



Deep learning technique to detect craniofacial anatomical abnormalities concentrated on middle and anterior of face in patients with sleep apnea

Shuai He^{a,1}, Yingjie Li^{b,1}, Chong Zhang^c, Zufei Li^a, Yuanyuan Ren^a, Tiancheng Li^{a,*}, Jianting Wang^{a,**}

^a Department of Otolaryngology Head and Neck Surgery, Beijing Chaoyang Hospital, Capital Medical University, China

^b School of Computer Science and Engineering, Beijing Technology and Business University, China

^c Department of Big Data Management and Application, School of International Economics and Management, Beijing Technology and Business University, China

ARTICLE INFO

Keywords:

Explainable deep learning
Craniofacial photographs
Obstructive sleep apnea

ABSTRACT

Objectives: The aim of this study is to propose a deep learning-based model using craniofacial photographs for automatic obstructive sleep apnea (OSA) detection and to perform design explainability tests to investigate important craniofacial regions as well as the reliability of the method.

Methods: Five hundred and thirty participants with suspected OSA are subjected to polysomnography. Front and profile craniofacial photographs are captured and randomly segregated into training, validation, and test sets for model development and evaluation. Photographic occlusion tests and visual observations are performed to determine regions at risk of OSA. The number of positive regions in each participant is identified and their associations with OSA is assessed.

Results: The model using craniofacial photographs alone yields an accuracy of 0.884 and an area under the receiver operating characteristic curve of 0.881 (95% confidence interval, 0.839–0.922). Using the cutoff point with the maximum sum of sensitivity and specificity, the model exhibits a sensitivity of 0.905 and a specificity of 0.941. The bilateral eyes, nose, mouth and chin, pre-auricular area, and ears contribute the most to disease detection. When photographs that increase the weights of these regions are used, the performance of the model improved. Additionally, different severities of OSA become more prevalent as the number of positive craniofacial regions increases.

Conclusions: The results suggest that the deep learning-based model can extract meaningful features that are primarily concentrated in the middle and anterior regions of the face.

1. Introduction

Obstructive sleep apnea (OSA) is a common sleep disorder that is estimated to affect 936 million adults [1]. It is a medical condition characterized by repetitive episodes of cessations and reductions in air flow, which results in intermittent hypoxemia, sympathetic activation, and sleep fragmentation [2]. OSA causes various daytime and nocturnal symptoms, such as excessive sleepiness, snoring, witnessed apneas, and decreased health-related quality of life. Additionally, OSA is negatively associated with cardiovascular, metabolic, and neurocognitive

impairment [3,4]. Therefore, early diagnosis and effective management are essential to prevent adverse health outcomes.

Restriction of craniofacial skeletal dimensions, enlargement of upper airway soft tissue, and excess regional adipose tissues can compromise the upper airway lumen, thus rendering the airway more susceptible to collapse during sleep. Radiological imaging techniques, such as cephalometric radiography, upper airway computed tomography (CT) scanning, and upper airway magnetic resonance imaging (MRI), have been used for two-dimensional (2D) or three-dimensional (3D) analysis to quantify skeletal and soft tissue features and to identify specific

* Corresponding author. Department of Otolaryngology Head and Neck Surgery, Beijing Chaoyang Hospital, Capital Medical University, 8th, Gongti South Road, Chaoyang District, Beijing, 100020, China.

** Corresponding author. Department of Otolaryngology Head and Neck Surgery, Beijing Chaoyang Hospital, Capital Medical University, 8th, Gongti South Road, Chaoyang District, Beijing, 100020, China.

E-mail addresses: litianchenglc@163.com (T. Li), ENT_wjt@163.com (J. Wang).

¹ Co-first authors.

anatomic risk factors for OSA. A larger tongue volume, thicker lateral pharyngeal walls, thicker soft palate, inferiorly posteriorly placed hyoid bone, and shorter mandible, maxilla, and cranial base lengths can increase the risk of sleep apnea [5–7]. However, these imaging modalities are generally limited to research applications owing to their labor intensity, radiation exposure, and the necessity for specific equipment.

Quantitative photographic analysis has been proposed as an alternative assessment technique for elucidating the relationship between craniofacial morphology and sleep apnea. Early studies using standard frontal and profile facial calibration photographs, manual annotation, and measurements demonstrated a range of distinct craniofacial features in patients with OSA. Furthermore, photographic measurements may be clinically useful for OSA risk stratification and screening in both Caucasian and Chinese adults [8,9]. However, manual annotation of images and photogrammetry may be time consuming and exhibit inter-reader variability. To mitigate this issue, a deep learning-based model can be used to capture important features hidden inside a raw image in a self-taught manner, thereby obviating the necessity for manual annotations and feature calculations [10]. In this regard, the input image flows through several layers of neural networks to learn the features of the data and outputs the probability of either class. Previously, we proposed a prediction model based on a deep convolutional neural network and craniofacial photographs from five angles, which can classify individuals with and without OSA (area under the receiver operating characteristic (AUC) curve, 0.9) [11]. However, before the model can be applied to real-world clinical conditions, a certain amount of explainability is required to gain the trust of clinicians and patients. In the current study, we conduct several tests to examine important craniofacial regions and render the deep-learning algorithm more transparent and understandable. We hypothesize that the deep learning-based model can identify features associated with OSA from different craniofacial regions and that the prediction efficiency can be improved by increasing the weight of regional features.

2. Methods

2.1. Participants

Participants older than 18 years who were initially referred to the sleep center for the investigation of OSA at the Department of Otolaryngology Head and Neck Surgery, Beijing Chaoyang Hospital, were invited consecutively and segregated randomly into training, validation, and test sets for model development and evaluation. Exclusion criteria included the presence of congenital or acquired craniofacial abnormalities, central apneas and/or central hypopneas ≥ 5 events/h and $\geq 50\%$ of the respiratory events in the polysomnography (PSG), and excessive facial hair. Written informed consent was obtained from each participant and ethical approval for the study was obtained from the Ethics Committee of Beijing Chaoyang Hospital.

