

Explainable artificial intelligence to quantify adenoid hypertrophy-related upper airway obstruction using 3D Shape Analysis

Claudia Trindade Mattos ^{a,b,*}, Lucie Dole ^c, Sergio Luiz Mota-Júnior ^{a,f}, Adriana de Alcantara Cury-Saramago ^a, Jonas Bianchi ^d, Heesoo Oh ^d, Karine Evangelista ^e, José Valladares-Neto ^e, Antonio Carlos de Oliveira Ruellas ^f, Juan Carlos Prieto ^c, Lucia Helena Soares Cividanes ^b

^a Department of Orthodontics, Faculdade de Odontologia, Universidade Federal Fluminense, Rua Mário Santos Braga, 30, 2^o andar, sala 214, Centro, Niterói, RJ, CEP 24020-140, Brazil

^b Department of Orthodontics and Pediatric Dentistry, School of Dentistry, University of Michigan, 1011 North University Avenue, Ann Arbor, Michigan, 48104, USA

^c Department of Psychiatry, School of Medicine, University of North Carolina, 333 S. Columbia Street, Suite 304, MacNider Hall, Chapel Hill, North Carolina, 27514, USA

^d Department of Orthodontics, University of the Pacific, Arthur A. Dugoni School of Dentistry, 155 Fifth Street, Third Floor, San Francisco, California, 94103, USA

^e Department of Orthodontics, School of Dentistry, Universidade Federal de Goiás, Avenida Universitária esquina com 1a Avenida, Goiânia, S/N, 74605-220, Brazil

^f Department of Orthodontics and Pediatric Dentistry, Faculdade de Odontologia, Universidade Federal do Rio de Janeiro, Rua Professor Rodolfo Paulo Rocco, 325, Ilha do Fundão, Rio de Janeiro, RJ, CEP 21941-617, Brazil

ARTICLE INFO

Keywords:

Artificial Intelligence (AI)
3D shape analysis
Adenoid hypertrophy
Upper airway obstruction
Cone-beam computed tomography (CBCT)

ABSTRACT

Objectives: To develop and validate an explainable Artificial Intelligence (AI) model for classifying and quantifying upper airway obstruction related to adenoid hypertrophy using three-dimensional (3D) shape analysis of cone-beam computed tomography (CBCT) scans.

Methods: 400 CBCT scans of patients aged 5–18 years were analyzed. Nasopharyngeal airway obstruction (NAO) ratio was calculated to label scans into four grades of obstruction severity, used as the ground truth. Upper airway surface meshes were used to train a deep learning model combining multiview and point-cloud approaches for 3D shape analysis and obstruction severity classification and quantification. Surface Gradient-weighted Class Activation Mapping (SurfGradCAM) generated explainability heatmaps. Performance was evaluated using area under the curve (AUC), precision, recall, F1-score, mean absolute error, root mean squared error, and correlation coefficients.

Results: The explainable AI model demonstrated strong performance in both classification and quantification tasks. The AUC values for the classification task ranged from 0.77 to 0.94, with the highest values of 0.88 and 0.94 for Grades 3 and 4, respectively, indicating excellent discriminative ability for identifying more severe cases of obstruction. The SurfGradCAM-generated heatmaps consistently highlighted the most relevant regions of the upper airway influencing the AI's decision-making process. In the quantification task, the regression model successfully predicted the NAO ratio, with a strong correlation coefficient of 0.854 ($p < 0.001$) and $R^2 = 0.728$, explaining a substantial proportion of the variance in NAO ratios.

Conclusions: The proposed explainable AI model, using 3D shape analysis, demonstrated strong performance in classifying and quantifying adenoid hypertrophy-related upper airway obstruction in CBCT scans.

Clinical significance: This AI model provides clinicians with a reliable, automated tool for standardized adenoid hypertrophy assessment. The model's explainable nature enhances clinical confidence and patient communication, potentially improving diagnostic workflow and treatment planning.

* Corresponding author at: Faculdade de Odontologia, Disciplina de Ortodontia, Rua Mário Santos Braga, 30, 2^o andar, sala 214, Centro, Niterói, RJ, CEP 24020-140, Brazil.

E-mail addresses: claudiamattos@id.uff.br (C.T. Mattos), lucie.dole@med.unc.edu (L. Dole), motaserj@id.uff.br (S.L. Mota-Júnior), adriancury@id.uff.br (A.A. Cury-Saramago), jbianchi@pacific.edu (J. Bianchi), hoh@pacific.edu (H. Oh), karineevangelista@ufg.br (K. Evangelista), jvalladares@uol.com.br (J. Valladares-Neto), antonioruellas@yahoo.com.br (A.C.O. Ruellas), jprieto@med.unc.edu (J.C. Prieto), luiciacev@umich.edu (L.H.S. Cividanes).

1. Introduction

Adenoid hypertrophy in children and adolescents is a well-established cause of partial or significant nasopharyngeal airway obstruction and a risk factor for pediatric obstructive sleep apnea (OSA) [1,2]. Orthodontists play an important role in screening for OSA [3,4] and mouth breathing [5,6], as these conditions can influence normal craniofacial growth and development [7], potentially impairing quality of life[8,9].

Image analysis is a valuable tool [10] for screening upper airway obstruction. While lateral cephalograms or cone beam computed tomography (CBCT) scans may be available for diagnostic purposes, maximizing their utility is essential for patient benefit [3]. Although CBCT scans are not requested solely for upper airway analysis, three-dimensional (3D) imaging provides a more accurate assessment of upper airway morphology and minimum cross-sectional areas compared to two-dimensional (2D) imaging [3,11]. These parameters are extremely important in evaluating obstruction. Therefore, investing in accessible tools for 3D upper airway analysis is an important goal for comprehensive assessment of these images when they are obtained.

Recently, artificial intelligence-aided diagnosis has gained significant research interest due to its potential for time-saving and enhanced data processing, which strengthens decision-making. Artificial intelligence (AI) models have been developed and tested to classify various conditions in dentistry with good precision and accuracy using 2D radiographs or photographs [12–16]. However, few AI models for classification tasks in dentistry using 3D imaging, particularly those considering shape or morphology, have been published[17,18]. Moreover, to ensure trustworthiness and future user adoption, clinicians need explainability models or maps to adequately supervise and understand AI decision-making in classification tasks [17,19].

A few studies have reported on adenoid hypertrophy diagnosis using AI models, primarily in lateral cephalograms [20–23]. A recent study proposed an automatic diagnosis method based on deep learning for adenoid hypertrophy detection on CBCT scans, with promising results [18]. However, although pioneering in this regard, their model considered only a binary classification of either presence or absence of hypertrophy. Additionally, no explainability for interpretable AI was assessed.

To address these limitations, the present study aims to develop and validate the first explainable AI model for classifying and quantifying upper airway obstruction related to adenoid hypertrophy using 3D shape analysis of CBCT scans. Specifically, the model classifies cases into four severity grades and introduces a novel quantification approach by predicting the percentage of obstruction. Additionally, it generates explainability heatmaps that make the AI decision-making process interpretable and trustworthy for clinicians.

2. Methods

This study was approved by the Institutional Review Board (IRB) of the University of Michigan School of Dentistry (HUM00251245). All methods were performed in accordance with the Helsinki Declaration guidelines on medical protocols and ethics. The IRB waived the requirement for informed consent. The sample comprised a secondary data analysis of de-identified cone-beam computed tomography (CBCT) scans obtained from patients treated at the dental schools of four different university centers: University of Michigan, Fluminense Federal University, University of the Pacific, and Goiás Federal University. All CBCT scans were acquired using iCat units (Imaging Sciences International, Hatfield, PA) with similar protocol across the different centers: 3.8 mA, 120 kV, exposure times ranging from 20 to 40 s, voxel size of 0.3 mm³ to 0.4 mm³, and field of view varying from 13 × 16 cm to 17 × 23 cm, depending on the patient's age and size.

