	94.0	0.97
AB	54.3	99.1	90.3	0.77	71.7	100.0	94.3	0.86
Normal	99.0	91.1	92.9	0.94	100.0	95.1	96.0	0.97
Mean	78.2 ± 16.8	93.9 ± 3.8	91.3 ± 2.4	0.86 ± 0.07	88.9 ± 11.8	97.2 ± 3.3	95.6 ± 1.5	0.94 ± 0.05

AUC, area under the curve.

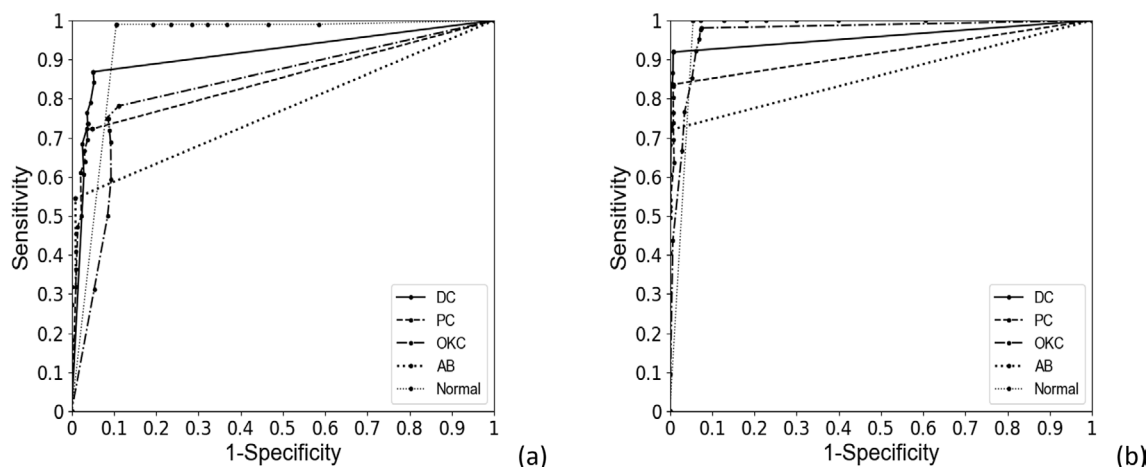


Figure 5 ROC curves from automatic classification of dentigerous cysts, periapical cysts, odontogenic keratocysts, ameloblastomas, and normal jaws without (a) and with (b) data set augmentation. ROC, receiver operating characteristic.

have been used to detect and classify diseases in CAD studies using a deep CNN. One study involved segmentation of teeth on panoramic images.¹⁶ Another evaluated the root morphology of the mandibular first molar on panoramic radiography.²¹ Sinusitis was detected by learning the form of the maxillary sinus on panoramic images.²³ Osteoporosis has also been diagnosed by evaluating cortical erosion of the mandibular inferior cortex from panoramic images.²⁵ However, there are few CNN-based CAD studies of radiolucent jaw lesions on panoramic radiographs.

We developed a modified CNN from YOLOv3 for automatically detecting and classifying odontogenic cysts and tumors of both jaws on panoramic images, as these types of cysts and tumors have the highest rate of occurrence in the oral and maxillofacial region. We increased the input resolution of the panoramic image, and changed the skip connections compared with the original YOLOv3, which could reduce information loss during training, and increase the performance the CNN. The developed CNN based on YOLOv3 showed overall detection performance of 0.88 of AP, 0.87 of precision, and 0.83 of recall. Classification sensitivity was 72% for ABs, 98% for OKCs, 91% for DCs, and 83% for PCs using augmented data set. A study reported automatic classification of ABs, OKCs, DCs, RCs, and SBCs using DIGITS, a pretrained deep learning model.²⁷ In the study, the total number of training data set was 210, and the number of test data set was 3 ameloblastomas, 6 odontogenic keratocysts, 8 dentigerous cysts, and 7 radicular cysts to evaluate the performance.²⁷ As the result, classification sensitivity was 60% for ABs, 13% for OKCs, 82% for DCs, and 77% for RCs.²⁷ On the other hand, we used a total of 1282 panoramic radiographs and data augmentation of 12 times. Generally, our CNN model showed higher sensitivities for these diseases than the model of Arijji *et al*, most likely because these authors used a smaller number of cysts and tumor images than we did for training and evaluation. In addition to this,

we classified odontogenic cysts and tumors in both the maxilla and mandible, while they only classified cysts and tumors in the mandible.²⁷ Another study reported binary classification of ABs and OKCs using a VGG-16 CNN, a pre-trained network using ImageNet. The number of original training data set and test data set was 200 and 50, respectively, for each disease. They used data augmentation of double using only horizontal flips for training data set, and the sensitivity and specificity for differential diagnosis was 81.8%, and 83.3%, respectively.²⁸ In our study, the sensitivity and specificity for ABs was 71.7%, and 100%, respectively, while that for OKCs was 98.4%, and 92.3%, respectively. Overall sensitivity was 88.9% and overall specificity was 97.2% using data augmentation in spite of multiple classification for four diseases. Generally, the deep learning model we developed based on the state-of-the-art network YOLOv3 and augmented data set had higher sensitivity and specificity than those of previous models in classifying cysts and tumors of both jaws on panoramic images.

