# Automated Radiographic Evaluation of Adenoid Hypertrophy Based on VGG-Lite

J.L. Liu[1] iD, S.H. Li[2] iD, Y.M. Cai[3], D.P. Lan[3], Y.F. Lu[1], W. Liao[1] iD, S.C. Ying[4], and Z.H. Zhao[1]

## Abstract

Adenoid hypertrophy is a pathological hyperplasia of the adenoids, which may cause snoring and apnea, as well as impede breathing during sleep. The lateral cephalogram is commonly used by dentists to screen for adenoid hypertrophy, but it is tedious and time-consuming to measure the ratio of adenoid width to nasopharyngeal width for adenoid assessment. The purpose of this study was to develop a screening tool to automatically evaluate adenoid hypertrophy from lateral cephalograms using deep learning. We proposed the deep learning model VGG-Lite, using the largest data set (1,023 X-ray images) yet described to support the automatic detection of adenoid hypertrophy. We demonstrated that our model was able to automatically evaluate adenoid hypertrophy with a sensitivity of 0.898, a specificity of 0.882, positive predictive value of 0.880, negative predictive value of 0.900, and F1 score of 0.889. The comparison of model-only and expert-only detection performance showed that the fully automatic method (0.07 min) was about 522 times faster than the human expert (36.6 min). Comparison of human experts with or without deep learning assistance showed that model-assisted human experts spent an average of 23.3 min to evaluate adenoid hypertrophy using 100 radiographs, compared to an average of 36.6 min using an entirely manual procedure. We therefore concluded that deep learning could improve the accuracy, speed, and efficiency of evaluating adenoid hypertrophy from lateral cephalograms.

**Keywords:** orthodontics, machine learning, artificial intelligence, diagnostic imaging, medical image classification, convolutional neural networks

## Introduction

The adenoid is an aggregate of lymphatic tissue located in the posterior nasopharyngeal airway. It commonly appears in early childhood between 6 and 10 y old and then disappears at 16 y old (Major et al. 2006; Yildirim et al. 2008). Adenoid hypertrophy, a pathological enlargement of the nasopharyngeal tonsils, is the most prevalent cause of nasal obstruction in childhood (Yildirim et al. 2008) and is associated with the "adenoid face" morphology featuring a narrow maxillary arch, posterior crossbite, retrognathic mandible, and large face height (Major et al. 2006). According to a recent meta-analysis, adenoid hypertrophy prevalence is 34% in the general pediatric population and varies between 42% and 70% in pediatric populations (Pereira et al. 2018). Adenoid hypertrophy may be caused by frequent upper airway infections, and it often leads to nasal congestion, breathing through the mouth, snoring, and otitis media with effusion (Chien et al. 2005).

Children with adenoid hypertrophy usually present in dental clinics with a chief complaint of occlusal disorder or dissatisfaction with their profile. Dentists typically obtain a lateral cephalogram in order to screen for the possibility of adenoid hypertrophy: the ratio of adenoid width to nasopharyngeal width (AN ratio) is measured on this image (Major et al. 2006),

[1]State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases, Department of Orthodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, Sichuan, China
[2]National Key Laboratory of Fundamental Science on Synthetic Vision, College of Computer Science, Sichuan University, Chengdu, Sichuan, China
[3]Department of Dental Technology, West China Hospital of Stomatology, Sichuan University, Chengdu, China
[4]College of Computer Science, Sichuan University, Chengdu, Sichuan, China

A supplemental appendix to this article is available online.

**Corresponding Authors:**
Z.H. Zhao, State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases, Department of Orthodontics, West China Hospital of Stomatology, Sichuan University, No. 14, 3rd Section, Renminnan Road, Chengdu, Sichuan 610041, China.
Email: zhzhao@scu.edu.cn
S.C. Ying, College of Computer Science, Sichuan University, No. 17 People's South Road, Chengdu, Sichuan 610041, China.
Email: yingsancong@scu.edu.cn
W. Liao, State Key Laboratory of Oral Diseases & National Clinical Research Center for Oral Diseases, Department of Orthodontics, West China Hospital of Stomatology, Sichuan University, No. 14, 3rd Section, Renminnan Road, Chengdu, Sichuan 610041, China.
Email: liaowenssw@126.com

which allows the clinician to determine whether the adenoid is pathologically enlarged or not (Elwany 1987). If deemed pathologically enlarged, the patient is typically subjected to more extensive otorhinolaryngology follow-up for definitive diagnosis and may accept adenoidectomy operation. Therefore, dentists should attach great importance to the concept of a correct early screen of adenoid hypertrophy, which is helpful for preventing any associated craniofacial consequences. During evaluation of children with suspected adenoid hypertrophy based on lateral cephalograms, dentists need to label landmarks on the cephalogram in order to measure the AN ratio, which is time-consuming, tedious, and experience dependent. According to the latest World Health Organization (WHO) report, the dentist-to-population ratio was 4.46:10,000 in China in 2017 (World Health Organization 2017). An excessive number of patients with a relatively small number of dentists has resulted in high workload. Therefore, there is an urgent need to develop a fully automated evaluation method to improve work efficiency and alleviate work burden of dentists.