2.1. Eligibility criteria

Patients aged 5 to 18 years with large field-of-view CBCT scans taken for clinical purposes were selected. Clinical purposes included diagnostic needs related to impacted or ectopic teeth and severe craniofacial disorders. Scans with artifacts produced by orthodontic appliances or metallic dental crowns were excluded.

2.2. 3D image analysis

Three-dimensional analysis was performed using InVivo 3D Imaging Software, version 6.5 (Anatomage, Santa Clara, CA, USA), and two open-source packages: 3D Slicer, version 5.6.2 (<http://www.slicer.com>) [24], and ITK-SNAP, version 4.0.2 (<http://www.itksnap.org>) [25].

2.3. Ground truth labeling

A quantitative ratio of the nasopharynx airway obstruction (NAO ratio) corresponding to the adenoid region was calculated to serve as the ground truth for labeling cases. The same researcher (CTM), an orthodontist with over ten years of experience in upper airway assessment, evaluated all images in the Digital Imaging and Communications in Medicine (DICOM) format using InVivo software. The head orientation process of the 3D image reconstruction involved: (1) adjusting the Frankfort horizontal plane (FHP) to align with the horizontal plane in both lateral and frontal perspectives; (2) positioning the transporionic plane perpendicular to the FHP and aligning it with the coronal plane in both right and left lateral views; and (3) aligning the midsagittal plane with the sagittal plane [26]. After orientation, the boundaries of the region of interest (ROI) were determined on the midsagittal plane slice using the software's airway module. The posterior border of the vomer served as the anterior and upper boundary, while the most inferior point of the first cervical vertebra was used as reference on the airway's posterior wall for the lower boundary. If the vomer's posterior border was not fully visible on the midsagittal plane, an alternative sagittal slice was selected to establish the limits. The software then automatically identified and segmented the airway within these limits. The researcher reviewed the automated airway segmentation on both axial and sagittal slices, making threshold adjustments as needed to ensure precise airway identification (Fig. 1). The minimum (MinCSA) and maximum cross-sectional areas (MaxCSA) within the ROI were identified, and the NAO ratio was calculated as follows: $(MaxCSA - MinCSA) / MaxCSA$, to determine the percentage of upper airway obstruction related to the adenoid region. The MaxCSA was used as a reference of normal airway passage and the ratio corresponds to the amount of obstruction present relative to what the normal passage would be. The ratio was used to categorize scans into four grades of upper airway obstruction due to adenoid hypertrophy, similar to the previously established scale used by otolaryngologists in nasopharyngoscopy [27] and already used in CBCT studies [28–30]. Scans with a NAO ratio of less than 25% were labeled as Grade 1; 25% to less than 50% as Grade 2; 50% to less than 75% as Grade 3; and 75% or greater as Grade 4. Examples of the grade categories are provided in Fig. 2.

2.4. Sample selection

To ensure a robust and balanced dataset for developing and validating our AI model, we employed a targeted selection approach, resulting in a final sample of 400 CBCT scans with appropriate representation across severity grades. A total of 1215 de-identified scans were initially screened. As Grades 3 and 4 can be considered diagnostic screenings for adenoid hypertrophy [28], all cases identified for these grades were included (128 for Grade 3 and 43 for Grade 4). After collecting all available samples, under-sampling techniques were employed to avoid extreme class imbalances by not adding additional instances to the majority class [31]. The sample was assessed in a randomly

distributed order. The sample selection process is visually represented in Supplementary Figure 1.

2.5. Dataset preparation and model training

The labeled CBCT scans were converted to the Neuroimaging Informatics Technology Initiative (NIFTI) single file compressed format (nii.gz extensions) using the 3D Slicer Batch Anonymizer module [32]. This step ensured compatibility with preprocessing tools and anonymized the data. Subsequently, automatic segmentations of the upper airways were obtained using the automatic multi-anatomical skull structure segmentation (AMASSS) algorithm on the Slicer Automated Dental Tools extension (Fig. 3) [33]. Each segmentation was reviewed for the need for manual corrections and was manually cropped perpendicularly to the airway passage using the ITK-SNAP scalpel mode by the same experienced operator (CTM). Only the region of interest was retained, with the anterior/upper limit at the posterior border of the vomer, and the lower limit defined by the point of the posterior pharyngeal wall at the same level as the lowest point of the first cervical vertebrae (Fig. 4). After cropping, segmentations were converted to surface meshes (format .vtk) using discrete marching cubes to convert the segmentation representation into a surface representation. Laplacian smoothing was applied to clean up the mesh and distribute vertices more evenly, regularizing triangles areas.

The surface meshes constituted the dataset used to train a deep learning model in a shape analysis severity classification task. The experiments were run on a single GPU with Python (version 3.9), PyTorch (version 2.0.1), and CUDA (version 11.7; Nvidia, Santa Clara, Calif) environment. The dataset was split into training (80%) and test (20%) sets. We then applied 5-fold cross-validation dividing the training data into internal training (64%) and internal validation (16%) to create five different train/validate splits aiming to robustly tune hyperparameters (such as learning rate, number of layers, number of trees, and number of leaf nodes) while avoiding overfitting. The internal validation set (16%) was used for early stopping during model training, validating hyperparameter choices from cross-validation and providing an unbiased assessment before testing. The test set (20%) remained completely

untouched until final model evaluation to determine the classification performance. To compensate for class imbalance, a weighted cost-sensitive approach was used during training [31]. The model training parameters are described in Table 1.

The model takes as input a 3D mesh defined by vertices and faces. The multiview branch renders the 3D mesh and captures 2D images from different viewpoints located on the unit sphere, building an image sequence. The viewpoints are chosen following an icosahedron subdivision. Each image has a resolution of (224,224) with 4 channels (RGB + depth) scaled between 0 and 1. The point-cloud branch takes the mesh as input and samples 4096 points uniformly across the surface.

The model architecture combined two approaches. The first approach involved a multiview method [34] that performed feature extraction using EfficientNet-BO, pre-trained on ImageNet, followed by multihead attention with 256 heads to combine information from each view, and feed-forward outputting a feature embedding of dimension 256. The second approach was based on a point-cloud representation [35], where the input was constructed by uniformly sampling 4096 points from the 3D mesh. The resulting point cloud was fed into an encoder-based attention mechanism that computed weighted direction vectors for each point using its 128 nearest neighbors and updated the point positions accordingly. This K-NN-based multihead attention mechanism was applied across three stages, progressively reducing the number of points. The processed point cloud was passed through feed-forward and attention layers, producing a feature embedding of dimension 256. The feature embeddings from both approaches were concatenated to form a unified representation. A fully connected layer was then applied to this concatenated embedding to output predicted class probabilities (Fig. 5). Softmax activation was used to compute the final predictions. To prevent overfitting, regularization techniques such as dropout and weight decay were incorporated during training. This architecture allowed the model to integrate complementary information from both modalities, providing robust predictions while maintaining generalizability.

