Improvement diagnostic accuracy of sinusitis recognition in paranasal sinus X-ray using multiple deep learning models

Hyug-Gi Kim[1], Kyung Mi Lee[1], Eui Jong Kim[1], Jin San Lee[2]

[1]Department of Radiology, [2]Department of Neurology, Kyung Hee University College of Medicine, Kyung Hee University Hospital, Seoul, Republic of Korea

Correspondence to: Kyung Mi Lee, MD, PhD. Assistant Professor, Department of Radiology, Kyung Hee University College of Medicine, Kyung Hee University Hospital, #23 Kyunghee-daero, Dongdaemun-gu, Seoul 02447, Republic of Korea. Email: bandilee@khu.ac.kr.

Background: Sinus X-ray imaging is still used in the initial evaluation of paranasal sinusitis, which is diagnosed by the opacification or air/fluid level in the sinuses and best seen in the Waters' view of the paranasal sinus (PNS). The objective of this study was to investigate the feasibility of recognizing the maxillary sinusitis features using PNS X-ray images, as well as to propose the most effective method of determining a reasonable consensus using multiple deep learning models.

Methods: A total of 4,860 patients, which included 2,430 normal and maxillary sinusitis subjects each, underwent Waters' view PNS X-ray scan. The datasets were randomly split into training (70%), validation (15%), and test (15%) subsets. We implemented a majority decision algorithm to determine a reasonable consensus using three multiple convolutional neural network (CNN) models: VGG-16, VGG-19, and ResNet-101. The performance of sinusitis detection was evaluated with quantitative accuracy (ACC) and activation maps.

Results: We compared the results of our approaches with ACC and activation maps. ACC [and area under the curve (AUC)] of the internal test dataset was evaluated as 87.4% (0.891), 90.8% (0.891), 93.7% (0.937), and 94.1% (0.948) for VGG-16, VGG-19, ResNet-101, and the majority decision, respectively. ACC (and AUC) of the external test dataset was evaluated as 87.58% (0.877), 87.58% (0.877), 92.12% (0.929), and 94.12% (0.942) for VGG-16, VGG-19, ResNet-101, and the majority decision, respectively. Majority decision algorithms can detect missing and correct lesions using a compensation function of the majority decision.

Conclusions: The majority decision algorithm showed high accuracy and significantly more accurate lesion detection compared with those of individual CNN models. The proposed deep learning method with PNS X-ray images can be used as an adjunct to classify maxillary sinusitis.

Keywords: Sinusitis; deep learning; convolutional neural network (CNN); paranasal sinus (PNS) X-ray; majority decision

Introduction

Paranasal sinusitis is an inflammation of the mucosal lining of the paranasal sinuses (PNS) and is a common clinical problem in general populations (1). The sinusitis is diagnosed by the opacification of the sinuses and air/fluid level best seen in the Waters' view of X-ray imaging or computed tomography (CT). Because CT is not routinely indicated in all patients with possible sinusitis, X-ray is still initially used for its diagnosis. However, there are several limitations to using PNS X-ray scans as a diagnostic tool. On plain radiographs, other facial bony structures overlap

the sinuses, and the rate of false-negative results is high (2,3). Furthermore, the posterior ethmoids are poorly visualized, and the ostiomeatal complex cannot be adequately assessed. Therefore, it is difficult to visualize these characteristics to discriminate sinusitis from normal cases.

Many studies using machine learning have reported ways of classifying these ambiguous characteristics using techniques such as traditional machine learning support vector machines (SVMs), K-means clustering, naïve Bayers classifiers, etc. (4-6). Traditional machine learning methods require expert knowledge and time-consuming hand-tuning to extract specific features. Therefore, with traditional machine learning, features that represent the characteristics to be extracted must be implemented using various segmentation methods: thresholding, adaptive thresholding, clustering, region-growing, etc. (7). To overcome these limitations, deep learning (a branch of machine learning) can be used to acquire useful representations of features directly from data. Especially, the convolutional neural network (CNN) model uses a deep learning architecture to create a powerful imaging classifier (8). For that reason, recently, it is widely used to analyze medical images, such as X-ray, CT, and magnetic resonance imaging (MRI) images (9).

Deep learning algorithms through millions of data points to find patterns and correlations that often go unnoticed by human experts. These unnoticed features can produce unexpected results. Particularly in the case of deep learning-based disease diagnosis with medical images, it is important to evaluate not only the accuracy of the image classification but also the features used by the model, such as the location and shape of the lesion. However, one of the limitations is that there is a lack of transparency in deep learning systems, known as the "black box" problem. Without an understanding of the reasoning process to evaluate the results produced by deep learning systems, clinicians may find it difficult to confirm the diagnosis with confidence.

The objective of this study to investigate the ability of multiple deep learning models to recognize the features of maxillary sinusitis in PNS X-ray images and to propose the most effective method of determining a reasonable consensus.