2.2. Data acquisition

The craniofacial photography protocol was identical to that used in our previous study [11]. Briefly, front and profile photographs of the face were captured while the participants stood in a natural head position with their eyes open. The frontal photograph shows the area between the top of the head and ears and the area equivalent to the cricoid cartilage. The profile photograph covered the area between the nose, top, and back of the head, and the area equivalent to the cricoid cartilage. Additionally, clinical variables, including sex, age (years), body mass index (BMI, kg/m^2), and neck circumference (cm) were determined simultaneously.

All participants underwent overnight polysomnography (Embla N7000, RemLogic Eastmed) comprising electroencephalography, electrooculography, electrocardiography, chin electromyography, nasal pressure transduction, thoracic and abdominal respiratory inductive

plethysmography, and pulse oximetry. Sleep and respiratory events were scored by technicians based on the standard criteria of the American Academy of Sleep Medicine [12]. Subsequently, apnea and hypopnea were classified as central, obstructive, or mixed, based on the presence of respiratory effort. The apnea-hypopnea index (AHI) was calculated as the number of apneas and hypopneas per hour of sleep. OSA was defined based on three thresholds: AHI ≥ 5 events/h, AHI ≥ 15 events/h, and AHI ≥ 30 events/h. The algorithm was trained and tested using three threshold values to predict the presence of different levels of OSA severity.

2.3. Model development

Because the original photographs were not of uniform size, they were unsuitable as direct inputs for deep neural networks. Thus, we pre-processed the photographs using OpenCV, which is an open-source computer vision library first released in 2000, using pretrained functions [13]. It can recognize and crop faces in frontal and profile views with their bounding box coordinates as well as generate uniformly sized photographs (512×512 pixels). Different features of the front and profile views complemented each other and provided effective information. The data were classified into OSA cases and controls based on a PSG-derived threshold value of AHI ≥ 5 , 15 or 30 events/h. Subsequently, they were assigned randomly into training, validation, and test sets at a ratio of 7:1:2. The training and validation sets were used to train and identify the optimal parameters of the network layers, and the test data samples were used to evaluate their performance and determine the regions that governed the output. Details regarding this random method are available in the literature [11].

OSA prediction is a binary classification problem in our study, and the classification process include two phases: 1. feature extraction and 2. classification. Deep convolutional neural networks were applied to learn discriminative features. In Phase 1, the important features hidden inside the two craniofacial photographs and four clinical variables (sex, age, BMI, and neck circumference) of each participant were recognized and fused into a dimensional feature map using an image encoder. For the image encoder, ResNet-101 was used as the backbone, which comprised four different blocks. The convolutional layers of the blocks were characterized by various channel scales, kernel sizes, and strides. The lower layers learned the image features, such as the basic shapes and edges, which were then further processed by subsequent higher layers in a more abstract manner. During this process, more complex features were identified in the input craniofacial photographs, whereas irrelevant features were filtered out. After the convolutional layer, the average and maximum pooling layers were incorporated to reduce the dimensions of the feature maps, thus reducing the number of parameters. In Phase 2, a high-level feature map was passed through the fully connected layers and a rectified linear unit (ReLU) to calculate the probabilities of disease occurrence via a softmax function (Supplementary material). Fig. 1 shows the deep neural network architecture for OSA classification using craniofacial photographs. When comparing the low-level features of the first layers with respect to the high-level features of the final layers, distinctive facial traits tend to emerge while identity-related features disappear.

The deep learning-based model was initially fitted to the training set; meanwhile, the validation set provided an evaluation of the model and enabled hyperparameter tuning. The prediction error was calculated by comparing the predicted probability with the ground truth based on PSG, and the parameters were adjusted accordingly to decrease the error. In our study, stochastic gradient descent was used for model training via 300 epochs using the Adam optimizer at a base learning rate of 10^{-3} , batch size of 64, momentum of 0.9, and weight decay of 0.0005. After the training was completed, the algorithm was locked.

The development of the model and preprocessing procedures were implemented using Python 3.7.16 and PyTorch 1.12.1.