Deep learning is a method of machine learning based on neural networks (LeCun et al. 2015). Deep learning has been shown to perform better than traditional machine learning algorithms on many computer vision tasks, such as classification, segmentation, and detection (Ronneberger et al. 2015; Redmon et al. 2016). Of these, classification is one of the tasks to which deep learning is most often applied (Ali et al. 2020; Kim and Kim 2020; Rajasree et al. 2020; Yousef et al. 2020), particularly for disease diagnosis based on medical images (Bhatele and Bhadauria 2020). Recently, many deep learning–based methods have been applied to automatic disease diagnosis. For example, 1 study applied the YOLO model to detect odontogenic cysts and tumors of the jaw in panoramic radiographs with a precision value of 0.707 and a recall value of 0.680 (Yang et al. 2020). A deep learning method has also been applied to orthodontic diagnosis, offering sensitivity, specificity, and accuracy over 90% for vertical and sagittal skeletal diagnosis (Yu et al. 2020). So far, only 1 study has applied deep learning for diagnosis of adenoid hypertrophy (Shen et al. 2020). Those investigators collected a total of 688 X-ray images of patients with adenoid hypertrophy and divided them into 3 groups for training (488 images), validation (64 images), and testing (116 images). After training, their deep learning model produced 4 landmarks for adenoid-nasopharynx measurement on X-ray images and showed a moderate F1 score of 0.624 under pretrained conditions or 0.545 not under pretrained conditions. Such landmark localization method based on deep learning algorithms has been widely investigated, and researchers were committed to minimize the errors and bias associated with landmark identification (Dot et al. 2020; Lee et al. 2020; Noothout et al. 2020). In view of the unavoidable localization error, we reasoned that improvement in methodology, which eliminated the process of landmark identification, was needed to improve the performance of the evaluation with less manual intervention.

Therefore, in the present study, we proposed a novel and simplified deep learning model, VGG-Lite, for fully automated evaluation of adenoid hypertrophy based on the convolutional neural network architecture VGG16 (Guari et al. 2019). We built a larger data set hitherto with 1,023 X-ray images for training, validation, and testing. We comprehensively evaluated the performance of our deep learning method for automated detection of adenoid hypertrophy. The present work extended the literature in several ways. Few studies have reported on direct detection of adenoid hypertrophy on X-ray images with deep learning, and none has used such a large data set, which contains more than 1,000 samples. The model that we proposed was able to accurately classify radiographs into those showing normal adenoid or pathological adenoid hypertrophy without any manual intervention. To our knowledge, the present study is the first attempt to comprehensively evaluate the performance of deep learning for direct detection of adenoid hypertrophy. By eliminating the landmark identification process, we expected to accelerate the evaluation process and develop a powerful tool for children who should be referred to otolaryngologists for surgery.

## Materials and Methods

### Study Population and Image Data Set

All experimental procedures involving human X-ray images were approved by the West China Hospital of Stomatology Ethics Committee (WCHSIRB-D-2020-409), in Chengdu, Sichuan, China. X-ray images were retrospectively examined for 2,016 patients who had initially visited our hospital between January and December 2019. Details about the cephalometric radiographs are given in the Appendix.

Figure 1A depicts the manual measurement for calculation of the adenoidal-nasopharyngeal ratio. Further detail on the cephalometric analysis of adenoid is available in the Appendix. According to the research of Elwany et al. (1987), a value of the AN ratio greater than 0.73 may be regarded as indicative of pathological adenoid hypertrophy. In our research, all X-ray images were manually labeled as "pathological adenoid hypertrophy" (AN ratio >0.73) or "normal adenoid" (AN ratio ≤0.73) by 2 experienced orthodontists (J.L. and W.L., with more than 5 y of experience) blinded to each other. To reduce the potential influence of subjectivity on outcome, the 2 orthodontists were trained several times prior to the trial to ensure consistency of data acquisition. A third senior orthodontic specialist (Z.Z., with 30 y of experience) was consulted in cases of disagreement. If the 3 experts still could not get an agreement, the confusing image would be excluded.