Methods

Data preparation

Data collection

Our institutional review board approved this retrospective study and waived the requirements for informed consent due to the retrospective study design and the use of anonymized patient imaging data. We included records from January 2014 to December 2017 for the internal dataset and from January 2018 to May 2018 for the external test dataset. A total of 4,860 subjects for the internal dataset, which included 2,430 normal and sinusitis subjects each, and 160 subjects for the external test dataset (temporal test dataset), which included 80 normal and sinusitis subjects each, underwent PNS X-ray imaging using Waters' view. The temporal test dataset included newly collected data from subjects with a 1-year examination interval. The temporal test dataset was used to evaluate objective performance to ensure that the accuracy of the data for a specific period is not represented. If an image of the same subject that repeatedly scanned is included in the same dataset, it can affect the results with underestimate or, conversely too low. All X-ray images were compared with PNS CT as the reference standard for sinusitis of deep learning. After reviewing the imaging database, including the PNS X-ray images, 60 subjects were excluded because of artifact caused by sinus surgery (n=23), fracture (n=7), cyst or mass (n=15), severe movement (n=10), and incorrect scanning (n=5). To prevent overfitting, we conserved as many images as possible, including those with appropriate artifacts.

Pre-processing steps

The pre-processing steps were conducted with resizing, patch, and augmentation steps. The first pre-processing step normalizes the size of the input images. Almost all the radiographs were rectangles of different heights and too large (median value of matrix size ≥1,800). Accordingly, we resized all images to a standardized 224×224 pixel square, through a combination of preserving their aspect ratios and using zero-padding. The investigation of deep learning efficiency depends on the input data; therefore, in the second processing step, input images were pre-processed by using a patch (a cropped part of each image). A patch was extracted using a bounding box so that it contained sufficient maxillary sinus segmentation for analysis. Finally, data augmentation was conducted for just the training dataset, using mirror images that were reversed left to right and rotated –30, –10, 10, and 30 degrees. *Figure 1* shows a representation of PNS X-ray images in the pre-processing step described above.

Image labeling and dataset distributions

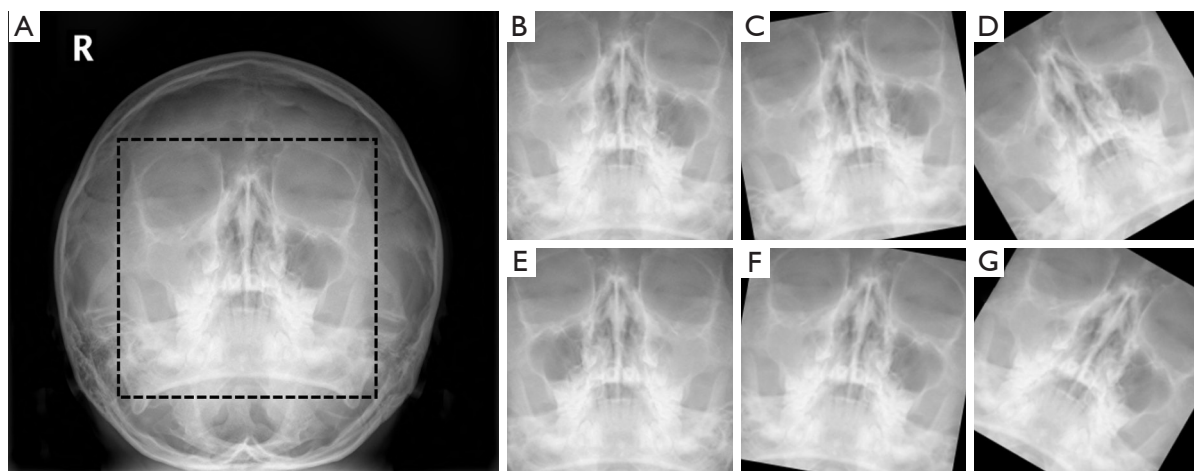All subjects were independently labeled twice as "normal"

Figure 1 Representation of the pre-processed PNS X-ray images. (A) Original PNS X-ray image; (B) patch that was extracted to contain sufficient maxillary sinus region from the original image (dotted bounding square box); (C,D,F,G) rotated images with –10 (C), –30 (D), 10 (F), and 30 degree (G) rotations; (E) mirrored image that was reversed left and right for data augmentation. PNS, paranasal sinus.

or "sinusitis" by two radiologists. Labeling was first evaluated with the original images on a picture archiving communication system (PACS) and secondly with the resized images that were used for the actual learning data. Datasets were defined as the internal dataset and temporal dataset, with the temporal dataset used to evaluate the test.

The internal dataset was randomly split into training (70%), validation (15%), and test (15%) subsets. The distribution of internal the test dataset consisted of 32% right maxillary sinusitis, 32% left maxillary sinusitis, and 34% bilateral maxillary sinusitis. The distribution of the temporal test dataset consists of 32.5% right maxillary sinusitis, 32.5% left maxillary sinusitis, and 35% bilateral maxillary sinusitis.

Majority decision algorithm

We implemented the majority decision algorithm to determine a reasonable consensus using three multiple CNN models in this study. Multiple CNN models consist of pre-trained VGG-16, VGG-19, and ResNet-101 architectures from ImageNet (http://www.image-net.org/) (10,11). An overview of the majority decision algorithm is shown in *Figure 2*. Briefly, each CNN model classified the input data and represented the major features for sinusitis using the activation map. Second, the majority decision was conducted based on the accuracy of a classification and the activation maps from the three models. The majority decision algorithm was performed using the criterion of

accuracy over 90% and a combination of intersection and union techniques of the activation map. The intersection method was applied to overlapping areas of the activation maps, and the union method was applied to non-overlapping areas. A model with an accuracy of under 90% was excluded for the evaluation of the majority decision.