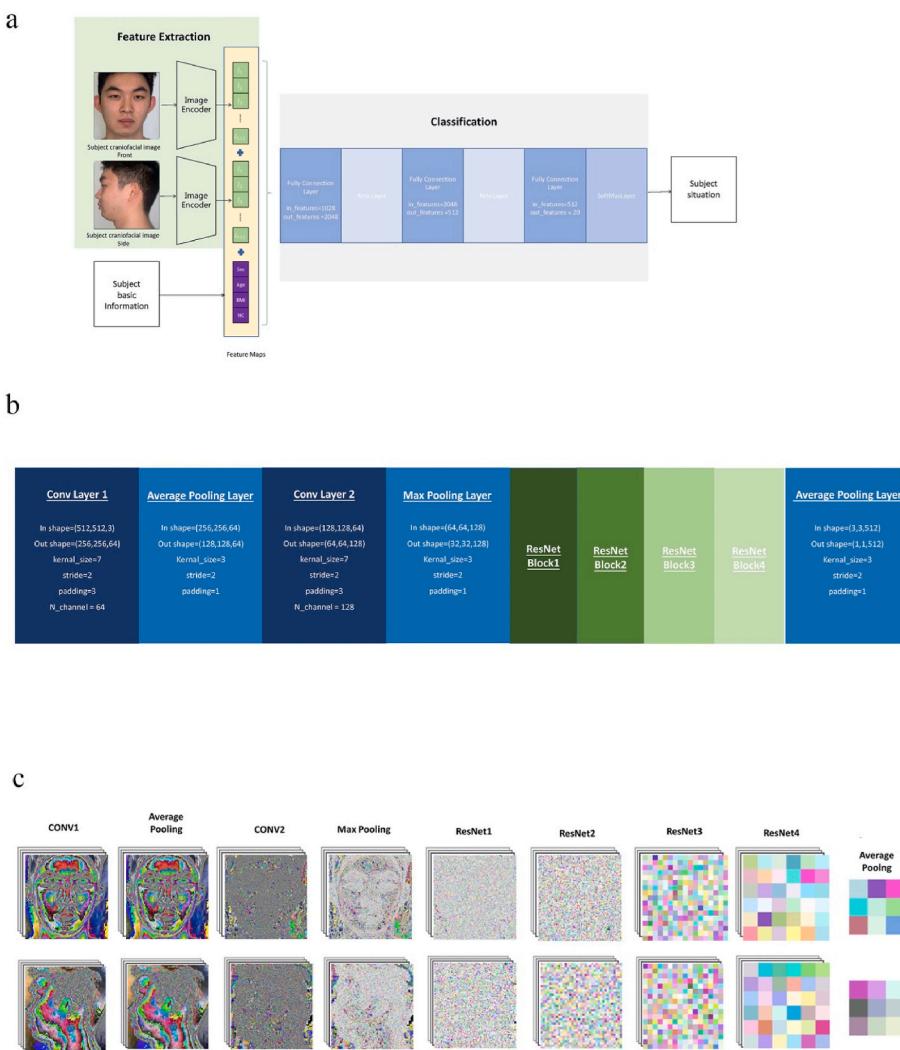


Fig. 1. Illustration of deep learning-based classifier. (a) Raw craniofacial photographs converted into a vector of features. Feature map (1028×1) passed through fully connected layers and finally output class-based likelihood of participant. (b) Architecture of image encoder. (c) A snapshot of a craniofacial photograph passing through the network.

2.4. Model evaluation and statistical analysis

The model performance was evaluated on the test set by calculating the accuracy, AUC curve, sensitivity, and specificity at the cutoff point selected from the receiver operating characteristic curve.

The data are expressed as mean \pm standard deviation or median (interquartile range) for continuous variables and as percentages for categorical variables. The Student's t-test and Mann-Whitney U test were performed to compare normal and skewed continuous variables, and the chi-square test was performed for categorical variables. Statistical significance was set at $P < 0.05$. Statistical analysis was performed using SPSS (version 22.0).

2.5. Discussion regarding model

Many deep learning-based models cannot explicitly explain their predictions and provide no information regarding facial regions that facilitate classification. Hence, several tests were conducted to better understand the model and improve its accuracy.

2.6. Photographic occlusion

We shielded nine sections of the frontal and profile craniofacial

photographs in the test set with a black patch (RGB, (0,0,0)), including the forehead, bilateral eyes, nose, bilateral cheeks, mouth and chin, ear, preauricular area, temporal-occipital area, and neck (Fig. 2). Subsequently, we input the occluded craniofacial photographs into the model and measured the contribution of the different craniofacial regions to the model based on the decrease in the AUC as compared with those of the original photographs.

Additionally, we calculated the mean values of the contributions from the frontal and profile photographs. The regions whose contributions were not above the average value were occluded by black patches to increase OSA-related information in the network's view. Newly weighted craniofacial photographs were generated by linearly fusing occluded and original photographs. We examined whether the weighted photographs could improve the model performance.

2.7. Visual explanation

To visually explain the areas of the image focused on the network, we generated saliency maps to show important craniofacial features in the photographs. The saliency mapping technique uses a backpropagation-based approach to highlight pixels that exert a greater effect on a decision [14]. We created saliency maps based on the average of 30 participants with the highest and lowest predictive probabilities.

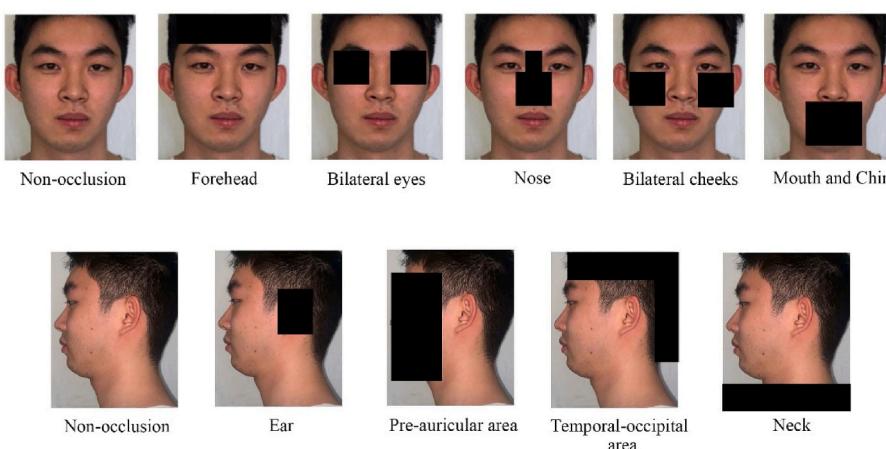


Fig. 2. Example of occluding nine different craniofacial regions.

2.8. Positive regions and severity of OSA relationship

To further examine the association between craniofacial characteristics and OSA, we assessed the relationship between positive craniofacial regions and OSA severity in the test set. As mentioned above, nine regions (e.g., neck, eyes, nose) of the participants' frontal and profile photographs were occluded, and if the occlusion decreased the predicted probability compared with non-occlusion situation, then the region was defined a positive region. For each participant, we determined the number of positive regions, segregated them into two groups (0–4 or 5–9 positive regions), and compared the prevalence of OSA using different AHI thresholds.

3. Results

3.1. Participant characteristics

Five hundred and thirty participants who satisfied the inclusion criteria were enrolled in this study, of whom 23.3% were female. In general, participants with an AHI between 0 and 5 events/h, between 5 and 15 events/h, between 15 and 30 events/h, and ≥ 30 events/h were 21.7%, 19.7%, 14.9%, and 43.7%, respectively. The mean age, BMI, and neck circumference were 40.55 ± 11.86 years, 27.03 ± 4.42 kg/m², and

39.83 ± 4.18 cm, respectively.

The participants were randomly segregated into training (N = 371), validation (N = 53), and test (N = 106) sets. Table 1 summarizes a comparison of the characteristics of the study populations in different sets. No statistically significant differences were indicated in terms of age, sex, BMI, AHI, and OSA status among the three sets.