Surface Gradient-weighted Class Activation Mapping (SurfGradCAM) was used to generate heatmaps[17] that highlighted the regions of the shape that most influenced the classification decision-

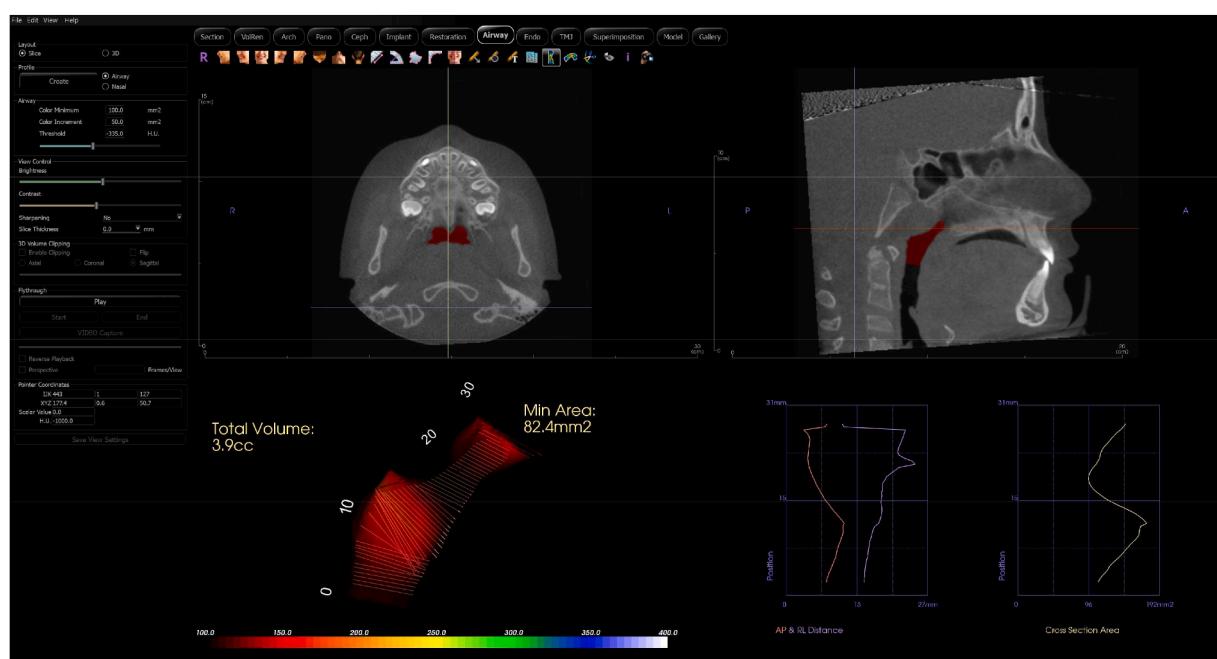


Fig. 1. Image of upper airway segmentation using InVivo software within the specified region of interest. The upper/anterior boundary is defined by the posterior border of the vomer, and the lower boundary by the most inferior point of the first cervical vertebrae. The software automatically segments the airway within these boundaries and calculates the minimum cross-sectional area. The maximum cross-sectional area can be identified in the graph on the lower right corner of the image.

making process, providing an explainability approach and a visual tool for clinicians.

Additionally, a quantification task was performed using a regression model in the same network to predict the NAO ratio, indicating the percentage of obstruction (Fig. 5).

2.6. Statistical analysis

InVivo NAO ratio measurements were repeated in 60 randomly selected scans by the same experienced researcher (CTM), and a second researcher (SLMJ) independently performed the entire measurement sequence. Intra- and inter-examiner reliability was tested using the intraclass correlation coefficient (ICC).

The algorithm's performance in classifying the four grades was

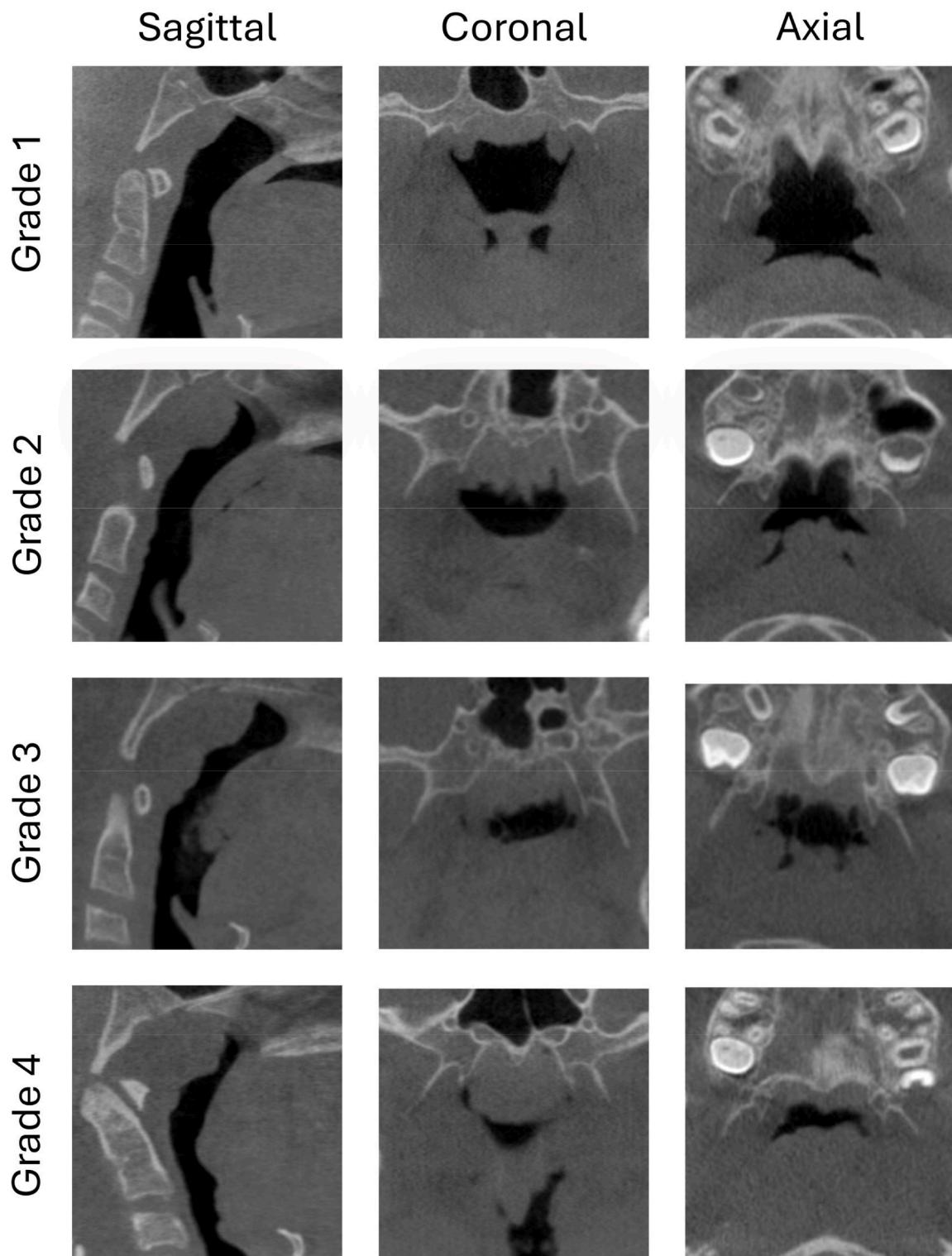


Fig. 2. Image of CBCT scans of illustrative cases of all adenoid-related upper airway obstruction in zoomed-in sagittal, coronal and axial views.



Fig. 3. Image of upper airway automatic segmentation using the AMASS tool on the Slicer Automated Dental Tools extension.



Fig. 4. Image of the upper airway segmentation after cropping the region of interest using the scalpel tool in ITK-SNAP.

Table I
Parameters used in model training.