Odontogenic cysts and tumors classified in this study share some radiologic features, thus differential diagnosis may be difficult.³⁷ It is common for DCs to wrap around the crown symmetrically, and it should not come into contact with the root as it originates from the cemento-enamel junction of the crown. If a cyst arising from the crown of an unerupted tooth extends towards the periapical side over the cemento-enamel junction of the unerupted tooth, then it is likely an OKC.³⁸ Separately, multilocular OKCs with scalloped borders may appear similar to ABs, but they tend to grow along the marrow space without significant cortical bone expansion compared to ABs, which usually show significant cortical expansion with well-defined cortical borders in the mandible and ill-defined margins in the maxilla.³⁹ It is more difficult to diagnose these radiolucent lesions correctly in the maxilla than in the mandible because of superimposition of structures such as the nasal cavity, maxillary sinus, hard palate, and inferior nasal concha in the maxilla.²⁹

Our CNN had 88.9% sensitivity for classification using augmented data set. It had a 100% sensitivity (false negative rate of 0) for classification of disease-free (normal) jaws. The sensitivity for OKCs was the highest at 98.4% (lowest false negative rate of 0.02) among the four diseases; the false negative was classified as normal. In contrast, the lowest sensitivity of 71.7% (highest false negative rate of 0.28) was obtained for ABs among the four diseases, with false diagnosis of ABs as OKCs (0.22) and normal (0.06). Fewer images were used for ABs learning than for the other three diseases, despite their heterogeneous radiographic appearance (multi- and unilocular). The second highest sensitivity of 91.4% (false negative rate of 0.09) was obtained for DCs among the four diseases, and DCs were falsely diagnosed as PCs (0.03), OKCs (0.03), or normal (0.03). DC lesions involving the third molar in the maxilla, those associated with the supernumerary tooth in the anterior maxilla, or those that overlapped with the wall of the maxillary sinus or anterior teeth were misdiagnosed. The sensitivity for PCs was 82.8% (false negative rate of 0.17). Our model failed to diagnose smaller lesions in the root apex of the tooth overlapping with other anatomical structures (0.1), and misdiagnosed them as DCs (0.03) or OKCs (0.04), especially in the maxilla. As a result, sensitivities for DCs and PCs were lower than that for OKCs as DCs, and PCs occur more commonly in the maxilla. Generally, sensitivity (false negative rate) for lesions located in the maxilla was lower (higher) than that for lesions in the mandible due to the overlap between lesions and other anatomical structures in the maxilla.

The classification specificity of 97.2% when using augmented data set was higher than the sensitivity of 88.9% for all diseases. Specificity for ABs was the highest (lowest false positive rate) at 100% (0.00); no diseases were falsely diagnosed as ABs. Specificity for DCs and PCs was also high at 99.2%. Specificity for OKCs was lowest at 92.3%. ABs were frequently misdiagnosed as OKCs, and the classification was least sensitive for ABs. This might be due to the smaller number of ABs used for learning compared with all the other cystic diseases in addition to the large variations in radiological appearance of ABs. Rectangular bounding boxes around the lesions should have included more normal anatomical structures to cover the larger boundary of ABs. As a result, our model showed poorer generalizability for ABs than the other cystic diseases. Nonetheless, our model showed high sensitivity for OKCs and high specificity for ABs. It is important to note that differential diagnosis between OKCs and ABs based on panoramic radiographs can be difficult even for experienced radiologists.

Overfitting of the training CNN model, which results in the model learning statistical regularity specific to the training data set, negatively impacts the model's ability to generalize to a new data set.² One solution to reduce overfitting is data augmentation, which is the process of modifying data through flipping, moving, cropping, rotating, and grayscale transformation.⁴⁰ Data augmentation can improve the performance of the CNN in medical image

analysis.⁴⁰ In this study, augmented technology increased the number of images in the dataset by 12-fold. Mean AP, precision, recall, and F1-score values for detection performance of the CNN generally increased after data augmentation. As the recall value increased, the precision value decreased more gradually with data augmentation than without data augmentation, which resulted in a higher AP with data augmentation than without data augmentation. Classification performance of the CNN improved from 78.2% sensitivity, 93.9% specificity, 91.3% accuracy, and 0.86 AUC to 88.9% sensitivity, 97.2% specificity, 95.6% accuracy, and 0.94 AUC for the four diseases when augmented data set was used. As the false positive rates increased, the sensitivity increased more steeply with data augmentation than without data augmentation, which resulted in a higher AUC with data augmentation than without data augmentation. In particular, the performance for diagnosing OKCs improved from 78.9% sensitivity, 89.4% specificity, 87.8% accuracy, and 0.83 AUC to 98.4% sensitivity, 92.3% specificity, 94.0% accuracy, and 0.97 AUC, and the performance for ABs from 54.3% sensitivity, 99.1% specificity, 90.3% accuracy, and 0.77 AUC to 71.7% sensitivity, 100% specificity, 94.3% accuracy, and 0.86 AUC.