3.2. Model performance

The AUC, sensitivity, specificity, and accuracy of the test set are summarized in Table 2. When the OSA was defined by AHI ≥ 5 events/h, the predictive model derived from the photographs achieved the best accuracy and AUC of 0.884 and 0.881, respectively (95% confidence interval [CI], 0.839–0.922). Using the cutoff point with the maximum sum of sensitivity and specificity, the model indicated a sensitivity of 0.905 and a specificity of 0.941. The results demonstrated that, relative to an AHI threshold of 5 events/h, the accuracy and AUC decreased to 0.878 and 0.872 (95%CI, 0.818–0.924), respectively, when an AHI threshold of 15 events/h was used, and to 0.874 and 0.853 (95%CI, 0.815–0.904), respectively, when an AHI threshold of 30 events/h was used. The accuracy of the model in predicting the presence of OSA increased when using a combination of photographs and clinical variables.

Table 1
Participant characteristics.

	Variables	Training (N = 371)	Validation (N = 53)	P-value ^a	Test (N = 106)	P-value ^b
AHI ≥ 5/h	Sex (M:F)	283:88	40:13	0.51	84:22	0.79
	Age (years)	40.23 ± 11.85	41.53 ± 12.53	0.51	41.01 ± 11.76	0.60
	NC (cm)	39.79 ± 4.23	39.89 ± 3.14	0.87	39.93 ± 4.51	0.95
	BMI (kg/m ²)	26.75 ± 4.47	27.07 ± 3.16	0.51	26.81 ± 4.89	0.78
	Participants with OSA (%)	77.9	79.2	0.49	78.3	0.97
	AHI (events/h)	22.1 (7.1, 53.5)	29.1 (8.6, 57.4)	0.49	20.6 (7.3, 51.2)	0.72
AHI ≥ 15/h	Sex (M:F)	289:82	39:14	0.49	79:27	0.41
	Age (years)	40.32 ± 11.60	40.42 ± 12.91	0.99	41.25 ± 12.42	0.85
	NC (cm)	39.86 ± 4.18	39.65 ± 4.11	0.93	39.80 ± 4.27	0.90
	BMI (kg/m ²)	26.71 ± 4.47	26.90 ± 4.23	0.56	27.03 ± 4.46	0.71
	Participants with OSA (%)	58.5	58.5	0.56	58.5	0.53
	AHI (events/h)	22.1 (7.2, 53.7)	20.7 (5.9, 38.8)	0.45	22.9 (7.2, 55.4)	0.60
AHI ≥ 30/h	Sex (M:F)	283:88	38:15	0.28	86:20	0.38
	Age (years)	40.21 ± 11.91	41.43 ± 10.62	0.41	41.13 ± 12.45	0.65
	NC (cm)	39.81 ± 4.37	40.11 ± 3.63	0.56	39.74 ± 3.76	0.84
	BMI (kg/m ²)	26.86 ± 4.72	26.83 ± 3.60	0.72	26.54 ± 3.76	0.92
	Participants with OSA (%)	43.4	45.3	0.45	43.4	0.96
	AHI (events/h)	21.9 (6.4, 54.5)	19.5 (8.2, 52.1)	0.82	31.31 ± 25.92	0.88

Data were presented as mean \pm standard deviation or median (interquartile range) or percentage (%).

BMI, body mass index; NC, neck circumstance; OSA, obstructive sleep apnea; AHI, apnea-hypopnea index; M, male; F, female.

^a P value was obtained by comparison of the training and validation sets.

^b P value was obtained by comparison of the training and test sets.

Table 2

The model's performance when using three AHI thresholds (≥ 5 , 15, and 30 events/h).

	Sensitivity	Specificity	Accuracy	AUC (95% CI)
AHI≥ 5				
Photographs	0.905	0.941	0.884	0.881(0.839-0.922)
Photographs and clinical variables	0.875	0.967	0.901	0.900(0.841-0.944)
AHI≥ 15				
Photographs	0.600	0.978	0.878	0.872(0.818-0.924)
Photographs and clinical variables	0.810	0.953	0.881	0.889(0.850-0.928)
AHI≥ 30				
Photographs	0.895	0.943	0.874	0.853(0.815-0.904)
Photographs and clinical variables	0.850	0.953	0.887	0.881(0.809-0.948)

AUC, Area Under the Receiver Operating Characteristic Curve; CI, Confidence Interval; AHI, Apnea-Hypopnea Index.

3.3. Occlusion results

We performed an occlusion test based on the model to predict OSA at an AHI threshold of 5 events/h. After occluding each of the nine craniofacial sections, the most decrease in the AUC of the frontal view was indicated for the bilateral eyes ($\Delta\text{AUC} = 0.082$), followed by the nose ($\Delta\text{AUC} = 0.061$), and mouth and chin ($\Delta\text{AUC} = 0.046$). Meanwhile, the most decrease in the AUC of the profile view was the pre-auricular area ($\Delta\text{AUC} = 0.113$), followed by the ear ($\Delta\text{AUC} = 0.074$). Changes in the AUCs after occluding the different craniofacial regions are plotted in Fig. 3.

The mean contribution values by the front and profile views were 0.045 and 0.054, respectively, and five regions indicated values above the mean contribution. This allowed us to create a weighted craniofacial photograph for each participant and analyze the effect of changing the input features on the output of the model. The results demonstrated that the prediction performance improved for both the frontal and profile-

weighted craniofacial photographs. The model exhibited the largest AUC of 0.953 when clinical variables were combined with weighted craniofacial photographs (Figs. 4 and 5).

3.4. Visualization of important features

The different features between participants with and without OSA were visualized using saliency maps. Fig. 6(a) shows that the model focuses on the midface, jaw, ear, and neck when classifying individuals with OSA. As shown in Fig. 6(b), the mouth area is emphasized in the classification of individuals without OSA.