Learning rate	1e-4
Batch size	32
Optimizer	AdamW with weight decay of 0.01
Loss function	Cross Entropy for the classification task Mean Square Error for the regression task
Monitoring and number of epochs	Use of Early Stopping with a patience of 30 and a stopping criterion monitoring the validation loss. Thus, the number of epochs can vary between settings. For the reported results, the number of epochs is 179 for the classification task and 297 for the regression task.
Dropout	0.1

evaluated using multiple metrics: precision, recall, F₁ score, area under the curve (AUC), and accuracy. We computed and reported both macro-averaged (unweighted) and weighted metrics to assess performance across classes. A confusion matrix was generated to visualize the distribution of predictions. Additionally, weighted Cohen's kappa with quadratic weighting was calculated to measure agreement between the algorithm predictions and ground truth labels [36].

The regression model was evaluated using mean absolute error, root

mean squared error, mean error, and regression and correlation coefficients between the predicted and the ground truth ratio.

3. Results

Four hundred patient datasets were collected for the study (average age, 10.8 years; age range, 5–18 years). The ICC for the NAO ratio repeated measures was 0.987 for the intra-examiner reliability and 0.901 for the inter-examiner assessment, confirming the consistency of the measurements used to define the ground truth.

Table 2 presents the performance of the deep learning model in classifying the severity of nasopharyngeal airway obstruction related to adenoid hypertrophy. The overall accuracy was 0.69. A good overall precision (0.79), corresponding to the positive predicted value (true positives divided by all positives in the model) was obtained. The confusion matrix (**Fig. 6A**) demonstrates the recall performance of the classifier algorithm for each severity grade. The cells on the main diagonal show the percentage correctly classified (accuracy) for each grade compared to the ground truth, with high sensitivity observed for Grades 2 and 3 and agreement above 87.5% within one grade difference of classification for the entire set. The area under the curve (AUC) values for each severity grade (**Fig. 6B**) indicate the model's ability to

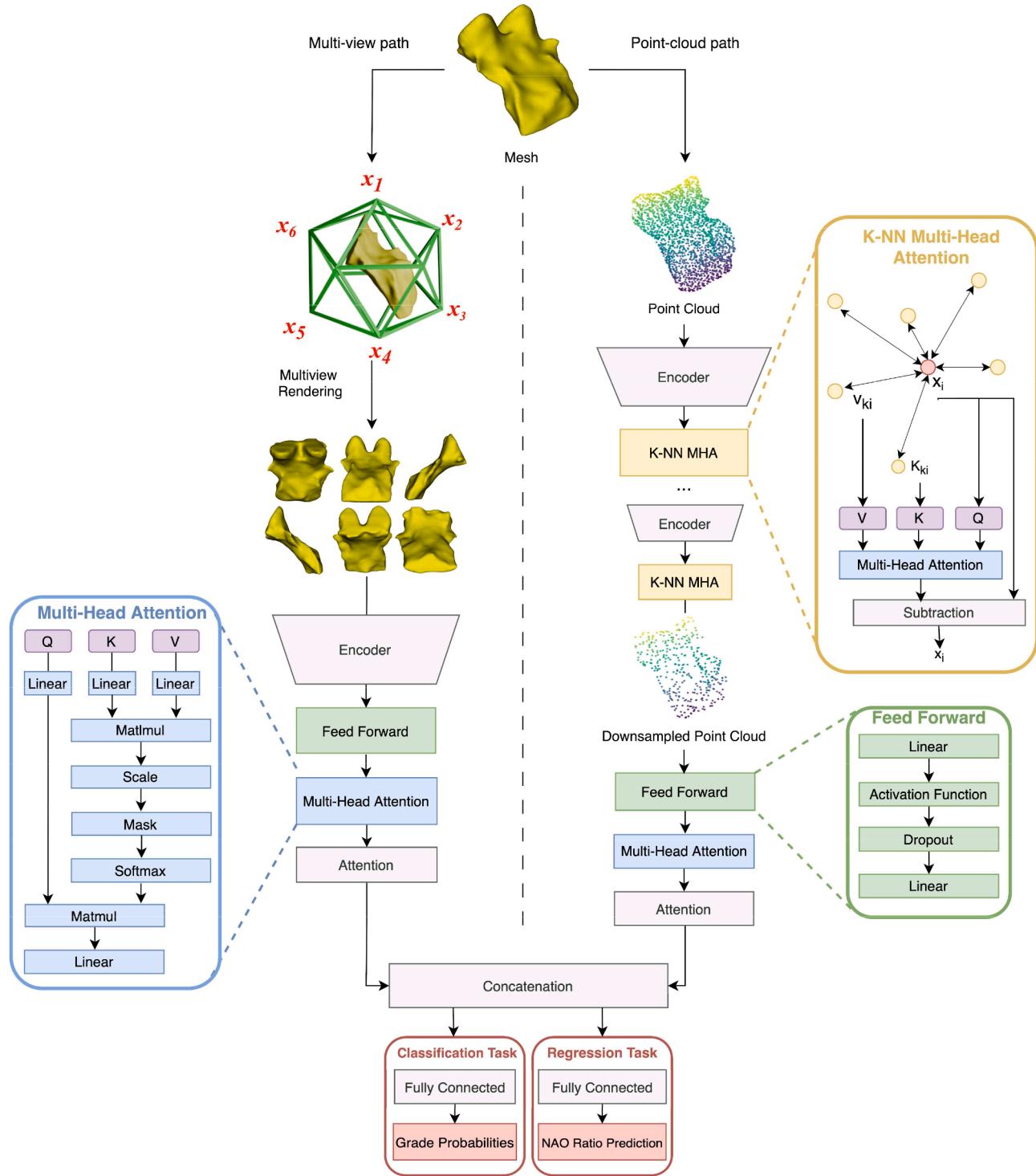


Fig. 5. Architecture combining multiview and point-cloud approaches, concatenated in a fully connected layer to improve the output. The model performs both prediction of obstruction severity grades and NAO ratio estimates.

discriminate between grades. Notably, the AUC values for Grades 3 and 4 were 0.88 and 0.94, respectively, demonstrating excellent performance in identifying more severe cases. The weighted Cohen's kappa between the algorithm's prediction and the ground truth was 0.750, indicating substantial agreement.

Heatmaps generated by the SurfGradCam algorithm for correctly classified cases are displayed in Fig. 7. For each case, heatmaps were created for all grades. Blue regions represent less important features, while red regions indicate the most important features influencing the AI

decision-making during classification. Red regions more precisely identify areas where adenoid tissue would invade the airway space or the most constricted region in the targeted grades matching the correct classification. In other targeted classes, red appears either in incorrect regions or is more widespread without focusing on the specific region mentioned.

Table 3 presents the errors and correlation for the regression task in predicting the NAO ratio compared to the ground truth. Fig. 8 shows the error plot for each grade and a scatterplot with the regression line of

Table II

Performance of the trained AI model for classification of severity of nasopharyngeal airway obstruction.

	Grade 1	Grade 2	Grade 3	Grade 4	macro average	weighted average
AUC	0.85	0.77	0.88	0.94	0.86	0.84
Precision	0.90	0.60	0.65	1.00	0.79	0.75
Recall	0.50	0.82	0.81	0.50	0.66	0.69
F1-score	0.64	0.69	0.72	0.67	0.68	0.68

AUC - area under the curve.

predicted versus ground truth ratios. Both MSE and RMSE indicate the magnitude of prediction errors in the regression, suggesting good performance and reasonable accuracy. The mean err suggests a slight tendency of the model to overpredict the ratio. The violin plot indicates that the model tends to overestimate the ratio in Grade 3 and to underestimate it in Grade 4.

The classification and regression algorithms are available on GitHub for use with the 3D Slicer open-source software (<https://github.com/DCBIA-OrthoLab/ShapeAXI/>). These algorithms will be deployed as a new tool called DOC-ShapeAXI (Dental Oral and Craniofacial Shape Analysis eXplainability and Interpretability) within the software.