VGG-16 and VGG-19 were tuned with the following parameters: factor for L2 regularization =0.004, max epochs =35, and size of mini-batch =10. The total training time for each model was 192 min 23 sec. ResNet-101 was tuned with the following parameters: factor for L2 regularization =0.001, max epochs =60, and size of mini-batch =64. Total training time was 372 min 23 sec.

All processing was performed on a personal computer equipped with an Intel Xeon E5-2643 3.40 GHz CPU, 256 GB memory, and Quadro M4000 D5 8GB GPU using MATLAB (MathWorks, R2018a, Natick, MA, USA).

Performance evaluation

The performance of sinusitis classification was evaluated with quantitative accuracy (ACC) and activation maps. First, the performance of the classification was evaluated using ACC and area under the curve (AUC) of receiver operating characteristic (ROC) curves. Second, the performance of recognizing features of maxillary sinusitis was evaluated using the activation map (or heat map) for each CNN model. Finally, the majority decision method was conducted to derive a reasonable consensus using the activation map
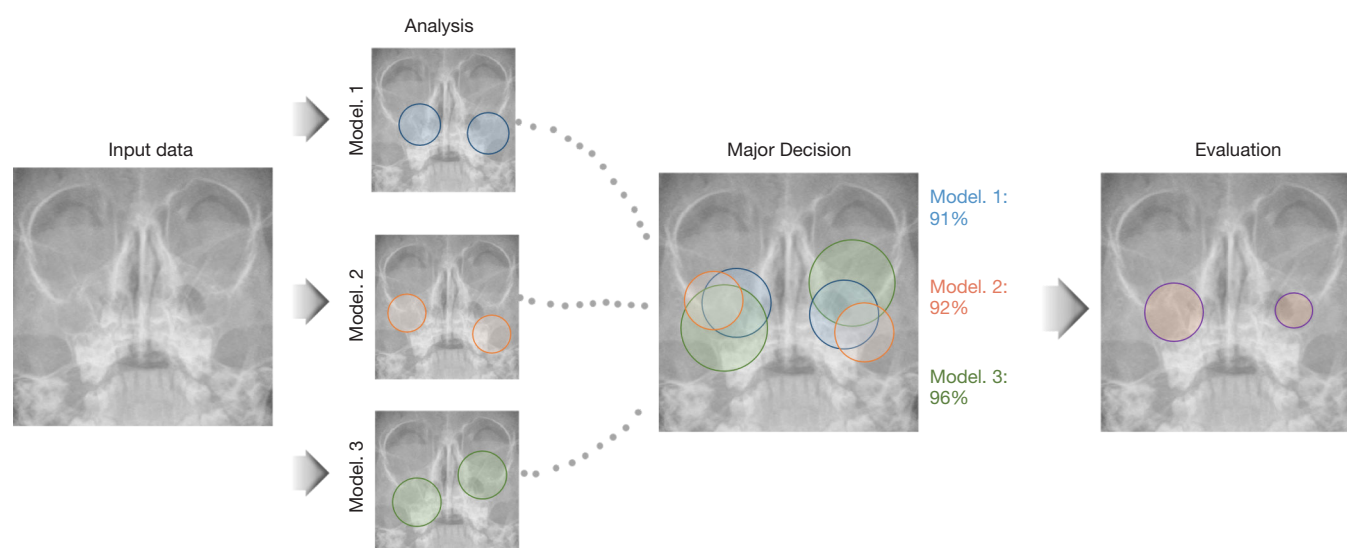
Figure 2 Representation of the process steps for the majority decision algorithm.

Table 1 Performance evaluation of training and validation datasets for each CNN model and the majority decision algorithm

Dataset/Model	VGG-16 (%)	VGG-19 (%)	ResNet-101 (%)	Majority decision (%)
Training set	99.8	99.8	99.9	99.9
Validation set	87.8	90.7	90.1	91.3

CNN, convolutional neural network.

and classification accuracy from the multiple CNN models. Two radiologists reviewed all classified images of the test dataset.

Statistical analysis of reproducibility for labeling and the AUC of the ROC curve was performed using MedCalc software (www.medcalc.org, Ostend, Belgium).

Results

Performance evaluation for classification

The reproducibility of the sinusitis labeling showed excellent agreement based on the two radiologists' evaluations (kappa value >0.85).

Table 1 shows the performance evaluation of the training and validation datasets for the VGG-16, VGG-19, ResNet-101 CNN models, and the majority decision with multiple CNN models. The ACCs of the training and validation datasets were evaluated as 99.8% and 87.8% for the VGG-16 model, 99.8% and 90.7% for the VGG-

19 model, and 99.9% and 90.1% for the ResNet-101 model, respectively. Performance of the majority decision was evaluated as 99.9% and 91.3% for the training and validation datasets.