Among the 2016 cephalometric radiographs, we excluded low-quality images (783 images) and eliminated 210 images in which the occipital slope was difficult to recognize. Finally, a total of 1,023 X-ray images of patients were included in the study (505 normal adenoids and 418 cases of pathological adenoid hypertrophy). Patients were randomly divided into 2 groups: a training set (923 images) and a testing set (100 images).

## Deep Learning Model for Diagnosis of Adenoid Hypertrophy

The VGG16 model has demonstrated excellent performance for image classification and can potentially be used to classify medical images (Krizhevsky et al. 2017; Guari et al. 2019). However, it is designed for big data sets like the ImageNet data set and has so many parameters that the model can easily overfit the training set (Mummadi et al. 2018). Details about the method of relieving overfitting and VGG-Lite configuration are summarized in Table 1 and the Appendix. The model architecture is shown in Figure 1B.

## Statistical Analysis and Evaluation Criteria

All statistical analyses were done with the IBM SPSS statistical software Version 20.0. To compare the performance of the 2 deep learning models (VGG16 and VGG-Lite), we used the metrics of sensitivity (SEN), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), F1 score, receiver operating characteristic (ROC) curves, and the area under ROC curve (AUC). The Appendix includes the details of metrics.

To compare the performance of VGG-Lite model and human experts, the other 3 orthodontic specialists (C.Z., D.J., and Y.L., with more than 5 y of experience) and VGG-Lite model detected the same 100 X-ray images selected from testing set. The agreements between model and experts were evaluated by Cohen's κ. Two weeks later, the 3 orthodontic specialists (C.Z., D.J., and Y.L.) detected the 100 radiographs selected from testing test with the aid of VGG-Lite model. For 1 radiograph, experienced orthodontists confirmed whether the screening result generated by the model was consistent with the subjective visual assessment of them. If not, the orthodontists would define the radiograph according to measurement of the AN ratio.

To assess the interobserver and intraobserver agreement, randomly selected 150 cephalometric radiographs (not in the testing set) were evaluated by the 3 experts (C.Z., D.J., and Y.L.) and then revaluated after 2 wk. Fleiss's κ (Fleiss 1971) was used to assess interobserver agreement, while Cohen's κ (Cohen 1960) was applied to assess intraobserver reliability.

# Results

## Deep Learning Model

Demographic and clinical characteristics of the patients included in the study are shown in Table 2. All experiments were performed using Python 3.6 and TensorFlow 1.9 on a single NVIDIA RTX 2080Ti (Abadi 2016). We proposed an advanced model—namely, VGG-Lite—for automatic diagnosis of adenoid hypertrophy based on VGG16. According to the
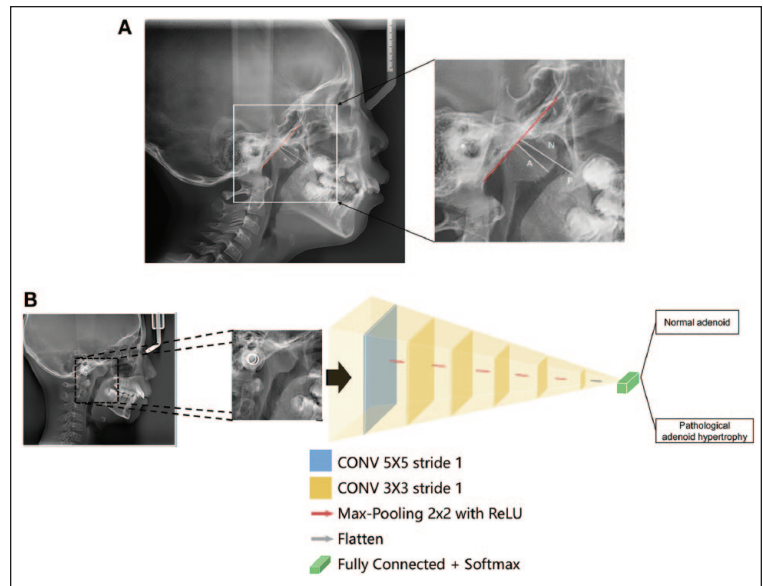


**Figure 1.** Overview of manual labeling process and the deep learning model architecture. (**A**) Example of manual measurement for calculation of the adenoidal-nasopharyngeal ratio (AN ratio, where A is the absolute size of the adenoid and N is the size of the nasopharyngeal space) on a standard lateral skull radiograph. The red line is tangential to the basiocciput. The adenoidal value A is obtained by drawing a perpendicular line to the red line at the point of maximal adenoid tissue. The nasopharyngeal value N is made between the posterior border of the hard palate (P) and the anteroinferior aspect of the spheno-basioccipital synchondrosis. (**B**) Architecture of the VGG-Lite model.