4. Discussion

This study is the first to demonstrate an explainable AI model for automatically classifying nasopharyngeal airway obstruction related to adenoid hypertrophy into four severity grades using 3D shape analysis of CBCT scans, while also pioneering the automatic quantification of obstruction percentage.

The combination of multiview and point-cloud approaches was selected after testing three different data representations (point cloud, multiview, and octree) and their combinations. Among these, the chosen approach demonstrated superior performance, effectively leveraging the

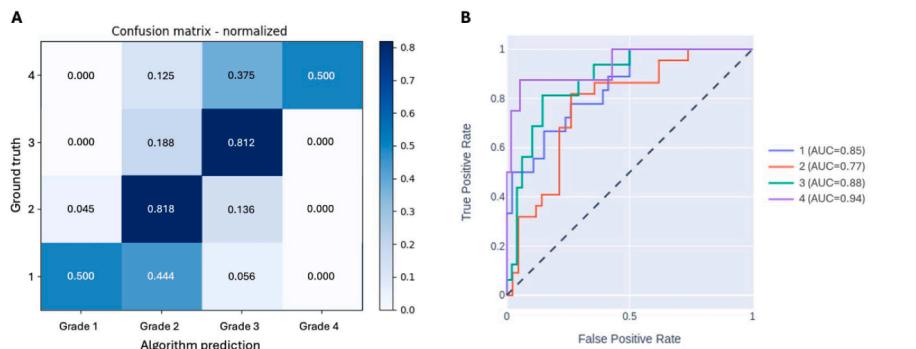


Fig. 6. **A**, Confusion matrix demonstrating the recall performance of the classifier algorithm. Rows represent the ground truth as assessed by the expert using the NAO ratio. Cells on the main diagonal show the percentage correctly classified compared to the ground truth. **B**, Receiver operating characteristic (ROC) curves with the area under the curve (AUC) showing the classification model performance (sensitivity vs. 1-specificity) for each severity grade.

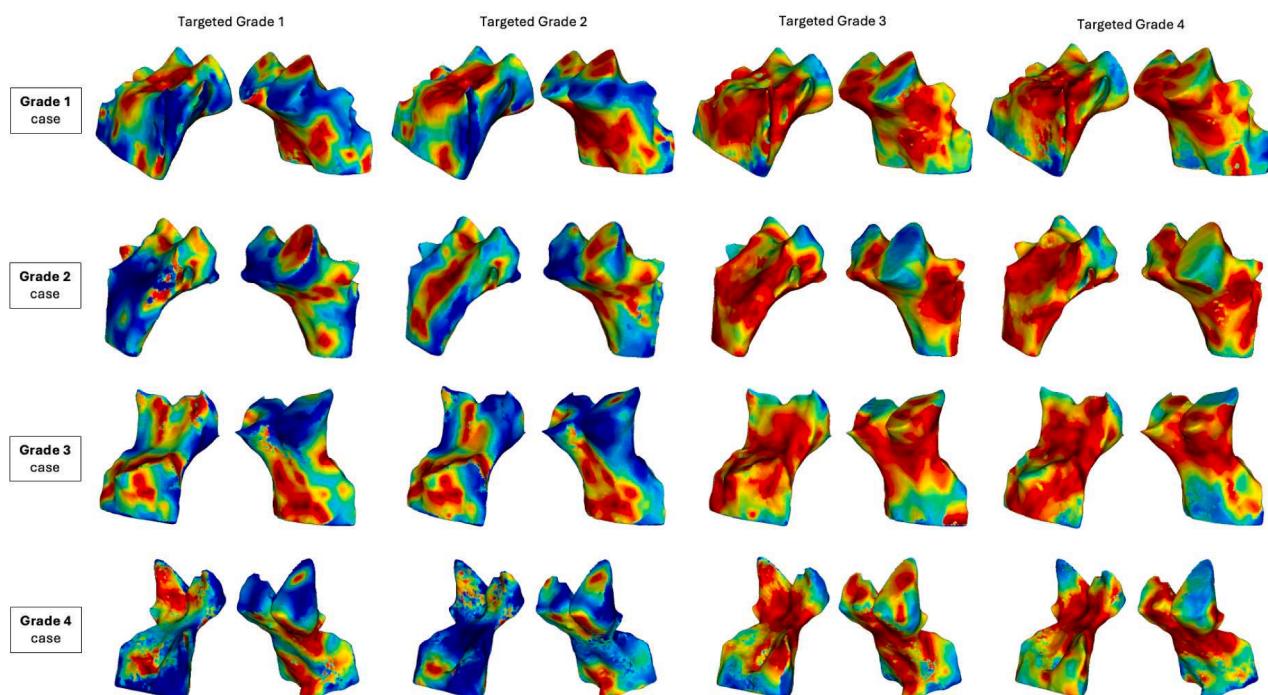


Fig. 7. SurfGradCAM explainability heatmaps for cases correctly predicted by the model as Grade 1 in the first row, Grade 2 in the second row, Grade 3 in the third row and Grade 4 in the fourth row. Columns show heatmaps generated for specific grades for each case, with red indicating the most important features and blue the least important features considered by the algorithm to determine the classification. Matching row x column grades show the most precise areas for severity classification.

Table III

Performance of the trained AI model in the regression task for predicting the nasopharynx airway obstruction (NAO) ratio.

Metric	Value
Mean absolute error	9.12
Root mean squared error	11.26
Mean error	0.13
Regression coefficient (R2)	0.728
Correlation coefficient	
R	0.854
CI	0.769 - 0.909
p-value	< 0.001

R2 - regression coefficient; R - correlation coefficient; CI - confidence interval.

complementary strengths of both modalities.

Our classification task results showed strong overall performance. The ROC curves provided a comprehensive description by combining sensitivity and specificity [37] and captured the intrinsic ability of a technology to detect or classify disease [38]. The AUC values ranged from 0.77 to 0.94, comparable to previous research [18] that reported AUC above 0.92 for automatic detection of adenoid hypertrophy in binary classification. Notably, the highest AUC values corresponded to Grades 3 and 4, highlighting the model's potential for identifying high-risk cases that warrant timely intervention. Furthermore, the model's misclassifications were predominantly within one grade-difference of the true grade, suggesting that even errors tends to be close to the true severity.

The heatmaps generated by the SurfGradCAM algorithm provide visual representation of the upper airway regions that most influenced the AI model's classification decisions. In clinical practice, assessment of adenoid hypertrophy-related upper airway obstruction typically focuses on specific anatomical regions where adenoid tissue may encroach upon the airway space, particularly the nasopharyngeal region, choanal opening, and posterior nasal passage. Our AI-generated heatmaps consistently highlight these clinically relevant regions, with red areas indicating the most important features influencing classification. For example, in correctly classified Grade 3 and Grade 4 cases, the heatmaps emphasize the nasopharyngeal region, aligning with expected sites of significant obstruction in these severe grades.

This strong correlation between the model's focus and clinically relevant regions enhances transparency in AI decision-making process and promotes clinician trust in the AI-assisted diagnostic tools. The SurfGradCAM algorithm, built on GradCAM++ methodology [39], offers an innovative way to visually represent the neural network's reasoning for 3D shape classification [17]. A more refined model may focus on smaller, more specific regions in the heatmaps [40]. In an era where AI adoption in healthcare is rapidly increasing, transparency and

explainability are essential for ensuring safety and reliability. The heatmaps not only validate the AI model's focus on appropriate regions but also provide an intuitive tool that supports clinical decision-making. This interpretability addresses the growing demand for trustworthy AI systems [19,41] and is essential for clinical acceptance and adoption.