Table 2 shows the performance evaluation of the internal and temporal test datasets for the VGG-16, VGG-19, ResNet-101 CNN models, and the majority decision algorithm with multiple CNN models. The ACC (and AUC) of the internal test dataset were evaluated as 87.4% (0.891), 90.8% (0.891), 93.7% (0.937), and 94.1% (0.948) for the VGG-16, VGG-19, ResNet-101 models, and the majority decision model, respectively. The ACC (and AUC) of the temporal test dataset were evaluated as 87.58% (0.877), 87.58% (0.877), 92.12% (0.929), and 94.12% (0.942) for the VGG-16, VGG-19, ResNet-101 models, and the majority decision model, respectively. The majority decision model showed the highest performance compared to the single CNN models for both the internal and temporal test datasets. *Figure 3* shows the results of ROC curve analysis with the internal (*Figure 3A*) and temporal

946

Kim et al. Sinusitis recognition using deep learning models

Table 2 Performance evaluation of internal and temporal test datasets for each CNN model and the majority decision algorithm

Model	ACC	AUC	SE	95% CI	Sensitivity	Specificity	P value
Internal test dataset							
VGG-16	87.4	0.891	0.0233	0.835 to 0.933	82.76	95.4	<0.0001
VGG-19	90.8	0.891	0.0237	0.835 to 0.933	87.36	90.8	<0.0001
ResNet-101	93.7	0.937	0.0183	0.890 to 0.968	89.66	97.7	<0.0001
Majority decision	94.1	0.948	0.0166	0.904 to 0.976	90.8	98.85	<0.0001
External test dataset (temporal test dataset)							
VGG-16	87.58	0.877	0.026	0.814 to 0.924	80.52	94.81	<0.0001
VGG-19	87.58	0.877	0.0266	0.814 to 0.924	85.71	89.61	<0.0001
ResNet-101	92.12	0.929	0.0206	0.876 to 0.964	88.31	97.4	<0.0001
Majority decision	94.12	0.942	0.0187	0.892 to 0.973	89.61	98.7	<0.0001

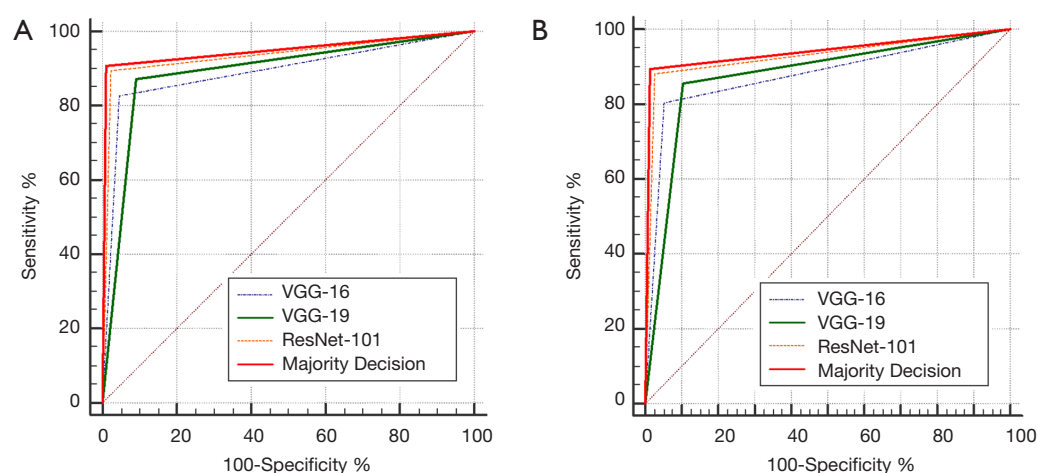CNN, convolutional neural network; ACC, accuracy; AUC, the area under the curve.

Figure 3 Comparison of ROC curve analysis of the multiple CNN models and the proposed majority decision for the classification of maxillary sinusitis. ROC, receiver operating characteristic; CNN, convolutional neural network.

test datasets (*Figure 3B*) for the multiple CNN models and the proposed majority decision algorithm. There are significant results between the proposed majority decision model and VGG-16 or VGG-19 (P=0.0011). However, there is no significant result between the proposed model and ResNet-101 (P=0.1573).

Performance evaluation to recognize the features of sinusitis

Figure 4 shows the performance evaluation to recognize the sinusitis features for each CNN model and the majority

decision model using the activation map from the test dataset. Test images were from patients with left maxillary sinusitis (*Figure 4A,B*), right maxillary sinusitis (*Figure 4C,D*), and bilateral maxillary sinusitis (*Figure 4E,F*). *Figure 4A*, *C*, and *E* were shown the results for internal test dataset and *Figure 4B*, *D*, and *F* were shown the results temporal test dataset. The three CNN models were evaluated with over 90% ACC.