**Table 1.** VGG-Lite Configuration.

| Structure of VGG-Lite | Configuration Information |
| --- | --- |
| Convolutional layer 1 | Kernel size = 5 × 5, stride = 1, kernel number = 10 |
| Max-pooling | Pooling size = 2 × 2 |
| Convolutional layer 2 | Kernel size = 3 × 3, stride = 1, kernel number = 20 |
| Max-pooling | Pooling size = 2 × 2 |
| Convolutional layer 3 | Kernel size = 3 × 3, stride = 1, kernel number = 30 |
| Max-pooling | Pooling size = 2 × 2 |
| Convolutional layer 4 | Kernel size = 3 × 3, stride = 1, kernel number = 40 |
| Max-pooling | Pooling size = 2 × 2 |
| Convolutional layer 5 | Kernel size = 3 × 3, stride = 1, kernel number = 50 |
| Max-pooling | Pooling size = 2 × 2 |
| Convolutional layer 6 | Kernel size = 3 × 3, stride = 1, kernel number = 60 |
| Fully connected layer with Softmax | Node number = 2 |

5-fold cross-validation method, we randomly selected one-fifth of the training set as a validation set. We performed the cross-validation procedure on this validation set during model training. In the training phase, we used a learning rate of 0.001 and a batch size of 150 in the Adam optimizer, and then we chose the "cross-entropy" as the loss function. The epoch number was set to 100 for model training. According to the
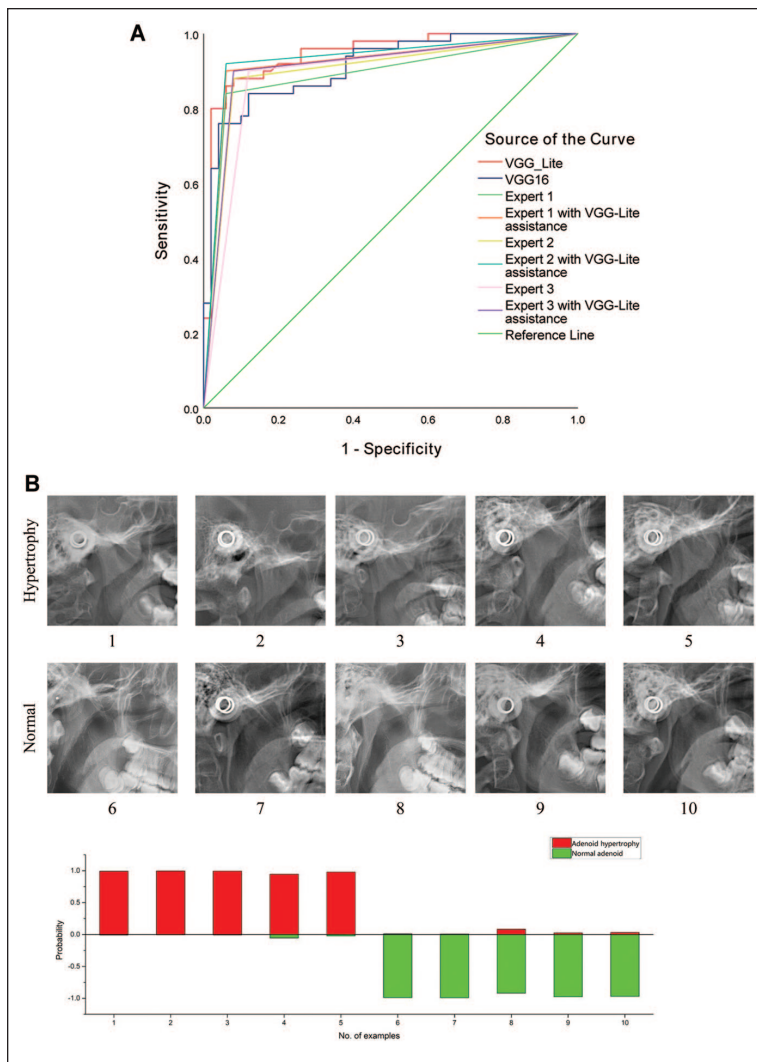
**Figure 2.** Overview of experimental results. (**A**) Receiver operating characteristic (ROC) curves and the area under the curves (AUCs) are represented in parentheses. Images of the adenoid region: images 1 to 5 are pathological samples; images 6 to 10, normal adenoid samples. (**B**) Representative examples of VGG-Lite prediction of adenoid hypertrophy based on lateral cephalograms. Probability of model prediction: red means that the model classified the sample as pathological; green, as normal adenoid.