3.5. Relationship between craniofacial regions and OSA

Compared with the 0–4 positive region group, the 5–9 positive region group with different apnea severities ($P < 0.05$) indicated a higher prevalence of OSA (Fig. 7). The different severities of OSA became more prevalent as the number of positive craniofacial regions increased.

4. Discussion

In this study, our deep learning-based model using craniofacial photographs facilitated the detection of OSA. The results showed that the highest sensitivity, specificity, accuracy, and AUC of the model were 90.5%, 94.1%, 88.4%, and 0.881, respectively, when $\text{AHI} \geq 5$ events/h was used to define the presence of OSA. Furthermore, the classification accuracy for detecting OSA decreased under the other two typically used cutoffs, i.e., 15 and 30 events/h. The deterioration in performance with the increase in AHI threshold values was similarly reported in another study, which adopted linear and geodesic measurements obtained from facial photographs to develop a predictive algorithm [15]. These results support the notion that patients with different OSA severity levels share a mutual craniofacial phenotype.

Craniofacial differences between participants with and without OSA can be identified using 2D or 3D photographic analyses, and measurements of craniofacial features can be used to predict OSA. A logistic regression model using four measurements (mandibular length, face width, eye width, and cervicomental angle) correctly classified 76.1% of

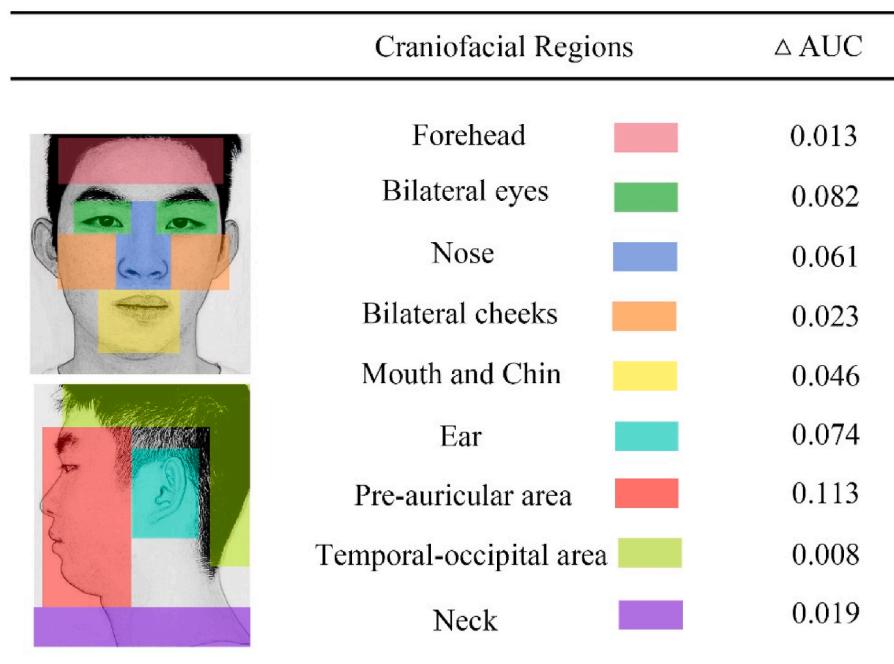


Fig. 3. ΔAUC was defined as the decrease in model performance after occluding a craniofacial region. The larger ΔAUC was obtained, the more important the occluded region was.

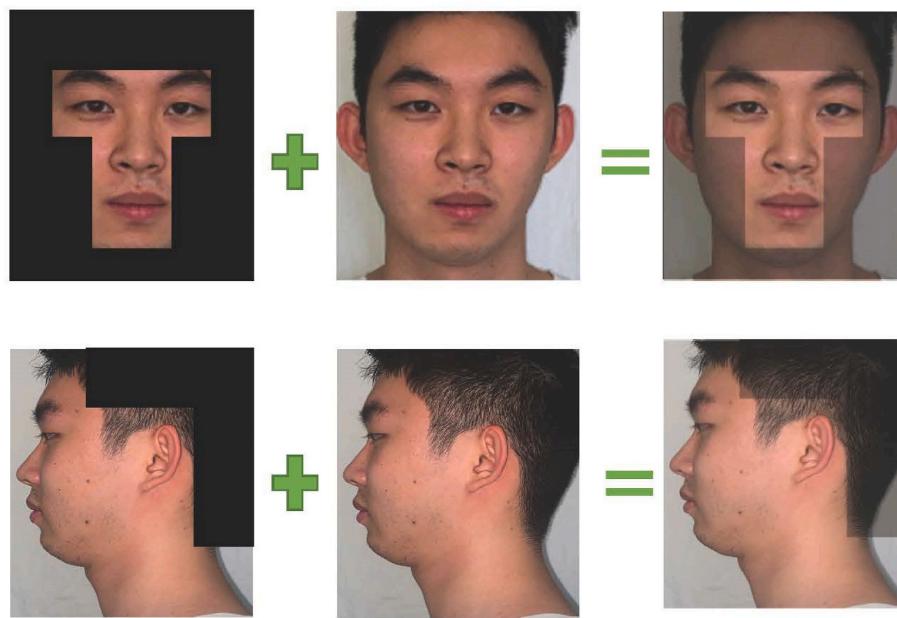


Fig. 4. Example of a regional weighed photographs.

