

RESEARCH

Open Access



Automatic dental age calculation from panoramic radiographs using deep learning: a two-stage approach with object detection and image classification

Kazuma Kokomoto^{1*}, Rina Kariya², Aya Muranaka², Rena Okawa², Kazuhiko Nakano² and Kazunori Nozaki^{1*}

Abstract

Background Dental age is crucial for treatment planning in pediatric and orthodontic dentistry. Dental age calculation methods can be categorized into morphological, biochemical, and radiological methods. Radiological methods are commonly used because they are non-invasive and reproducible. When radiographs are available, dental age can be calculated by evaluating the developmental stage of permanent teeth and converting it into an estimated age using a table, or by measuring the length between some landmarks such as the tooth, root, or pulp, and substituting them into regression formulas. However, these methods heavily depend on manual time-consuming processes. In this study, we proposed a novel and completely automatic dental age calculation method using panoramic radiographs and deep learning techniques.

Methods Overall, 8,023 panoramic radiographs were used as training data for Scaled-YOLOv4 to detect dental germs and mean average precision were evaluated. In total, 18,485 single-root and 16,313 multi-root dental germ images were used as training data for EfficientNetV2 M to classify the developmental stages of detected dental germs and Top-3 accuracy was evaluated since the adjacent stages of the dental germ looks similar and the many variations of the morphological structure can be observed between developmental stages. Scaled-YOLOv4 and EfficientNetV2 M were trained using cross-validation. We evaluated a single selection, a weighted average, and an expected value to convert the probability of developmental stage classification to dental age. One hundred and fifty-seven panoramic radiographs were used to compare automatic and manual human experts' dental age calculations.

Results Dental germ detection was achieved with a mean average precision of 98.26% and dental germ classifiers for single and multi-root were achieved with a Top-3 accuracy of 98.46% and 98.36%, respectively. The mean absolute errors between the automatic and manual dental age calculations using single selection, weighted average, and expected value were 0.274, 0.261, and 0.396, respectively. The weighted average was better than the other methods and was accurate by less than one developmental stage error.

Conclusion Our study demonstrates the feasibility of automatic dental age calculation using panoramic radiographs and a two-stage deep learning approach with a clinically acceptable level of accuracy.

*Correspondence:

Kazuma Kokomoto
kokomoto.kazuma.dent@osaka-u.ac.jp
Kazunori Nozaki
nozaki.kazunori.dent@osaka-u.ac.jp

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords Artificial intelligence, Machine learning, Medical informatics applications, Pediatric dentistry, Orthodontics, Dental informatics

Introduction

Growth and development assessment in children is essential for making appropriate diagnosis and treatment decisions in orthodontic and pediatric dentistry [1]. Dental age and chronological age are different ways of measuring a person's developmental stage. Chronological age refers to a person's actual age based on their date of birth. This is the most commonly used way of measuring age and is used to determine when a person reaches certain developmental stages. Dental age, on the other hand, is an estimate of a person's age based on the development of their teeth. While chronological age is just a fixed number that does not change, dental age can vary depending on a person's growth and thus provide more important and individual information. In addition, dental age is useful in deciding when to initiate orthodontic treatment or whether a child's dental development is delayed or advanced since chronological age is not always equal to dental age [2, 3].

The various dental age calculation methods can be categorized as morphological methods, biochemical methods, and radiological methods [4, 5]. Morphological methods are based on measurement of actual teeth and regression formulas are used for calculation. Biochemical methods are based on the racemization of amino acids [6]. Radiological methods are commonly used since they are non-invasive and reproducible compared to other methods [7–11]. Estimating dental age is feasible when radiographs are available, by assessing the growth stage of permanent teeth and converting this stage into an estimated age using a lookup table, or by measuring the distance from landmarks such as the tooth, root, or pulp, and inputting these values into regression equations [4, 7].

However, these processes are predominantly manual and require considerable time. The average time for manually calculating dental age is 10 min [12], making it inconvenient to be performed for every patient in daily clinical practice. Therefore, automatic dental age calculation is expected to save time in treatment planning by eliminating time-consuming but crucial routine tasks and increasing the interaction time between dentists and patients [13].

Recently, deep learning with Convolutional Neural Networks (CNN) and Artificial Intelligence (AI) in computer vision has been developed, which can automatically extract imaging features with original pixel information as input data. In previous studies, deep learning methods

for object detection, image classification, and image segmentation were widely used in dentistry [14–19]. Furthermore, the field of chronological age calculation has seen a growing interest in applying these techniques [20–23]. Studies have demonstrated that CNN models can surpass the accuracy of manual methods in classifying chronological age based on dental images. However, few studies included multiple deep learning techniques and addressed germ detection or developmental stage classification. In addition, those methods have been developed for chronological age calculation, not for dental age calculation.