**Table 2.** Clinical and Demographic Characteristics of the Study Patients.

| Characteristic | Training Set (*n* = 923) | Testing Set (*n* = 100) |
|---|---|---|
| Age, median (range), y | 9.04 (4–11) | 9.08 (5–11) |
| Sex, *n* (%) | | |
| Male | 412 (44.6) | 42 (42.0) |
| Female | 511 (55.4) | 58 (58.0) |
| Clinical evaluation, *n* (%) | | |
| Pathological adenoid hypertrophy | 468 (50.7) | 50 (50.0) |
| Normal adenoid | 455 (49.3) | 50 (50.0) |

performance of the validation set, the highest performance appeared from 50 to 60 epochs. We saved the model when the highest performance appeared first. Representative examples

of automatic diagnosis for adenoid hypertrophy are shown in Figure 2B. The VGG16 achieved a sensitivity of 0.8750, specificity of 0.8462, PPV of 0.840, NPV of 0.8800, and F1 score of 0.8571. The VGG-Lite achieved a sensitivity of 0.8980, specificity of 0.8824, PPV of 0.8800, NPV of 0.9000, and F1 score of 0.8889 (Table 3). The ROC curves and the AUCs are provided in Figure 2A with VGG-Lite of 0.946 and VGG16 of 0.913.

### Comparison of Model Performance and Expert-Only Detection

The deep learning model showed a strong ability to learn features from X-ray images through manually annotated samples. Our model showed a very close performance to expert-only diagnosis, based on the F1 score (Table 3). The κ values were found to be 0.86 (C.Z. vs. model), 0.90 (D.J. vs. model), and 0.88 (Y.L. vs. model). Our model showed values of SEN, PPV (reflected in the F1 score), SPEC, and NPV similar to those of human experts. The execution time of the deep learning model was 0.07 min for observing 100 images, while a human expert needed an average of 36.6 min to observe 100 X-ray images and perform a diagnosis (Table 3). Our experimental results demonstrated that the fully automatic method was about 522 times faster than the human expert.

### Comparison of Human Experts with or without Deep Learning Assistance

To comprehensively evaluate the applicability of our deep learning model, we compared human-only and human-machine diagnosis of adenoid hypertrophy. Deep learning–assisted human experts (*n* = 3) required an average of 23.3 min for detection in the testing set (100 images), which was significantly more efficient than manual diagnosis only (36.6 min). Furthermore, deep learning slightly improved the accuracy of diagnosis: F1 score was increased by 2.78% with deep learning assistance (Table 3).

### Interobserver and Intraobserver Agreement

Fleiss κ value of 3 human experts was 0.956. Cohen κ values were 0.973 (expert 1, C.Z.), 0.960 (expert 2, D.J.) and 0.947 (expert 3, Y.L.).

## Discussion

Children with adenoid hypertrophy usually present in dental clinics with a chief complaint of occlusal disorder or dissatisfaction with their profile. The golden standard method defined in the diagnosis of adenoid hypertrophy was nasal fiberoptic endoscopy (Abdollahi-Fakhim et al. 2008). Although it can

**Table 3.** Performance of Human Experts and Model for Adenoid Hypertrophy Detection in the Testing Set.

| | | Pathological | | Normal | | | | | | | Time to Diagnosis, min |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Operator | Method | True Positive | False Positive | True Negative | False Negative | SEN | SPEC | PPV | NPV | F1 Score | |
| VGG16 | Automatic | 42 | 8 | 44 | 6 | 0.8750 | 0.8462 | 0.8400 | 0.8800 | 0.8571 | 0.13 |
| VGG-Lite | Automatic | 44 | 6 | 45 | 5 | 0.8980 | 0.8824 | 0.8800 | 0.9000 | 0.8889 | 0.07 |
| Expert 1 | Manual | 42 | 8 | 47 | 3 | 0.9333 | 0.8545 | 0.8400 | 0.9400 | 0.8842 | 36.7 |
| | VGG-Lite assisted | 45 | 5 | 47 | 3 | 0.9375 | 0.9038 | 0.9000 | 0.9400 | 0.9184 | 21.1 |
| Expert 2 | Manual | 44 | 6 | 46 | 4 | 0.9167 | 0.8846 | 0.8800 | 0.9200 | 0.8980 | 37.2 |
| | VGG-Lite assisted | 46 | 4 | 47 | 3 | 0.9388 | 0.9216 | 0.9200 | 0.9400 | 0.9293 | 24.9 |
| Expert 3 | Manual | 45 | 5 | 44 | 6 | 0.8824 | 0.8980 | 0.9000 | 0.8800 | 0.8911 | 35.9 |
| | VGG-Lite assisted | 45 | 5 | 46 | 4 | 0.9184 | 0.9020 | 0.9000 | 0.9200 | 0.9091 | 23.8 |

NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SPEC, specificity.

provide invaluable diagnostic information, it is not suitable for screening in dental clinics since it is too expensive and time-consuming compared with X-ray examination. Many studies proved the high reliability of cephalometric radiographs when detecting adenoid hypertrophy (Moideen et al. 2019; Soldatova et al. 2020). Hence, in this study, we proposed a practical and simplified deep learning model, VGG-Lite, for automated evaluation of adenoid hypertrophy based on cephalometric radiographs. We show that VGG-Lite is able to detect adenoid hypertrophy with high accuracy and reduce the time to diagnosis, which could be served as a reliable screening tool for referral.

Many studies have reported that deep learning methods have impressive learning capacity and classification accuracy for automatic diagnosis of diseases such as cancer, caries lesions, and diabetic retinopathy and for other medical applications (Cantu et al. 2020; Liu et al. 2020; Sarao et al. 2020). However, few studies have focused on diagnosis of adenoid hypertrophy based on deep learning. In this article, we observed that our deep learning model showed good ability to learn and detect adenoid hypertrophy after training with professionally annotated cephalometric images. Our data showed that our method performed similarly to human experts based on standard classification metrics with faster detection speed.

A previous study applied deep learning for keypoint localization on lateral cephalometric images for automatic diagnosis of adenoid hypertrophy (Shen et al. 2020). Those researchers measured the AN ratio according to landmarks generated by deep learning. In that work, deep learning achieved direct classification with an F1 score of 0.624. We reasoned that relying on keypoint localization rather than direct classification may be caused by the limited performance of the model with a small, imbalanced data set (68 normal samples vs. 488 hypertrophic samples). Errors in the keypoints method are widespread, necessitating time-consuming manual correction. Hence, the method of direct classification produced by deep learning appears to be more practical and efficient for clinical applications. To obtain the best performance of the deep learning model, we assembled a balanced data set with comparable numbers of nonpathological and pathological X-ray images for training and testing (505 normal samples vs. 518 hypertrophic

samples). Another model for diagnosis of adenoid face from photographs has been proposed based on traditional machine learning (Hu et al. 2019). Those researchers used an existing face detection software, Dlib, to extract the facial feature points in patient photographs. Subsequently, machine learning models were trained using these points and used to predict adenoid face in photographs of new patients. The best performance of traditional machine learning (support vector machine) applied by Hu et al. (2019) achieved a sensitivity of 0.882 and a specificity of 0.703. Our model showed a better performance in both sensitivity (0.898) and specificity (0.882). Moreover, our model did not require extra steps of feature extraction for training or prediction. In the present study, we assembled what appears to be the largest data set so far to support automatic diagnosis of adenoid hypertrophy. We also verified the effectiveness of deep learning for assisting human experts.

In the past 5 y, many physicians have started to apply deep learning to various medical applications. Deep learning technology is expected to be possibly one of the most clinically applicable classification methods (Casalegno et al. 2019; Yu et al. 2020). The main reason is that manual error correction is easier in classification tasks than in other types of tasks, such as segmentation and keypoint localization. With time, more samples with incorrect prediction can be found and added to the training set to improve the performance of the deep learning model. Therefore, model performance may further improve with clinicians' usage time. At the same time, classification tasks may be more amenable to deep learning than nonclassification tasks such as segmentation and keypoint localization. This condition might be explained by problems with nondeterministic error. For deep learning, it is impossible to use manual error correction to improve model performance in segmentation and other tasks. Hence, in this study, we decided to use deep learning to classify X-ray images to obtain the diagnostic conclusions rather than perform other types of tasks to diagnose. The structure of the model is required to be comprehensively considered for different data sets. Data sets in the medical field are different from and smaller than many data sets collected in daily life. At the same time, typical deep learning models have a large number of parameters. Hence, overfitting problems occur widely in various applications of

deep learning if typical models are trained using small data sets. To avoid problems of overfitting, we developed in this work a Lite model based on VGG16. Our model with fewer parameters performed better than the VGG16 model with more parameters.