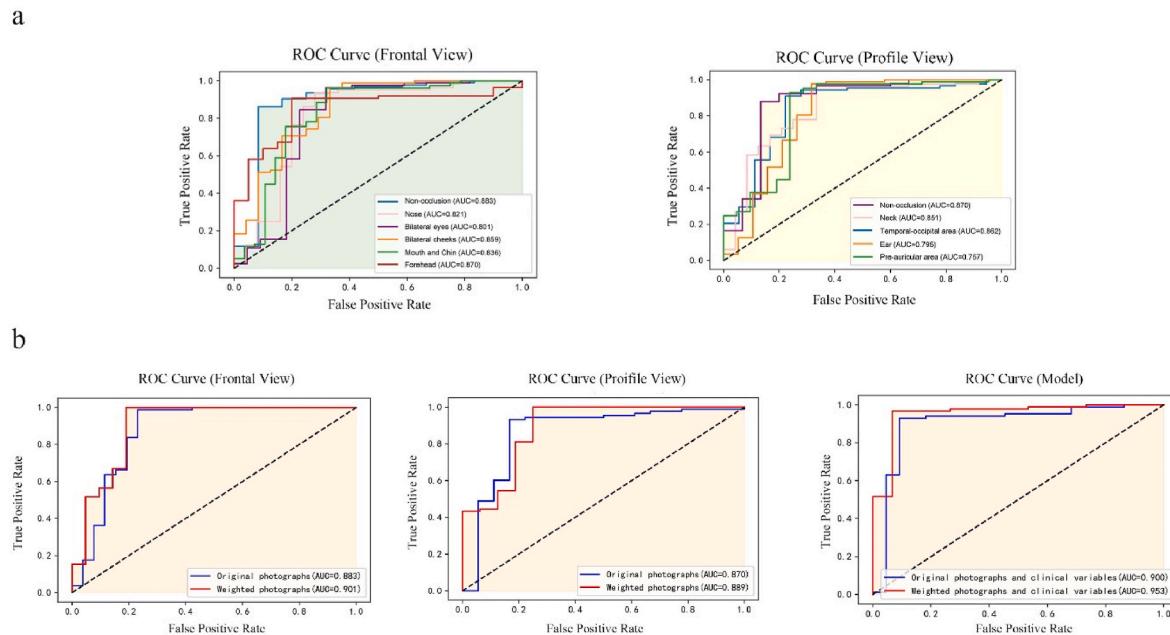


Fig. 5. (a) The receiver operating characteristics curves (ROC curves) of the model using different occluded regions of craniofacial photographs. (b) The ROC curves of the model using the weighted photographs alone, and using both weighted photographs and clinical variables.

participants and achieved an AUC of 0.82 [8]. In another study, photographic measurements of the oropharyngeal structures (tongue area, uvula area, frenulum length, and retrusion distance) were used to establish a prediction model. It classified 82.7% of the participants correctly and indicated a sensitivity of 85.6%, a specificity of 84.3% at the best cutoff point, and an AUC of 0.90 [16]. Additionally, 3D photographs were used to better delineate the shape and contour of the face and provide a precise analysis. Measurements of craniofacial linear distances, angles, areas, and volumes between 3D photographs and CT scans showed stronger agreement than those between 2D photographs and CT scans; however, the relationship between the measurements with AHI was weak for both 2D and 3D photographs [17]. Recently, two studies combined various classifiers (k-nearest neighbors, extreme

gradient boosting, and convolutional neural networks) with 3D craniofacial scans to predict AHI values and OSA presence. The results showed the accuracy and AUC were approximately 67% and 0.70, respectively [18,19]. It appears that 3D photographs of the face did not further facilitate the identification of patients with OSA but might be a clinically useful method for evaluating craniofacial profiles. Meanwhile, 2D photographs are more readily available and inexpensive, thus rendering them more accessible to a larger population. Therefore, we developed an OSA prediction method based on 2D craniofacial analysis.

The results of the current study showed that the bilateral eyes, nose, mouth and chin, ear, and pre-auricular area contributed more to our model than other craniofacial regions. The important regions generally focus on the middle and anterior facial regions. However, the

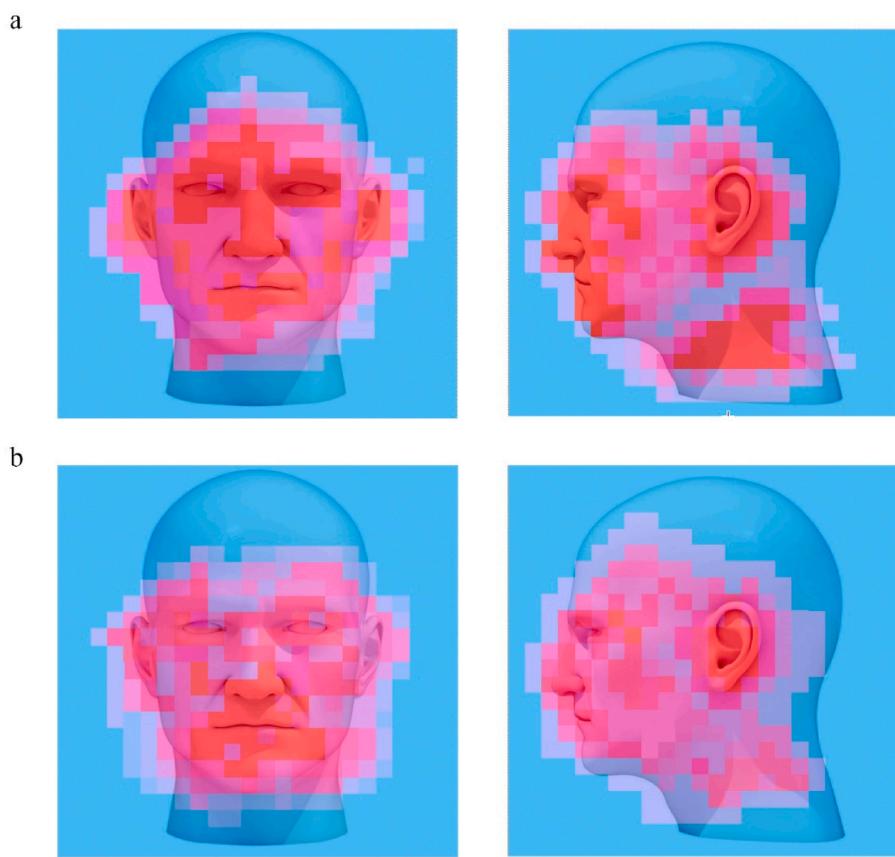


Fig. 6. (a) The average saliency maps with the highest predicted probabilities. (b) The average saliency maps with the lowest predicted probabilities.