The quantification task outcomes were extremely relevant, as accurately predicting obstruction percentage enables more personalized assessment and targeted treatment planning compared to broad severity categories alone. Our pioneering deep learning model showed good performance in predicting this ratio, with an acceptable mean error.

Utilizing accessible open-source tools [33] for automatic multi-scan segmentation significantly reduced processing time [42–44], enhancing the feasibility of incorporating this analysis into routine clinical practice. However, the subsequent steps for manual definition of the region of interest and identifying cross-sectional areas may remain time-consuming.

A strength of our study is that it is the first to report the training of an AI model for the quantification of adenoid hypertrophy-related obstruction in the pharyngeal airways. Our sample size eliminated the need for data augmentation techniques that were necessary in prior research [18] with smaller datasets. While data augmentation can increase training dataset size through various transformations to existing data [45,46], it may introduce challenges in maintaining data integrity and properties, potentially leading to biases [47,48]. Limitations of our study include sample imbalances, particularly for Grade 4 (severe nasopharyngeal airway obstruction), which is less common compared to other severity levels. However, specific strategies were implemented to avoid bias and increase generalizability. The exclusion of scans with artifacts produced by orthodontic appliances or metallic dental crowns may reduce generalizability of our findings. Additionally, it is important to acknowledge that imaging analysis represents just one aspect of upper airway assessment, and CBCT scans, being static, do not provide information on the dynamic respiratory cycle [3].

Future research should focus on validating these results in larger, more diverse cohorts, with either full-head or partial scans to ensure broader applicability. Prospective studies investigating the tool's influence on treatment decisions and patient outcomes would provide valuable insights. Additionally, the model's utility for longitudinal monitoring of adenoid hypertrophy progression or regression warrants further investigation, especially in cases where radiation exposure is justifiable by concomitant features or deformities, offering opportunities for more personalized management. Future research should also explore alternative imaging modalities to CBCT for assessing the dynamic respiratory cycle.

The integration of AI technologies into orthodontic practices presents both opportunities and challenges. While these tools offer potential for improving patient care and optimizing clinical workflows, they are

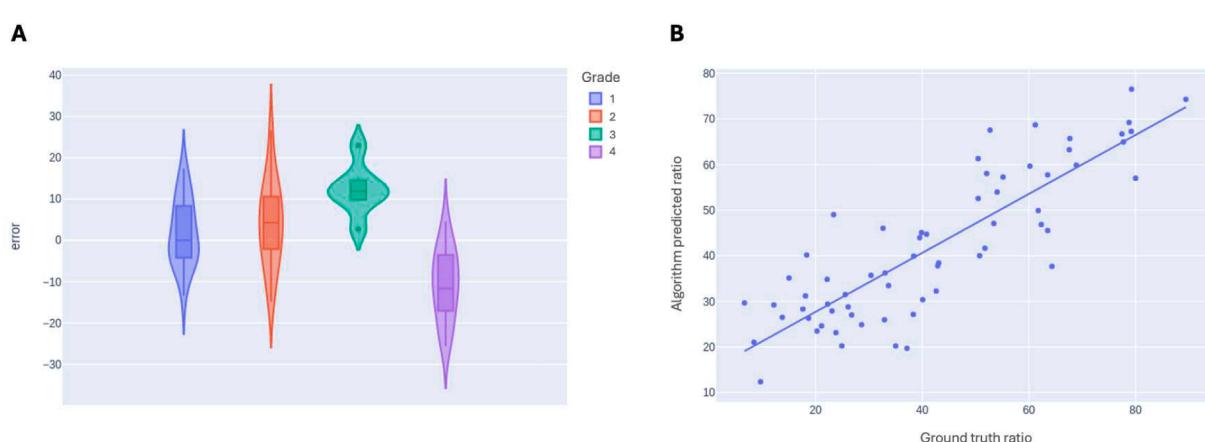


Fig. 8. A, Violin plot representing errors for each grade. B, Scatterplot with regression line between the ground truth ratio and the algorithm-predicted ratio.

intended to support, rather than replace, expert clinical judgment. The successful adoption of AI systems requires close collaboration between developers, researchers, and clinicians to ensure reliability, transparency, and alignment with clinical needs and ethical principles. The promising results of this study underscore the potential of explainable AI and 3D shape analysis to support clinical decision-making and improve patient care. By carefully balancing the application of AI with clinician oversight, these tools may be developed and deployed in a manner that prioritizes patient well-being and upholds the highest standards of clinical excellence in orthodontic care.

5. Conclusions

The proposed explainable AI model, utilizing 3D shape analysis, demonstrated strong performance in both classifying and quantifying the severity of adenoid hypertrophy-related upper airway obstruction in CBCT scans. The model's high AUC values, particularly for more severe obstruction grades, highlight its potential as a valuable tool for screening and diagnostic purposes. The regression model's ability to accurately predict the nasopharyngeal airway obstruction ratio further underscores the AI's capability to provide quantitative insights into the extent of obstruction. The explainability heatmaps consistently highlighted the most relevant anatomical regions influencing the AI's decision-making process, offering a promising approach to enhance the interpretability and trustworthiness of AI-assisted diagnostics. These findings suggest that the integration of explainable AI and 3D shape analysis could significantly improve the efficiency and accuracy of assessing adenoid hypertrophy-related upper airway obstruction.

Statement

During the preparation of this work the authors used ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Funding

This work was supported by the National Institute of Dental and Craniofacial Research (NIDCR). The grant number is R01DE024550.

CRediT authorship contribution statement

Claudia Trindade Mattos: Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lucie Dole:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Sergio Luiz Mota-Júnior:** Writing – review & editing, Visualization, Validation, Methodology, Investigation. **Adriana de Alcantara Cury-Saramago:** Writing – review & editing, Validation, Resources. **Jonas Bianchi:** Writing – review & editing, Validation, Resources, Methodology, Investigation. **Heesoo Oh:** Writing – review & editing, Validation, Resources. **Karine Evangelista:** Writing – review & editing, Validation, Resources, Investigation. **José Valladares-Neto:** Writing – review & editing, Validation, Resources. **Antonio Carlos de Oliveira Ruellas:** Writing – review & editing, Validation, Methodology. **Juan Carlos Prieto:** Writing – review & editing, Validation, Software, Resources, Methodology, Formal analysis, Conceptualization. **Lucia Helena Soares Cevidanes:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

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.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jdent.2025.105689.