The VGG-16, VGG-19, and ResNet-101 models detected sinusitis features using activation maps. Multiple CNN models were analyzed using the intersection and union methods from each activated map. Left sinusitis and right sinusitis features were analyzed using only the
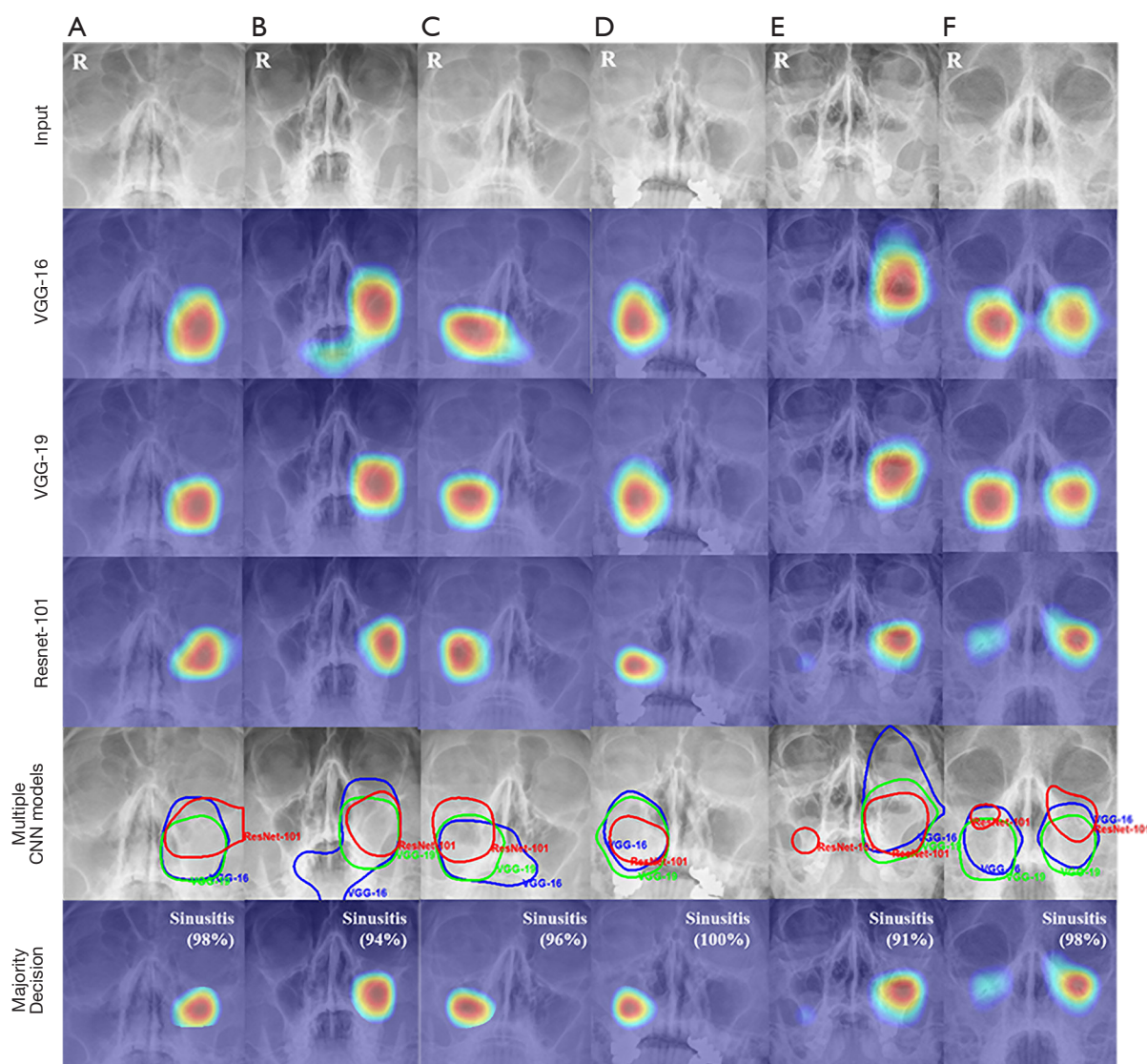
Figure 4 Performance evaluation to recognize the sinusitis features for each CNN model and the majority decision algorithm. Input patched images from patients with left maxillary sinusitis subjects (A,B), patients with right maxillary sinusitis (C,D), and patients with bilateral maxillary sinusitis (E,F). Internal test dataset (A, C, and E) and temporal test dataset subjects (B, D, and F). CNN, convolutional neural network.

intersection method (*Figure 4A,B,C,D*). Both sinusitis features were analyzed using intersection and union methods (*Figure 4E,F*).

The majority decision algorithm was shown to be significantly more accurate at lesion detection compared with the individual CNN models, and the activated features of sinusitis were visually confirmed to be closely correlated with the features evaluated by the radiologists using PNS X-ray images.