This study aimed to fill the gaps between classical manual calculations and modern AI technologies in the field of dental age calculation. We proposed a novel and completely automatic dental age calculation method using panoramic radiographs and two-stage deep learning combined with object detection and image classification, trained with voluminous images. Additionally, we evaluated its accuracy by comparing automatic and expert manual calculations and whether our proposed method could be clinically acceptable.

Materials and methods

Dataset

This study was retrospective and observational in nature. All images used in this study were obtained from patients who received dental treatment between January 2000 and December 2018 at Osaka University Dental Hospital, Department of Pediatric Dentistry, Osaka, Japan. All images were anonymized and had no metadata such as patient name, chronological age, sex, dentition, or disease due to ethics. Roughly speaking, the datasets contained a relatively high number of images that showed healthy dentition. Our proposed process is illustrated in Fig. 1.

Germ detection

We utilized Scaled-YOLOv4 [24] as our germ detector, which is an improved version of YOLOv4 [25] and has achieved state-of-the-art object detection. Scaled-YOLOv4 has performed well with larger models and input image sizes [24]; therefore, we used the second largest model, Scaled-YOLO v4 P6, for germ detection with an input size of 1280×1280 pixels, as much as our computational resource allowed.

To train Scaled-YOLOv4, 8023 panoramic radiographs were used as training data. These images included primary, mixed, and permanent dentitions. Four pediatric