To date, oral and maxillofacial medicine has benefited little from advancements in CAD. Here, we developed a CNN model to automatically detect and classify cysts and tumors of both jaws on panoramic radiographs with high accuracy using data augmentation. The developed CAD can help general dental clinicians diagnose and treat patients more efficiently, and thus improve diagnostic performance and patient care. In future studies, the use of a more complex annotation method, such as lesion segmentation rather than a rectangular bounding box, and the use of more images for training, especially images of the maxilla, will increase the classification performance of the CNN for diseases of both jaws.

Conclusion

The deep CNN method we developed for automatically diagnosing odontogenic cysts and tumors in both jaws on panoramic radiographs showed high sensitivity, specificity, accuracy, and AUC when using augmented data, despite the limited number of panoramic images available.

Acknowledgment

This work was funded by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2C2008365), and this work was also supported by the Technology Innovation Program (10063389) funded by the Ministry of Trade, Industry & Energy (MOTIE), Korea.

REFERENCES

- Wang C-W, Huang C-T, Lee J-H, Li C-H, Chang S-W, Siao M-J, et al. A benchmark for comparison of dental radiography analysis algorithms. *Med Image Anal* 2016; **31**: 63–76. doi: <https://doi.org/10.1016/j.media.2016.02.004>
- Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018; **9**: 611–29. doi: <https://doi.org/10.1007/s13244-018-0639-9>
- Yonekura A, Kawanaka H, Prasath VBS, Aronow BJ, Takase H. Automatic disease stage classification of glioblastoma multiforme histopathological images using deep convolutional neural network. *Biomed Eng Lett* 2018; **8**: 321–7. doi: <https://doi.org/10.1007/s13534-018-0077-0>
- Billah M, Waheed S. Gastrointestinal polyp detection in endoscopic images using an improved feature extraction method. *Biomed Eng Lett* 2018; **8**: 69–75. doi: <https://doi.org/10.1007/s13534-017-0048-x>
- Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015; **61**: 85–117. doi: <https://doi.org/10.1016/j.neunet.2014.09.003>
- Kallenberg M, Petersen K, Nielsen M, Ng AY, Igel C, AY N, Diao PF, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Trans Med Imaging* 2016; **35**: 1322–31. doi: <https://doi.org/10.1109/TMI.2016.2532122>
- Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med Image Anal* 2018; **43**: 98–111. doi: <https://doi.org/10.1016/j.media.2017.10.002>
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115–8. doi: <https://doi.org/10.1038/nature21056>
- Gao XW, Hui R, Tian Z. Classification of CT brain images based on deep learning networks. *Comput Methods Programs Biomed* 2017; **138**: 49–56. doi: <https://doi.org/10.1016/j.cmpb.2016.10.007>
- Teramoto A, Fujita H, Yamamuro O, Tamaki T. Automated detection of pulmonary nodules in PET/CT images: ensemble false-positive reduction using a convolutional neural network technique. *Med Phys* 2016; **43**: 2821–7. doi: <https://doi.org/10.1118/1.4948498>
- Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Publisher correction: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019; **25**: 530. doi: <https://doi.org/10.1038/s41591-019-0359-9>
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems* 2012; **1**: 1097–105.
- Lee H, Park M, Kim J. Cephalometric landmark detection in dental X-ray images using convolutional neural networks. *Medical Imaging 2017: Computer-Aided Diagnosis*. 2017; **101341W**: 10134.
- Ronneberger O, Fischer P, Brox T. Dental X-ray image segmentation using a U-shaped deep Convolutional network. *International Symposium on Biomedical Imaging* 2015;: 1–13.
- Miki Y, Muramatsu C, Hayashi T, Zhou X, Hara T, Katsumata A, et al. Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput Biol Med* 2017; **80**: 24–9. doi: <https://doi.org/10.1016/j.combiomed.2016.11.003>
- Jader G, Fontineli J, Ruiz M, Abdalla K, Pithon M, Oliveira L. Deep instance segmentation of teeth in panoramic X-ray images. *31st SIBGRAPI Conference on Graphics, Patterns and Images* 2018; **2018**: 400–7.
- Ben AR, Ridha E, Mourad Z. Detection and classification of dental caries in X-ray images using deep neural networks. *Eleventh Int Conf Softw Eng Adv* 2016;: 223–7.