Performance evaluation to recognize the features of normal

Figure 5 shows the performance evaluation to recognize the features of normal subjects for each CNN model and the majority decision model using the activation map from the test dataset. *Figure 5A*, *B*, and *C* were shown the results for internal test dataset and *Figure 5D*, *E*, and *F* were shown the results of the temporal test dataset. The three CNN models
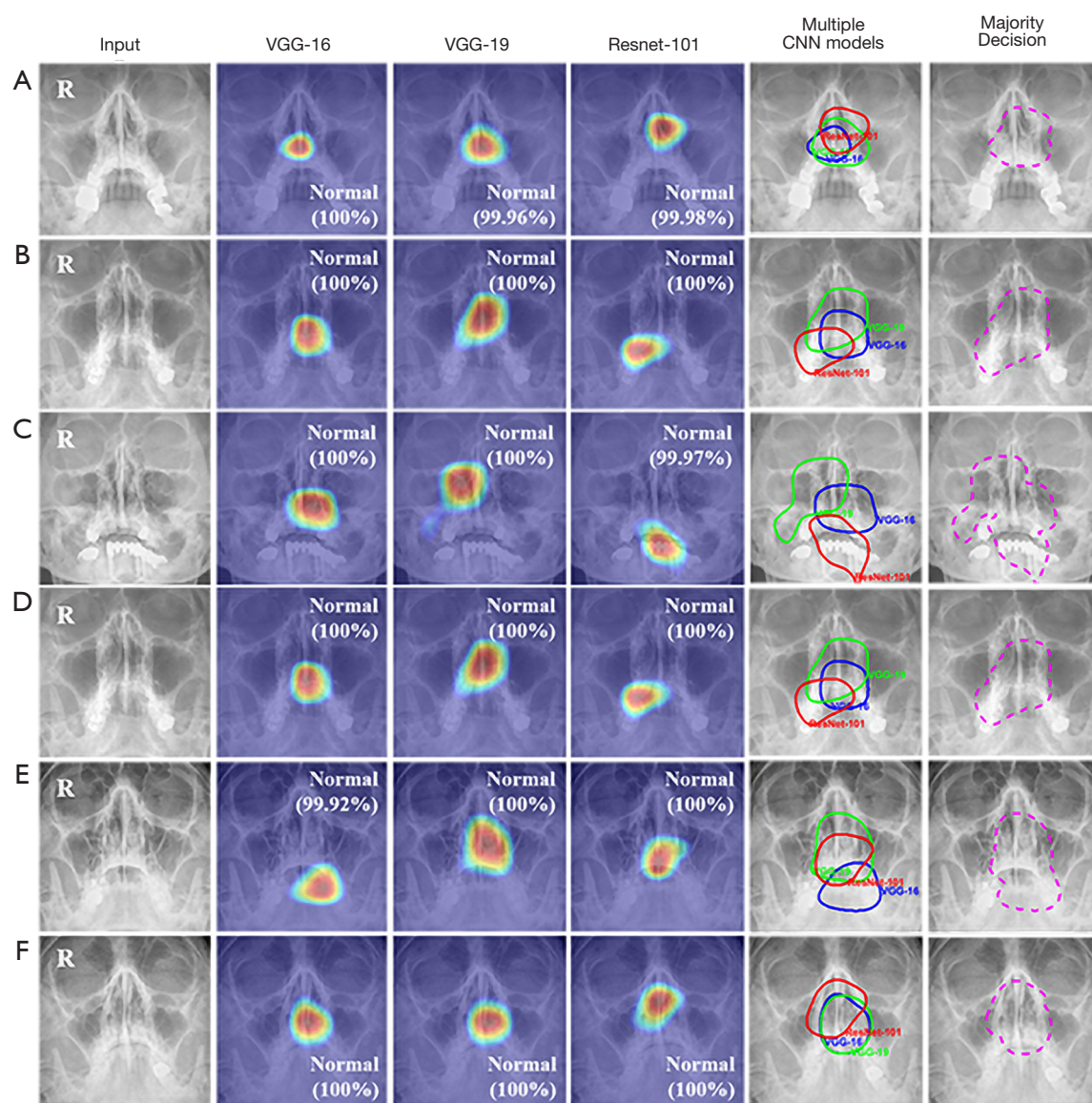
Figure 5 Performance evaluation to recognize the features of normal subjects for each CNN model and the majority decision algorithm. Internal test dataset (A, B, and C) and temporal test dataset subjects (D, E, and F). CNN, convolutional neural network.

were evaluated with over 90% ACC.

The VGG-16, VGG-19, and ResNet-101 models detected features of normal subjects using activation maps. Multiple CNN models were analyzed using the union methods from each activated map. In the case of normal subjects, nasal septal area, midline portion of the maxillary sinus, was activated. All models, including the majority decision algorithm, were evaluated as 'normal' with high accuracy.

Discussion

To apply deep learning systems for disease assessment in medical imaging, it is important to ensure that the classification achieves high accuracy in test datasets as well as reasonable feature extraction of the target lesions. Most of the recently published papers on medical imaging using deep learning have focused on achieving high accuracy with a single deep learning model (12-15). However, a single

deep learning model is not sufficient to provide a clear feature-based evaluation result because each model has its advantages and disadvantages. In this study, to solve that limitation, a majority decision algorithm was developed and applied with multiple deep learning models to evaluate maxillary sinusitis in PNS X-ray images.