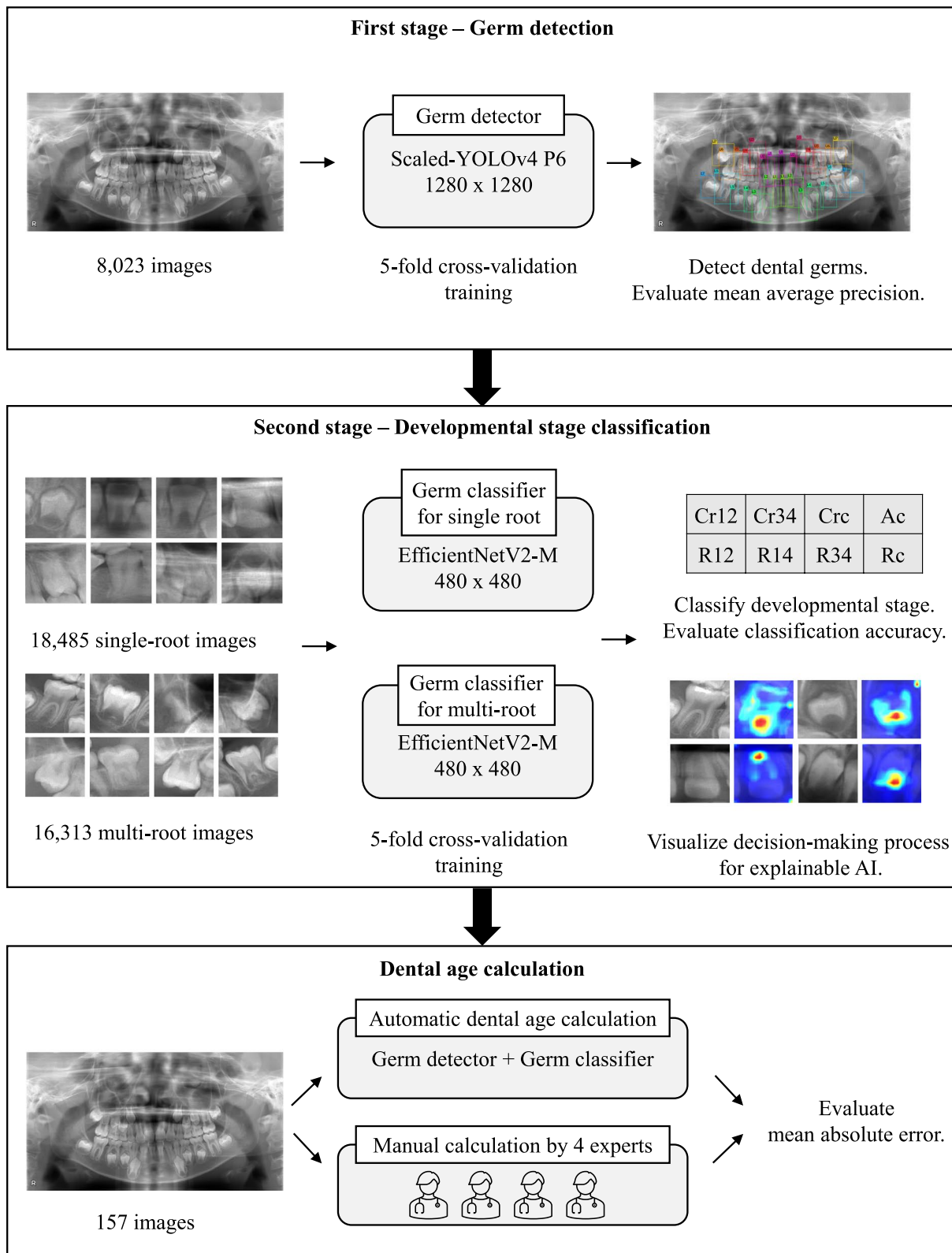


Fig. 1 Our pipeline processes for automatic dental age calculation and evaluation methods

dentists were presented with panoramic radiographs and instructed to draw box boundaries around all dental germs and class-label those boxes. The class labels were based on a palmer system [26]. The correspondence between number and tooth; 1 was the central incisor, 2 was the lateral incisor, 3 was the canine, 4 was the first premolar, 5 was the second premolar, 6 was the first molar, 7 was the second molar, and 8 was the third molar. We added upper case prefixes “U” and “L” to identify the upper and lower tooth, respectively. For instance, U1 denotes the upper central incisor, and L6 denotes the lower first molar.

Our model’s performance was evaluated using average precisions with a 0.5 Intersection over Unit threshold, which is a common metric in object detection deep learning models, including the YOLO families [24, 25, 27–31], known as AP₅₀. We performed the 5-fold cross-validation to assess the generalization of our model and prevent overfitting [32]. The training datasets were split into five, and Scaled-YOLOv4 was trained using four of them. The remaining fold was used to calculate the AP₅₀ of the trained model. We replicated this five times and evaluated the average of the out-of-fold predictions.

Developmental stage classification

In total, 18,485 single- and 16,313 multi-root dental germ images were prepared using our germ detector and used as the training dataset. Four pediatric dentists were presented with dental germ images and instructed to classify developmental stages in Japanese children as Cr1/2 (1/2 crown formation), Cr3/4 (3/4 crown formation), Crc (crown complete), R1/4 (1/4 root formation), R1/2 (1/2 root formation), R3/4 (3/4 root formation), Rc (complete root formation), and Ac (apex closed) [8]. Ci (initial calcification) and Cco (coalescence of cusps) were excluded because the images were too few to train the model. Each of the four pediatric dentists annotated the distinct images once and did not re-annotate any image annotated by the other dentists.

We utilized EfficientNetV2 to classify dental germ images [33]. EfficientNetV2 is an improved version of EfficientNet [34] and a CNN-based image classification model that achieves state-of-the-art performance on the ImageNet dataset with better accuracy and efficiency than previous famous models, such as ResNet [35], DenseNet [36], and Xception [37]. EfficientNetV2 scales up from EfficientNetV2-S to EfficientNetV2-M/L, and classification performance also improves as the model scales up. However, the computational complexity increases exponentially as EfficientNetV2 scales up; therefore, we chose the intermediate EfficientNetV2-M as our germ classification model. All germ images were resized to 480 × 480 pixels to train EfficientNetV2-M. We

performed 5-fold stratified cross-validation so that each fold could have the same proportion of developmental stages. The classification accuracy of out-of-fold predictions was evaluated using the same procedure as germ detection. We evaluated the Top-1 accuracy and Top-3 accuracy. The former is a metric of model prediction performance that must match the single developmental stages. The latter is the classification correctness, in which the top three highest probabilities of model predictions matched the target developmental stages.

It is also crucial to know the decision-making process of AI for interpretability and explainability [38, 39]. We applied Gradient-weighted Class Activation Mapping (Grad-CAM) [40] to analyze how our model could classify dental germs and whether the procedure was similar to that used by dentists.