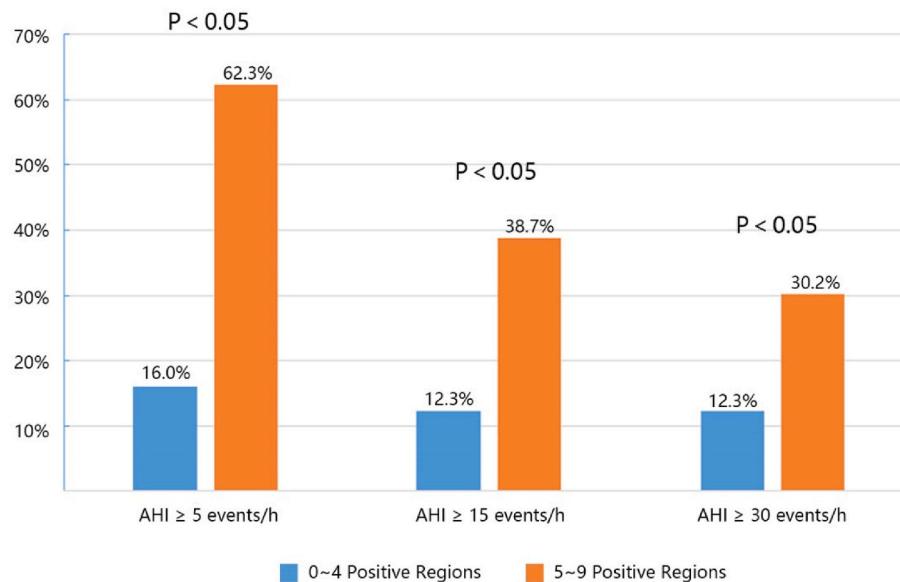


Fig. 7. The relationship between the number of positive regions and OSA prevalence.

performance of the model improved when the weight of the corresponding regions was increased, thus indicating the presence of OSA-related risk factors in these regions. This result is consistent with the classical photographic measurement of patients with OSA observed by clinicians as well as with those identified in prior imaging studies. For example, patients with OSA have larger intercanthal and nose widths, shorter and more retruded jaws, and wider faces as compared with

patients without OSA. Notably, some measures may be protective against OSA in the presence of clinical risk factors such as older age, obesity, and male sex [20,21]. Studies using CT scans to examine the craniofacial structures demonstrated that nasal width, jaw reposition angle, and frontomaxillary suture distance were correlated with AHI, and that the intermolar distance as well as the posterior nasal spine to hyoid distance were related to upper airway collapsibility [22,23].

These significant measurements were included in the regions identified using the deep-learning technique. The visualization of the photographic features was similar to that of the occlusion test, which showed that patients with OSA may exhibit abnormal craniofacial complexes across the cranial base, maxilla, and mandible. The deep learning-based model should be able to reveal important risk factors related to OSA from the highlighted areas. Compared with the front view, the profile view may provide additional valuable information in the neck area. We speculate that not only the neck size, but also the positional relationship between the neck and mandible can be reflected in the profile view. Both the occlusion and visual display maps demonstrated that the ears were a potential risk area, although few studies have been conducted to investigate the relationship between this area and OSA. Our model may have learned some features that are associated with OSA but are beyond human perception or understanding. Additionally, we identified a relationship between the number of positive craniofacial regions considered meaningful for OSA detection and OSA prevalence. An individual that shows more positive regions is at a greater risk of having OSA. Collectively, these tests explain the scientific basis and reliability of the deep learning-based model.

Another approach of applying craniofacial photographs is to define disease phenotyping and investigate anatomical disharmony relevant to OSA pathogenesis. Sutherland et al. provided evidence that surface facial dimensions (measured from photographs or MRI), such as the midface width and lower face height, reflected the size of the upper airway structures [24,25]. Although we used deep learning to determine the risk areas related to OSA, the method adopted could not differentiate among skeletal structures, soft tissues, and regional adipose tissues in a sophisticated manner; thus, it cannot truly assess the anatomical imbalance in patients with OSA. More research is required to analyze the association between the extracted representative features hidden inside craniofacial photographs and the upper airway structures, or other physiological traits such as pharyngeal critical pressure, pharyngeal opening pressure, and loop gain. In addition to anatomical risk factors, patients with OSA may have a sleepful or less attractive facial appearance. Whether the deep learning algorithm extracts specific features that can distinguish changes in facial appearance requires further investigation.

5. Conclusion

Craniofacial photographic analysis using deep learning may enable automatic and rapid OSA risk stratification in outpatient clinics or may serve as a tool for OSA screening. The high-stakes decision-making of deep learning-based methods is intertwined with medicine, and our results showed that the proposed model successfully captured OSA-associated craniofacial features.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Shuai He: Conceptualization, Methodology, Data collection, Statistical analysis, Writing – original draft. **Yingjie Li:** Conceptualization, Software, Technical development and support, Writing – original draft. **Chong Zhang:** Software, Technical development and support. **Zufei Li:** Conceptualization, Clinical support. **Yuanyuan Ren:** Conceptualization, Clinical support. **Tiancheng Li:** Conceptualization, Design, Methodology, Writing – review & editing. **Jianting Wang:** Conceptualization, Design, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge their participants and colleagues for their support of this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.sleep.2023.09.025>.