In this study, image resizing, patch extraction, and rotation for data argumentation were performed as preprocessing steps (*Figure 1*). Using patched images is an efficient method to detect lesions accurately as well as to achieve high classification accuracy in medical images (16). Image rotation was performed only within –30 to 30 degrees of the original images from the training dataset. Data argumentation has the advantage of improving the learning power, but learning that includes unnecessary data is meaningless. For example, the angle of rotation for data augmentation is also significant in the evaluation of an actual image and is determined by the range of head movements in the actual PNS X-ray scan. In this study, the angle of rotation was limited to ±30 degrees for that reason. If images with 90- or 180-degree rotations are used for learning, the wrong features might be activated. This is because a feature from a region that is not in the position of the actual maxillary sinus can be learned incorrectly, due to rotation angle being too large.

Major objectives in deep learning are high accuracy and preventing overfitting problems. There are several techniques to prevent overfitting problems: more training data, adding stronger regularization, data augmentation, and reducing the complexity of the model. The proposed majority decision algorithm in this study can be used as a method to reduce overfitting. Comparing the results of the performance evaluation in *Table 2*, the majority decision algorithm was shown to have the least difference in accuracy between the training and test datasets (VGG-16: 12.4%, VGG-19: 9%, ResNet-101: 6.2%, majority decision: 5.8% for the internal test dataset and VGG-16: 12.22%, VGG-19: 12.22%, ResNet-101: 7.09%, majority decision: 5.78% for the temporal test dataset). This difference is one way to evaluate overfitting performance.

For the evaluation of accuracy, the majority decision algorithm was shown to be the most efficient model to classify maxillary sinusitis in PNS X-ray images (*Tables 1,2*). Compared to the VGG-16 and VGG-19 models, the ACC of the majority decision algorithm was 6.7% and 3.3% higher, respectively, for the internal test dataset. The ACC of the ResNet-101 model and the majority decision algorithm did not differ significantly for test dataset. The majority

decision algorithm also provides a function to compensate for incorrect classifications by an individual model. This compensation function was shown clearly in the combination of activation maps by the multiple CNN model.

The patient with bilateral sinusitis in *Figure 4E* was shown clearly to have left maxillary sinusitis and a mild degree of right maxillary sinusitis. The VGG-16 and VGG-19 models detected the left maxillary sinusitis but did not detect the mild maxillary sinusitis. The ResNet-101 model did detect the right mild maxillary sinusitis. In that case, the missing lesion from the first two models was detected using the compensation function of the majority decision algorithm. In the case with left sinusitis shown in *Figure 4A*, the ResNet-101 model activated a larger region. As a result, the activation map also detected the left eye region beyond the maxillary sinus. The VGG-16 and VGG-19 models only activated within the left maxillary sinus. In that case, the correct lesion was detected using the compensation function of the majority decision algorithm. The results from the temporal test dataset showed similar accuracy.

A review of the activation map in the test datasets for each model showed that the VGG-16 and VGG-19 models detected a relatively larger area than the actual lesion. Conversely, the ResNet-101 model detected a relatively smaller region than actual lesion, but it could also detect a lesion that was missed by the VGG-16 or VGG-19 models. The majority decision algorithm was shown to be effective in making a reasonable decision by compensating for the advantages and disadvantages of each deep learning model.

As a result of reviewing the classification accuracy with an activation map for normal subjects, a normal feature recognized the surrounding regions, not the maxillary sinus region (*Figure 5*). The reason is shown that CNN models recognize the differences in the characteristics between normal and sinusitis as a major feature of relatively bright signal intensity in the maxillary sinus. Previous papers showed only abnormal cases with a heat map; however, our study wants to secure reliability by showing the normal cases with multiple algorithms.

The selection criteria for the multiple CNN models included a high accuracy of classification and activation maps of the lesions. In addition to the VGG and ResNet models used in this study, to select multiple CNN models, Alexnet (17) and GoogleNet (18) were also evaluated for performance. However, the Alexnet and GoogleNet models were found to have less than a 90% ACC for classification, and they did not detect the sinusitis features. By contrast, the VGG and ResNet models showed reasonable feature

extraction for sinusitis. One of the reasons for this difference is the kernel size that is used to extract features in an image.

The VGG-16 and VGG-19 architecture consist of large kernel-size filters with multiple 3×3 kernel-size filters, one after another. Within a given receptive field (the effective area size of an input image on which output depends), multiple stacked smaller sized kernels are better than a single larger sized kernel because multiple non-linear layers increase the depth of the network, enabling it to learn more complex features at a lower cost. As a result, the 3×3 kernels in the VGG architecture help to retain more fine details of an image (10). The ResNet architecture is similar to the VGG model, consisting mostly of 3×3 filters. Additionally, the ResNet model has a network depth of as large as 152. Therefore, it achieves better accuracy than VGG and GoogleNet, while being computationally more efficient than VGG (11). While the VGG and ResNet models achieve phenomenal accuracy, their deployment on even the most modest sized GPUs is a problem because of the massive computational requirements, both in terms of memory and time.