References

- [1] M.M. Hasuneh, A.A. Toubasi, B. Khraisat, H. Aldabbas, M. Al-Ide, Risk factors of obstructive sleep apnea (OSA) in pediatric patients: a systematic review and meta-analysis, *J. Pediatr. Health Care* 38 (5) (2024) 717–726, <https://doi.org/10.1016/j.jpedhc.2024.05.008>.
- [2] S. Selvadurai, G. Voutsas, E.J. Propst, N.E. Wolter, I. Narang, Obstructive sleep apnea in children aged 3 years and younger: rate and risk factors, *Paediatr. & Child Health* 25 (7) (2020) 432–438, <https://doi.org/10.1093/pch/pxz097>.
- [3] R.G. Behrents, A.V. Shelgikar, R.S. Conley, C. Flores-Mir, M. Hans, M. Levine, J. A. McNamara, J.M. Palomo, B. Pliska, J.W. Stockstill, J. Wise, S. Murphy, N. J. Nagel, J. Hitner, Obstructive sleep apnea and orthodontics: an American association of orthodontists white paper, *Am. J. Orthod. Dentofacial Orthop.* 156 (1) (2019) 13–28.e1, <https://doi.org/10.1016/j.ajodo.2019.04.009>.
- [4] N.C.F. Fagundes, C. Flores-Mir, Pediatric obstructive sleep apnea—Dental professionals can play a crucial role, *Pediatr. Pulmonol.* 57 (8) (2021) 1860–1868, <https://doi.org/10.1002/ppul.25291>.
- [5] J.G. Costa, G.S. Costa, C. Costa, O.V. Vilella, C.T. Mattos, A.A. Cury-Saramago, Clinical recognition of mouth breathers by orthodontists: a preliminary study, *Am. J. Orthod. Dentofacial Orthop.* 152 (5) (2017) 646–653, <https://doi.org/10.1016/j.jajodo.2017.03.025>.
- [6] M.C. Pacheco, C.F. Casagrande, L.P. Teixeira, N.S. Finck, M.T. de Araujo, Guidelines proposal for clinical recognition of mouth breathing children, *Dental Press J. Orthod.* 20 (4) (2015) 39–44, <https://doi.org/10.1590/2176-9451.20.4.039-044.oar>.
- [7] B. Basheer, K.S. Hegde, S.S. Bhat, D. Umar, K. Baroudi, Influence of mouth breathing on the dentofacial growth of children: a cephalometric study, *J. Int. Oral Health* 6 (6) (2014) 50–55.
- [8] M. Bergeron, A.L. Duggins, A.P. Cohen, B.A. Leader, S.L. Ishman, The impact of persistent pediatric obstructive sleep apnea on the Quality of Life of Patients' families, *Int. J. Pediatr. Otorhinolaryngol.* 129 (2020), <https://doi.org/10.1016/j.ijporl.2019.109723>.
- [9] R.B. Leal, M.C. Gomes, A.F. Granville-Garcia, P.S.A. Goes, V.A. de Menezes, Development of a questionnaire for measuring health-related quality of life among children and adolescents with mouth breathing, *Am. J. Rhinol. Allergy* 29 (6) (2015) e212–e215, <https://doi.org/10.2500/ajra.2015.29.4258>.
- [10] W.C. Hsu, K.T. Kang, C.C.J. Yao, C.H. Chou, W.C. Weng, P.L. Lee, Y.J. Chen, Evaluation of upper airway in children with obstructive sleep apnea using cone-beam computed tomography, *Laryngoscope* 131 (3) (2020) 680–685, <https://doi.org/10.1002/lary.28863>.
- [11] M.D. Abé-Nickler, S. Pörtner, P. Sieg, S.G. Hakim, No correlation between two-dimensional measurements and three-dimensional configuration of the pharyngeal upper airway space in cone-beam computed tomography, *J. Craniomaxillofac. Surg.* 45 (3) (2017) 371–376, <https://doi.org/10.1016/j.jcms.2017.01.004>.
- [12] J. Neumayr, E. Frenkel, J. Schwarzmaier, N. Ammar, A. Kessler, F. Schwendicke, J. Kühnisch, H. Dujic, External validation of an artificial intelligence-based method for the detection and classification of molar incisor hypomineralisation in dental photographs, *J. Dent.* 148 (2024), <https://doi.org/10.1016/j.jdent.2024.105228>.
- [13] R.S. Fedato Tobias, A.B. Teodoro, K. Evangelista, A.F. Leite, J. Valladares-Neto, B. S. de Freitas Silva, F.P. Yamamoto-Silva, F.T. Almeida, M.A.G. Silva, Diagnostic capability of artificial intelligence tools for detecting and classifying odontogenic cysts and tumors: a systematic review and meta-analysis, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 138 (3) (2024) 414–426, <https://doi.org/10.1016/j.oooo.2024.03.004>.
- [14] E. Bardideh, F. Lal Alizadeh, M. Amiri, M. Ghorbani, Designing an artificial intelligence system for dental occlusion classification using intraoral photographs: a comparative analysis between artificial intelligence-based and clinical diagnoses, *Am. J. Orthod. Dentofacial Orthop.* 166 (2) (2024) 125–137, <https://doi.org/10.1016/j.ajodo.2024.03.012>.
- [15] A. Yari, P. Fasih, M. Hosseini Hooshiar, A. Goodarzi, S.F. Fattahi, Detection and classification of mandibular fractures in panoramic radiography using artificial intelligence, *Dentomaxillofac. Radiol.* 53 (6) (2024) 363–371, <https://doi.org/10.1093/dmfr/twe018>.
- [16] B. Tiryaki, K. Torenek-Agirman, O. Miloglu, B. Korkmaz, İ.Y. Ozbek, E.A. Oral, Artificial intelligence in tongue diagnosis: classification of tongue lesions and normal tongue images using deep convolutional neural network, *BMC Med. Imaging* 24 (1) (2024), <https://doi.org/10.1186/s12880-024-01234-3>.
- [17] F. Miranda, V. Choudhari, S. Barone, L. Anchling, N. Hutin, M. Gurgel, N. Al Turkestani, M. Yatabe, J. Bianchi, A. Aliaga-Del Castillo, P. Zupelari-Gonçalves, S. Edwards, D. Garib, L. Cevidanes, J. Prieto, Interpretable artificial intelligence for classification of alveolar bone defect in patients with cleft lip and palate, *Sci. Rep.* 13 (1) (2023), <https://doi.org/10.1038/s41598-023-43125-7>.
- [18] W. Dong, Y. Chen, A. Li, X. Mei, Y. Yang, Automatic detection of adenoid hypertrophy on cone-beam computed tomography based on deep learning, *Am. J. Orthod. Dentofacial Orthop.* 163 (4) (2023) 553–560.e3, <https://doi.org/10.1016/j.jajodo.2022.11.011>.