- Srivastava MM, Kumar P, Pradhan L, Varadarajan S. Detection of tooth caries in Bitewing radiographs using deep learning. *arXiv e-print* 2017;. arXiv:1711.07312 p.
- Choi J, Eun H, Kim C. Boosting proximal dental caries detection via combination of variational methods and Convolutional neural network. *J Signal Process Syst* 2018; **90**: 87–97. doi: <https://doi.org/10.1007/s11265-016-1214-6>
- Lee J-H, Kim D-H, Jeong S-N, Choi S-H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 2018; **77**: 106–11. doi: <https://doi.org/10.1016/j.jdent.2018.07.015>
- Hiraiwa T, Arijji Y, Fukuda M, Kise Y, Nakata K, Katsumata A, et al. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. *Dentomaxillofac Radiol* 2019; **48**: 20180218–7. doi: <https://doi.org/10.1259/dmfr.20180218>
- Lee J-H, Kim D-H, Jeong S-N, Choi S-H. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontal Implant Sci* 2018; **48**: 114–23. doi: <https://doi.org/10.5051/jpis.2018.48.2.114>
- Murata M, Arijji Y, Ohashi Y, Kawai T, Fukuda M, Funakoshi T, et al. Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. *Oral Radiol* 2019; **35**: 301–7. doi: <https://doi.org/10.1007/s11282-018-0363-7>
- Leo LM, Simla AJ. 2D to 3D conversion of dental images using deep neural network. *Journal of Chemical and Pharmaceutical Sciences* 2017; **10**: 1432–6.
- Lee J-S, Adhikari S, Liu L, Jeong H-G, Kim H, Yoon S-J. Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study. *Dentomaxillofac Radiol* 2019; **48**: 20170344–8. doi: <https://doi.org/10.1259/dmfr.20170344>
- Hwang J-J, Jung Y-H, Cho B-H, Heo M-S. An overview of deep learning in the field of dentistry. *Imaging Sci Dent* 2019; **49**: 1–7. doi: <https://doi.org/10.5624/isd.2019.49.1.1>
- Arijji Y, Yanashita Y, Kutsuna S, Muramatsu C, Fukuda M, Kise Y, et al. Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2019; **128**: 424–30. doi: <https://doi.org/10.1016/j.oooo.2019.05.014>
- Poedjiastoeti W, Suebnukarn S. Application of Convolutional neural network in the diagnosis of jaw tumors. *Health Inform Res* 2018; **24**: 236–41. doi: <https://doi.org/10.4258/hir.2018.24.3.236>
- Ohashi Y, Arijji Y, Katsumata A, Fujita H, Nakayama M, Fukuda M, et al. Utilization of computer-aided detection system in diagnosing unilateral maxillary sinusitis on panoramic radiographs. *Dentomaxillofac Radiol* 2016; **45**: 20150419–5. doi: <https://doi.org/10.1259/dmfr.20150419>
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; **9**: 676–82. doi: <https://doi.org/10.1038/nmeth.2019>
- Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017;: 2117–25.
- Redmon J, Farhadi A. YOLOv3: an incremental improvement. *arXiv e-print* 2018;. arXiv:1804.02767 p.
- Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection. *Advances in neural information processing systems* 2013;: 2553–61.
- Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition* 2014;: 580–7.

35. Deng X, Liu Q, Deng Y, Mahadevan S. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Inf Sci* 2016; **340-341**: 250–61. doi: <https://doi.org/10.1016/j.ins.2016.01.033>
36. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978; **8**: 283–98. doi: [https://doi.org/10.1016/S0001-2998\(78\)80014-2](https://doi.org/10.1016/S0001-2998(78)80014-2)
37. Scholl RJ, Kellett HM, Neumann DP, Lurie AG. Cysts and cystic lesions of the mandible: clinical and radiologic-histopathologic review. *Radiographics* 1999; **19**: 1107–24. doi: <https://doi.org/10.1148/radiographics.19.5.g99se021107>
38. Tsukamoto G, Sasaki A, Akiyama T, Ishikawa T, Kishimoto K, Nishiyama A, et al. A radiologic analysis of dentigerous cysts and odontogenic keratocysts associated with a mandibular third molar. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2001; **91**: 743–7. doi: <https://doi.org/10.1067/moe.2001.114157>
39. Eversole LR, Leider AS, Strub D. Radiographic characteristics of cystogenic ameloblastoma. *Oral Surg Oral Med Oral Pathol* 1984; **57**: 572–7. doi: [https://doi.org/10.1016/0030-4220\(84\)90320-7](https://doi.org/10.1016/0030-4220(84)90320-7)
40. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of Big Data* 2019; **6**: 1. doi: <https://doi.org/10.1186/s40537-019-0197-0>