There are several limitations to this study. First, the external test dataset in multiple medical centers did not be included for reproducibility. In the case of X-ray equipment, there is a relatively small difference in performance compared to other medical imaging equipment, depending on the manufacturer or model. In this study, therefore, the external test dataset in other medical centers did not be included. In the case of a local medical center using relatively old equipment; however, an additional performance evaluation is also required to utilize artificial intelligence (AI) assistive software. Second, the proposed majority decision algorithm was optimized to evaluate only maxillary sinusitis. Therefore, there is a limitation to evaluate sinusitis in frontal, ethmoid, and sphenoid. In order to utilize AI based assistive software in the future, further study is underway because it is necessary to evaluate sinusitis at other locations as well as maxillary. Third, it lacks pattern recognition and representation methods that can solve black-box in deep learning. It needs to determine a reasonable consensus for solving the black-box problem. The feature recognition based activation map was used to solve the black-box problem in deep learning. As it can be shown from the results, not only classification but also lesion localization can be expressed as a result. It helps medical doctors make a reasonable inference about the deep learning analysis. However, it is not enough to understand

all deep leaning procedures. For example, it is difficult to understand the pattern of each learned CNN model. By understanding the pattern recognition capabilities of each model, we can understand the advantages and disadvantages of each model and achieve the optimization of the overall AI system. To overcome this limitation, a feature connectivity representation should be available for each layer to determine which feature weights are strong (19). In addition to feature representation, text-based description algorithm can be applied to overcome the black-box limitation in a medical application using the convolutional recurrent neural network (CRNN) that is the combination CNN and recurrent neural network (RNN) (20,21).

A majority decision algorithm with multiple CNN models was shown to have high accuracy and significantly more accurate lesion detection ability compared to individual CNN models. The proposed deep learning method using PNS X-ray images can be used as an adjunct tool to help improve the diagnostic accuracy of maxillary sinusitis.

Acknowledgments

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: Our institutional review board approved this retrospective study and waived the requirements for informed consent due to the retrospective study design and the use of anonymized patient imaging data.

References

1. Fagnan LJ. Acute sinusitis: a cost-effective approach to diagnosis and treatment. Am Fam Physician 1998;58:1795-802, 805-6.
2. Kirsch CFE, Bykowski J, Aulino JM, Berger KL,

Choudhri AF, Conley DB, Luttrull MD, Nunez D Jr, Shah LM, Sharma A, Shetty VS, Subramaniam RM, Symko SC, Cornelius RS. ACR Appropriateness Criteria((R)) Sinonasal Disease. J Am Coll Radiol 2017;14:S550-S559.

3. Aaløkken TM, Hagtvedt T, Dalen I, Kolbenstvedt A. Conventional sinus radiography compared with CT in the diagnosis of acute sinusitis. Dentomaxillofac Radiol 2003;32:60-2.

4. Settouti N, Bechar ME, Chikh MA. Statistical Comparisons of the Top 10 Algorithms in Data Mining for Classification Task. International Journal of Interactive Multimedia and Artificial Intelligence 2016;4:46-51.

5. Wu XD, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou ZH, Steinbach M, Hand DJ, Steinberg D. Top 10 algorithms in data mining. Knowl Inf Syst 2008;14:1-37.

6. Dawson-Elli N, Lee SB, Pathak M, Mitra K, Subramanian VR. Data Science Approaches for Electrochemical Engineers: An Introduction through Surrogate Model Development for Lithium-Ion Batteries. J Electrochem Soc 2018;165:A1-A15.

7. Shih FY, Cheng SX. Automatic seeded region growing for color image segmentation. Image and Vision Computing 2005;23:877-86.

8. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998;86:2278-324.

9. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide. Radiology 2019;290:590-606.

10. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv 2014.

11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition 2016:770-8.

12. Kim Y, Lee KJ, Sunwoo L, Choi D, Nam CM, Cho J, Kim J, Bae YJ, Yoo RE, Choi BS, Jung C, Kim JH.

Deep Learning in Diagnosis of Maxillary Sinusitis Using Conventional Radiography. Invest Radiol 2019;54:7-15.

13. Larson DB, Chen MC, Lungren MP, Halabi SS, Stence NV, Langlotz CP. Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. Radiology 2018;287:313-22.

14. Mannil M, von Spiczak J, Manka R, Alkadhi H. Texture Analysis and Machine Learning for Detecting Myocardial Infarction in Noncontrast Low-Dose Computed Tomography: Unveiling the Invisible. Invest Radiol 2018;53:338-43.

15. Cicero M, Bilbily A, Colak E, Dowdell T, Gray B, Perampaladas K, Barfett J. Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. Invest Radiol 2017;52:281-7.

16. Zagoruyko S, Komodakis N. Learning to Compare Image Patches via Convolutional Neural Networks. Computer Vision and Pattern Recognition 2015.

17. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the Advances in Neural Information Processing Systems 2012:1097-105.

18. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going Deeper with Convolutions. Conference on Computer Vision and Pattern Recognition (CVPR) 2015.

19. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Object detectors emerge in Deep Scene CNNs. International Conference on Learning Representations (ICLR) 2015.

20. Donahue J, Hendricks LA, Rohrbach M, Venugopalan S, Guadarrama S, Saenko K, Darrell T. Long-term Recurrent Convolutional Networks for Visual Recognition and Description. 2016.

21. Karpathy A, Fei-Fei L. Deep Visual-Semantic Alignments for Generating Image Descriptions. Computer Vision and Pattern Recognition (CVPR) 2015.