The reason why we applied classification model VGG rather than an image detection model like YOLO was argued as follows: the output of the VGG model is the probability of every category in terms of the X-ray image as input, but the image detection model needs human experts to record the location of the adenoid region on every X-ray image and label the detection result of every region. Regarding the detection of adenoid hypertrophy, the region of adenoid is fixed on the center of original X-ray images. We can crop this region on the X-ray images easily instead of dynamically detecting the region of adenoid using the YOLO model with potential region detection error. Besides, in this study, our model was trained by the region of adenoid, which was cropped from the original X-ray images. In image processing, a method based on the region of interest (ROI) is a common way for model construction. The ROI-based method has some advantages. First, compared to the original X-ray image, ROI provides enough information for evaluation with less data as input. The ROI with a smaller size reduces memory overhead. Second, deep learning is a resource-consuming method in memory and calculation. The input image with high resolution requires more computational workload and memory resource. Therefore, we suggest that extra preprocessing, such as ROI extraction, is required as it is resource-friendly.

Nevertheless, our study presents several limitations. First, some images were annotated incorrectly during the building of the data set, so further manual data cleaning is necessary for this work. This should be feasible, given that our data set has only around 1,000 X-ray images. Future work on automatic data cleaning will be very meaningful for medical applications based on deep learning when an extremely large data set is used. Second, our 1,023 X-ray images were produced by Pax-400C and Morita X550 at 2 resolutions. Our model may not be robust at other resolutions, which should be addressed in the future through appropriate expansion of the training set with images at other resolutions. Third, similar to other dental studies, our research had to use multiple human annotators to independently evaluate the adenoid on cephalometric images, since radiologic evaluation of different orthodontists unavoidably tends to be subjective (Schwendicke et al. 2021). Therefore, inevitable subjectivity in radiologic evaluation may affect the results, which posed a limitation on our study.

## Conclusions

We show that deep learning methods can automatically detect pathological adenoid hypertrophy with higher accuracy and faster speed than manual diagnosis. A Lite model based on a typical deep learning structure showed good performance for automatic diagnosis of adenoid hypertrophy based on medical images. Deep learning may be a useful tool for relieving dentists' workload of early screen and improving diagnostic accuracy in adenoid hypertrophy.

## Author Contributions

J.L. Liu, contributed to conception, design, and data analysis, drafted and critically revised the manuscript; S.H. Li, contributed to conception, design, data analysis, and interpretation, drafted the manuscript; Y.M. Cai, D.P. Lan, Y.F. Lu, contributed to data acquisition, drafted the manuscript; W. Liao, S.C. Ying, Z.H. Zhao, contributed to conception and design, critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ORCID iDs

J.L. Liu https://orcid.org/0000-0001-5554-0870

S.H. Li https://orcid.org/0000-0002-6418-7818

W. Liao https://orcid.org/0000-0002-6861-0054

## Data and Materials Availability

The data sets generated or analyzed during the current study are not publicly available in order to preserve patient confidentiality but are available from the corresponding authors on reasonable request.

## References

Abadi M. 2016. Tensorflow: learning functions at scale. Paper presented at: ICFP 2016 Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming. doi:10.1145/2951913.2976746

Abdollahi-Fakhim S, Naderpoor M, Shahid N, Javadrashid R, Mashrab O, Ravaghi M. 2008. Assessment of adenoid size in children. Res J Biol Sci. 3(7):747–749.

Ali F, El-Sappagh S, Islam SMR, Kwak D, Ali A, Imran M, Kwak KS. 2020. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Inf Fusion. 63:208–222.

Bhatele KR, Bhadauria S. 2020. Brain structural disorders detection and classification approaches: a review. Artif Intell Rev. 53(5):3349–3401.

Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, Elhennawy K, Schwendicke F. 2020. Detecting caries lesions of different radiographic extension on bitewings using deep learning. J Dent. 100:103425.

Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schurmann F, Krejci I, Markram H. 2019. Caries detection with near-infrared transillumination using deep learning. J Dent Res. 98(11):1227–1233.

Chien CY, Chen AM, Hwang CF, Su CY. 2005. The clinical significance of adenoid-choanae area ratio in children with adenoid hypertrophy. Int J Pediatr Otorhinolaryngol. 69(2):235–239.

Cohen J. 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas. 20(1):37–46.

Dot G, Rafflenbeul F, Arbotto M, Gajny L, Rouch P, Schouman T. 2020. Accuracy and reliability of automatic three-dimensional cephalometric landmarking. Int J Oral Maxillofac Surg. 49(10):1367–1378.

Elwany S. 1987. The adenoidal-nasopharyngeal ratio (AN ratio): its validity in selecting children for adenoidectomy. J Laryngol Otol. 101(6):569–573.

Fleiss JL. 1971. Measuring nominal scale agreement among many raters. Psychol Bull. 76(5):378–382.

Guari Q, Wang YJ, Ping B, Li DS, Du JJ, Qin Y, Lu HT, Wan XC, Xiang J. 2019. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J Cancer. 10(20):4876–4882.