- [19] M. Reyes, R. Meier, S. Pereira, C.A. Silva, F.-M. Dahlweid, H.v. Tengg-Kobligk, R. M. Summers, R. Wiest, On the Interpretability of Artificial Intelligence in Radiology: challenges and Opportunities, *Radiol. Artif. Intell.* 2 (3) (2020), <https://doi.org/10.1148/ryai.2020190043>.
- [20] Y. Shen, X. Li, X. Liang, H. Xu, C. Li, Y. Yu, B. Qiu, A deep-learning-based approach for adenoid hypertrophy diagnosis, *Med. Phys.* 47 (5) (2020) 2171–2181, <https://doi.org/10.1002/mp.14063>.
- [21] T. Zhao, J. Zhou, J. Yan, L. Cao, Y. Cao, F. Hua, H. He, Automated adenoid hypertrophy assessment with lateral cephalometry in children based on artificial intelligence, *Diagnostics* 11 (8) (2021), <https://doi.org/10.3390/diagnostics11081386>.
- [22] J.L. Liu, S.H. Li, Y.M. Cai, D.P. Lan, Y.F. Lu, W. Liao, S.C. Ying, Z.H. Zhao, Automated radiographic evaluation of adenoid hypertrophy based on VGG-Lite, *J. Dent. Res.* 100 (12) (2021) 1337–1343, <https://doi.org/10.1177/00220345211009474>.
- [23] Y. Jeong, Y. Nang, Z. Zhao, K. Ahmed, Automated evaluation of upper airway obstruction based on deep learning, *BioMed Res. Int.* 2023 (1) (2023), <https://doi.org/10.1155/2023/8231425>.
- [24] 3D Slicer, 2024. <https://download.slicer.org>. Accessed 19 Jul 2024.
- [25] ITK-SNAP. www.itksnap.org. (Accessed 19 Jul 2024).
- [26] E.V. Biggs, E. Benavides, J.A. McNamara Jr., L.H.S. Cevidanes, F. Copello, R. R. Lints, J.P. Lints, A.C.O. Ruellas, Three-dimensional Evaluation of the Carriere Motion 3D Appliance in the treatment of Class II malocclusion, *Am. J. Orthod. Dentofacial Orthop.* 164 (6) (2023) 824–836, <https://doi.org/10.1016/j.ajodo.2023.05.031>.
- [27] G. Bravo, A. Ysunza, J. Arrieta, M.C. Pamplona, Videonasopharyngoscopy is useful for identifying children with Pierre Robin sequence and severe obstructive sleep apnea, *Int. J. Pediatr. Otorhinolaryngol.* 69 (1) (2005) 27–33, <https://doi.org/10.1016/j.ijporl.2004.07.009>.
- [28] M.P. Major, M. Witmans, H. El-Hakim, P.W. Major, C. Flores-Mir, Agreement between cone-beam computed tomography and nasoendoscopy evaluations of adenoid hypertrophy, *Am. J. Orthod. Dentofacial Orthop.* 146 (4) (2014) 451–459, <https://doi.org/10.1016/j.ajodo.2014.06.013>.
- [29] C. Pacheco-Pereira, N.A. Alsufyani, M.P. Major, C. Flores-Mir, Accuracy and reliability of oral maxillofacial radiologists when evaluating cone-beam computed tomography imaging for adenoid hypertrophy screening: a comparison with nasopharyngoscopy, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* 121 (6) (2016) e168–e174, <https://doi.org/10.1016/j.oooo.2016.03.010>.
- [30] C. Pacheco-Pereira, N.A. Alsufyani, M. Major, G. Heo, C. Flores-Mir, Accuracy and reliability of orthodontists using cone-beam computerized tomography for assessment of adenoid hypertrophy, *Am. J. Orthod. Dentofacial Orthoped.* 150 (5) (2016) 782–788, <https://doi.org/10.1016/j.ajodo.2016.03.030>.
- [31] J.L. Leevy, T.M. Khoshgoftaar, R.A. Bauder, N. Seliya, A survey on addressing high-class imbalance in big data, *J. Big Data* 5 (1) (2018), <https://doi.org/10.1186/s40537-018-0151-6>.
- [32] SlicerBatchAnonymize. <https://github.com/DCBIA-OrthoLab/SlicerBatchAnonymize>. (Accessed Jul 19 2024).
- [33] M. Gillot, B. Baquero, C. Le, R. Deleat-Besson, J. Bianchi, A. Ruellas, M. Gurgel, M. Yatabe, N. Al Turkestani, K. Najarian, R. Soroushmehr, S. Pieper, R. Kikinis, B. Paniagua, J. Gryak, M. Ishida, C. Massaro, L. Gomes, H. Oh, K. Evangelista, C. M. Chaves Junior, D. Garib, F. Costa, E. Benavides, F. Soki, J.-C. Fillion-Robin, H. Joshi, L. Cevidanes, J.C. Prieto, Automatic multi-anatomical skull structure segmentation of cone-beam computed tomography scans using 3D UNETR, *PLoS One* 17 (10) (2022), <https://doi.org/10.1371/journal.pone.0275033>.
- [34] J.C. Prieto, F. Miranda, M. Gurgel, L. Anchling, N. Hutin, S. Barone, N. Al Turkestani, A. Aliaga, M. Yatabe, J. Bianchi, L. Cevidanes, S. Wu, H. Yoshida, ShapeAXI: shape analysis explainability and interpretability, *Proc. SPIE Int. Soc. Opt. Eng.* (2024), <https://doi.org/10.1117/12.3007053>.
- [35] L. Zou, Z. Huang, N. Gu, G. Wang, MSSPA-GC: multi-Scale Shape Prior Adaptation with 3D Graph Convolutions for Category-Level Object Pose Estimation, *Neural Netw.* 166 (2023) 609–621, <https://doi.org/10.1016/j.neunet.2023.07.037>.
- [36] O. Rainio, J. Teuhola, R. Klén, Evaluation metrics and statistical tests for machine learning, *Sci. Rep.* 14 (1) (2024), <https://doi.org/10.1038/s41598-024-56706-x>.
- [37] C.E. Metz, Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems, *J. Am. Coll. Radiol.* 3 (6) (2006) 413–422, <https://doi.org/10.1016/j.jacr.2006.02.021>.
- [38] R.F. Wagner, C.E. Metz, G. Campbell, Assessment of medical imaging systems and computer aids: a tutorial review, *Acad. Radiol.* 14 (6) (2007) 723–748, <https://doi.org/10.1016/j.acra.2007.03.001>.
- [39] A. Chattopadhyay, A. Sarkar, P. Howlader, V.N. Balasubramanian, Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 839–847, <https://doi.org/10.1109/WACV.2018.00097>.
- [40] Q. Chen, J. Huang, H.S. Salehi, H. Zhu, L. Lian, X. Lai, K. Wei, Hierarchical CNN-based occlusal surface morphology analysis for classifying posterior tooth type using augmented images from 3D dental surface models, *Comput. Methods Programs Biomed.* 208 (2021), <https://doi.org/10.1016/j.cmpb.2021.106295>.
- [41] S. Ali, F. Akhlaq, A.S. Imran, Z. Kastrati, S.M. Daupota, M. Moosa, The enlightening role of explainable artificial intelligence in medical & healthcare domains: a systematic literature review, *Comput. Biol. Med.* 166 (2023), <https://doi.org/10.1016/j.combiomed.2023.107555>.
- [42] T. Ogawa, R. Enciso, W.H. Shintaku, G.T. Clark, Evaluation of cross-section airway configuration of obstructive sleep apnea, *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* 103 (1) (2007) 102–108, <https://doi.org/10.1016/j.tripleo.2006.06.008>.
- [43] P. Maken, A. Gupta, M.K. Gupta, A systematic review of the techniques for automatic segmentation of the human upper airway using volumetric images, *Med. Biol. Eng. Comput.* 61 (8) (2023) 1901–1927, <https://doi.org/10.1007/s11517-023-02842-x>.
- [44] R. Leonardi, A. Lo Giudice, M. Farronato, V. Ronisvalle, S. Allegrini, G. Musumeci, C. Spampinato, Fully automatic segmentation of sinonasal cavity and pharyngeal airway based on convolutional neural networks, *Am. J. Orthod. Dentofacial Orthop.* 159 (6) (2021) 824–835.e1, <https://doi.org/10.1016/j.ajodo.2020.05.017>.
- [45] H. Mohammad-Rahimi, M. Nadimi, M.H. Rohban, E. Shamsoddin, V.Y. Lee, S. R. Motamedian, Machine learning and orthodontics, current trends and the future opportunities: a scoping review, *Am. J. Orthod. Dentofacial Orthop.* 160 (2) (2021) 170–192.e4, <https://doi.org/10.1016/j.ajodo.2021.02.013>.
- [46] A.-J. Rousseau, M. Geubbelsmans, T. Burzykowski, D. Valkenborg, Deep learning, *Am. J. Orthod. Dentofacial Orthop.* 165 (3) (2024) 369–371, <https://doi.org/10.1016/j.ajodo.2023.12.003>.
- [47] T. Islam, M.S. Hafiz, J.R. Jim, M.M. Kabir, M.F. Mridha, A systematic review of deep learning data augmentation in medical imaging: recent advances and future research directions, *Healthcare Analytics* 5 (2024), <https://doi.org/10.1016/j.health.2024.100340>.
- [48] B. Lin, M. Cheng, S. Wang, F. Li, Q. Zhou, Automatic detection of anteriorly displaced temporomandibular joint discs on magnetic resonance images using a deep learning algorithm, *Dentomaxillofac. Radiol.* 51 (3) (2022), <https://doi.org/10.1259/dmfr.20210341>.