References

- [1] Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MSM, Morrell MJ, Nunez CM, Patel SR, Penzel T, Pepin JL, Peppard PE, Sinha S, Tiflik S, Valentine K, Malhotra A. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med* 2019;7:687–98.
- [2] Veasey SC, Rosen IM. Obstructive sleep apnea in adults. *N Engl J Med* 2019;380:1442–9.
- [3] Werli KS, Otuyama LJ, Bertolucci PH, Rizzi CF, Guilleminault C, Tufik S, Poyares D. Neurocognitive function in patients with residual excessive sleepiness from obstructive sleep apnea: a prospective, controlled study. *Sleep Med* 2016;26:6–11.
- [4] Dewan NA, Nieto FJ, Somers VK. Intermittent hypoxemia and OSA. *Chest* 2015;147:266–74.
- [5] Schwab R, Pasirstein M, Pierson R, Mackley A, Hachadoorian R, Arens R, Maislin G, Pack AI. Identification of upper airway anatomic risk factors for obstructive sleep apnea with Volumetric magnetic resonance imaging. *Am J Respir Crit Care Med* 2003;168:522–30.
- [6] Chi L, Comyn FL, Mitra N, Reilly MP, Wan F, Maislin G, Chmiewski L, Thorne-FitzGerald MD, Victor UN, Pack AI. Identification of craniofacial risk factors for obstructive sleep apnea using three-dimensional MRI. *Eur Respir J* 2011;38:348–58.
- [7] Schorr F, Kayamori F, Hirata RP, Danzi-Saunders NJ, Gebrim EM, Moriya HT, Malhotra A, Lorenzi-Filho G, Genta PR. Different craniofacial characteristics predict upper airway collapsibility in Japanese-Brazilian and white men. *Chest* 2016;149:737–46.
- [8] Lee RW, Petocz P, Prvan T, Chan AS, Grunstein RR, Cistulli PA. Prediction of obstructive sleep apnea with craniofacial photographic analysis. *Sleep* 2009;32:46–52.
- [9] Sutherland K, Lee RW, Petocz P, Chan TO, Ng S, Hui DS, Cistulli PA. Craniofacial phenotyping for prediction of obstructive sleep apnoea in a Chinese population. *Respirology* 2016;21:1118–25.
- [10] Chen X, Wang X, Zhang K, Fung KM, Thai TC, Moore K, Mannel RS, Liu H, Zheng B, Qiu Y. Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal* 2022;79:102444.
- [11] He S, Su H, Li Y, Xu W, Wang X, Han D. Detecting obstructive sleep apnea by craniofacial image-based deep learning. *Sleep Breath* 2022;26:1885–95.
- [12] Berry RB, Budhiraja R, Gottlieb DJ, Gozal D, Iber C, Kapur VK, Marcus CL, Mehra R, Parthasarathy S, Quan SF, Redline S, Strohl KP, Davidson Ward SL, Tangredi MM, American Academy of Sleep Medicine. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. Deliberations of the sleep apnea definitions task force of the American Academy of sleep medicine. *J Clin Sleep Med* 2012;8:597–619.
- [13] Levin AA, Klimov DD, Nechunaev AA, Prokhorchenko LS, Mishchenkov DS, Nosova AG, Astakhov DA, Poduraev YV, Panchenkov DN. Assessment of experimental OpenCV tracking algorithms for ultrasound videos. *Sci Rep* 2023;13:6765.
- [14] Van der Velden BHM, Kuijff HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal* 2022;79:102470.
- [15] Eastwood P, Gilani SZ, McArdle N, Hillman D, Walsh J, Maddison K, Goonewardene M, Mian A. Predicting sleep apnea from three-dimensional face photography. *J Clin Sleep Med* 2020;16:493–502.
- [16] He S, Li Y, Xu W, Kang D, Li H, Wang C, Ding X, Han D. The predictive value of photogrammetry for obstructive sleep apnea. *J Clin Sleep Med* 2021;17:193–202.
- [17] Lin SW, Sutherland K, Liao YF, Cistulli PA, Chuang LP, Chou YT, Chang CH, Lee CS, Li LF, Chen NH. Three-dimensional photography for the evaluation of facial profiles in obstructive sleep apnoea. *Respirology* 2018;23:618–25.
- [18] Monna F, Ben Messaoud R, Navarro N, Baillieul S, Sanchez L, Loiodice C, Tamisier R, Joyeux-Faure M, Pépin JL. Machine learning and geometric morphometrics to predict obstructive sleep apnea from 3D craniofacial scans. *Sleep Med* 2022;95:76–83.
- [19] Hanif U, Leary E, Schneider L, Paulsen R, Morse AM, Blackman A, Schweitzer P, Kushida CA, Liu S, Jennnum P, Sorensen H, Mignot E. Estimation of apnea-hypopnea index using deep learning on 3-D craniofacial scans. *IEEE J Biomed Health Inform* 2021;25:4185–94.

- [20] Lee RW, Chan AS, Grunstein RR, Cistulli PA. Craniofacial phenotyping in obstructive sleep apnea—a novel quantitative photographic approach. *Sleep* 2009; 32:37–45.
- [21] Rizzatti FG, Mazzotti DR, Mindel J, Maislin G, Keenan BT, Bittencourt L, Chen NH, Cistulli PA, McArdle N, Pack FM, Singh B, Sutherland K, Benediktsdottir B, Fietze I, Gislason T, Lim DC, Penzel T, Sanner B, Han F, Li QY, Schwab R, Tufik S, Pack AI, Magalang UJ. Defining extreme phenotypes of OSA across international sleep centers. *Chest* 2020;158:1187–97.
- [22] Gurgel M, Cevidaner L, Pereira R, Costa F, Ruellas A, Bianchi J, Cunali P, Bittencourt L, Junior CC. Three-dimensional craniofacial characteristics associated with obstructive sleep apnea severity and treatment outcomes. *Clin Oral Invest* 2022;26:875–87.
- [23] Thuler E, Seay EG, Woo J, Lee J, Jafari N, Keenan BT, Dedhia RC, Schwartz AR. Transverse maxillary deficiency predicts increased upper airway collapsibility during drug-induced sleep endoscopy. *Otolaryngol Head Neck Surg* 2023;7:258.
- [24] Sutherland K, Schwab RJ, Maislin G, Lee RW, Benediktsdottir B, Pack AI, Gislason T, Juliussen S, Cistulli PA. Facial phenotyping by quantitative photography reflects craniofacial morphology measured on magnetic resonance imaging in Icelandic sleep apnea patients. *Sleep* 2014;37:959–68.
- [25] Lee RW, Sutherland K, Chan AS, Zeng B, Grunstein RR, Darendeliler MA, Schwab RJ, Cistulli PA. Relationship between surface facial dimensions and upper airway structures in obstructive sleep apnea. *Sleep* 2010;33:1249–54.