Hu X, Zhang QY, Yang JJ, Wang Q, Lei Y, Wu JL. 2019. Photographic analysis and machine learning for diagnostic prediction of adenoid hypertrophy. Paper presented at: Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control; IEEE; New York.

Kim Y, Kim D. 2020. A CNN-based 3D human pose estimation based on projection of depth and ridge data. Pattern Recognit. 106:107462.

Krizhevsky A, Sutskever I, Hinton GE. 2017. Imagenet classification with deep convolutional neural networks. Commun ACM. 60(6):84–90.

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature. 521(7553):436–444.

Lee JH, Yu HJ, Kim MJ, Kim JW, Choi J. 2020. Automated cephalometric landmark detection with confidence regions using Bayesian convolutional neural networks. BMC Oral Health. 20(1):270.

Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, Kanada K, Marinho GD, Gallegos J, Gabriele S, et al. 2020. A deep learning system for differential diagnosis of skin diseases. Nat Med. 26(6):900–908.

Major MP, Flores-Mir C, Major PW. 2006. Assessment of lateral cephalometric diagnosis of adenoid hypertrophy and posterior upper airway obstruction: a systematic review. Am J Orthod Dentofacial Orthop. 130(6):700–708.

Moideen SP, Mytheenkunju R, Nair AG, Mogarnad M, Afroze MKH. 2019. Role of adenoid-nasopharyngeal ratio in assessing adenoid hypertrophy. Indian J Otolaryngol Head Neck Surg. 71(Suppl 1):469–473.

Mummadi SR, Al-Zubaidi A, Hahn PY. 2018. Overfitting and use of mismatched cohorts in deep learning models: preventable design limitations. Am J Respir Crit Care Med. 198(4):544–545.

Noothout JMH, De Vos BD, Wolterink JM, Postma EM, Smeets PAM, Takx RAP, Leiner T, Viergever MA, Isgum I. 2020. Deep learning-based regression and classification for automatic landmark localization in medical images. IEEE Trans Med Imaging. 39(12):4011–4022.

Pereira L, Monyror J, Almeida FT, Almeida FR, Guerra E, Flores-Mir C, Pacheco-Pereira C. 2018. Prevalence of adenoid hypertrophy: a systematic review and meta-analysis. Sleep Med Rev. 38:101–112.

Rajasree R, Columbus CC, Shilaja C. 2020. Multiscale-based multimodal image classification of brain tumor using deep learning method. Neural Comput Appl [epub ahead of print 18 Sep 2020]. doi:10.1007/s00521-020-05332-5

Redmon J, Divvala S, Girshick R, Farhadi A; IEEE. 2016. You only look once: unified, real-time object detection. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition; IEEE; New York.

Ronneberger O, Fischer P, Brox T. 2015. U-net: convolutional networks for biomedical image segmentation. Paper presented at: Medical Image Computing and Computer-Assisted Intervention, Pt III; Springer International Publishing AG; Cham, Switzerland.

Sarao V, Veritti D, Lanzetta P. 2020. Automated diabetic retinopathy detection with two different retinal imaging devices using artificial intelligence: a comparison study. Graefes Arch Clin Exp Ophthalmol. 258(12):2647–2654.

Schwendicke F, Singh T, Lee J-H, Gaudin R, Chaurasia A, Wiegand T, Uribe S, Krois J. 2021. Artificial intelligence in dental research: checklist for authors, reviewers, readers. J Dent. 107:103610.

Shen Y, Li XH, Liang X, Xu H, Li CF, Yu YQ, Qiu BS. 2020. A deep-learning-based approach for adenoid hypertrophy diagnosis. Med Phys. 47(5):2171–2181.

Soldatova L, Otero HJ, Saul DA, Barrera CA, Elden L. 2020. Lateral neck radiography in preoperative evaluation of adenoid hypertrophy. Ann Otol Rhinol Laryngol. 129(5):482–488.

World Health Organization. 2017. Dentists (per 10 000 population) [accessed 2021 Mar 24]. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/dentists-(per-10-000-population).

Yang H, Jo E, Kim HJ, Cha IH, Jung YS, Nam W, Kim JY, Kim JK, Kim YH, Oh TG, et al. 2020. Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs. J Clin Med. 9(6):1839.

Yildirim N, Sahan M, Karslioglu Y. 2008. Adenoid hypertrophy in adults: clinical and morphological characteristics. J Int Med Res. 36(1):157–162.

Yousef M, Hussain KF, Mohammed US. 2020. Accurate, data-efficient, unconstrained text recognition with convolutional neural networks. Pattern Recognit. 108:107482.

Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW, Choi J. 2020. Automated skeletal classification with lateral cephalometry based on artificial intelligence. J Dent Res. 99(3):249–256.