Dental age calculation

In the germ classification stage, we obtained the probability of each dental germ’s developmental stage. We evaluated a single selection, a weighted average, and an expected value to convert the probability to dental age. The Ac stage was excluded from calculation because this stage refers to the end of dental germ development and does not have a dental age [8]. The single selection chooses one developmental stage with the highest probability and converts it to the respective dental age [8]. The weighted average considers the probability of each developmental stage. We used the three highest probabilities to calculate the weighted average, as follows:

$$\text{weighted average} = \frac{x_1p_1 + x_2p_2 + x_3p_3}{p_1 + p_2 + p_3}$$

where p_1 , p_2 , and p_3 are the top three probabilities of the dental germ’s developmental stage, as obtained using the germ classifier, and x_1 , x_2 , and x_3 are the dental ages of the corresponding developmental stages. The expected value considers all probabilities and was calculated as follows:

$$\text{expected value} = \frac{\sum_{i=1}^n x_i p_i}{\sum_{i=1}^n p_i}$$

where n is the number of developmental stages used for calculation, p_i are the probabilities of the dental germ’s developmental stage, and x_i are the dental ages of the corresponding developmental stages.

After converting the probability to dental age, the simple average of the dental age of each dental germ was regarded as the overall dental age of each panoramic radiograph [41]. To analyze the accuracy of the overall dental age, 157 panoramic radiographs that were not included in previous training datasets were used, and automatic dental age calculation was performed using

one of the 5-fold cross-validated germ detector and germ classifier models. Four pediatric dentists manually calculated the overall dental age using the same radiographs, and the mean absolute errors between the experts' and the automatic calculation were evaluated. Since all the images used in this study were anonymized and had no metadata in terms of sex, we calculated dental age of males and females from the same panoramic radiograph and evaluated an average of both.

Results

The germ detector's performance is presented in Table 1. Scaled-YOLOv4 P6 with an input image size of 1280×1280 achieved detection accuracy with an AP₅₀ of 98.26%. Examples of germ cell detection with and without congenitally missing teeth are presented in Figs. 2, 3, and 4. Notably, the dental germs were correctly detected. Primary tooth conditions, such as healthy, caries, composite filling, metal crown, root canal filling, and orthodontic materials, did not affect detection performance. Examples of germ detection failures are illustrated in Fig. 4. Most dental germs were accurately detected; however, some dental germs were not.

The performance of the germ classifier for single- or multi-root dental germ images with 5-fold cross-validation is summarized in Table 1. The germ classifier for single- and multi-root dental germs achieved the highest Top-1 classification accuracies of 68.31 and 71.54% and the Top-3 accuracies of 98.46 and 98.36%, respectively. The confusion matrices for germ classification are presented in Fig. 5. The germ classifier tended to misclassify the actual stages as adjacent stages. The Grad-CAM images of the germ classifier are illustrated in Fig. 6. The germ classifier can recognize the shape or form of the dental germ like human experts.

The mean absolute errors between the automatic and manual overall dental age calculations by the four experts using single selection, weighted average, and expected

value to convert the probability of each dental germ's developmental stage to each dental age are described in Table 2. The weighted average was better than the other methods for the conversion to dental age.

Discussion

In this study, our dental germ detector using Scaled-YOLOv4 P6 with an input size of 1280×1280 achieved a very high AP₅₀ of over 98% by cross-validation, as presented in Table 1. The training data, which was much larger than that of previous studies [14, 15], was sufficient for our model to learn the features of images. Generally, there were two choices for object identification from the image: semantic segmentation and object detection. Since pixel-level annotation for semantic segmentation was costly and erroneous and object detection was better at handling overlapping objects [19], we selected object detection for germ detection.

In addition, since the method of obtaining panoramic radiography with optimal quality has been established [42], our models could achieve high performance by learning dental germ features, including background images, overlapping with other objects, or relative position to other dental germs. Therefore, Scaled-YOLOv4 or older YOLO families [25, 27] may have sufficiently detected dental germs, and the newest, but computationally time-consuming models, such as YOLOv7 [43], were not necessary.

However, using state-of-the-art image models for dental germ-stage classification is important. Despite EfficientNet V2's exceptional performance, the Top-1 accuracy of our germ classifier was approximately 70%, as presented in Table 1. This might be because overlapping with other objects or background images, thought to be good for germ detection models, negatively affects the germ classification model. Therefore, we utilized one of the state-of-the-art but computationally expensive models for germ classification. Our germ classification models are considered to be similar to human experts and are clinically applicable with reasonable accuracy. Our models focus on the crown shape or root formation of the dental germ to classify developmental stages, as illustrated in Fig. 6 like human experts. In addition, our model tends to misclassify adjacent stages using the confusion matrix in Fig. 5. This tendency was observed in previous research and also in real-world dentists because the adjacent stages of the dental germ looks similar and the many variations of the morphological structure can be observed between developmental stages [8, 44]. This is why we achieved an exceptional Top-3 accuracy of 98%. It is also the reason we adopted the Top-3 weighted averages to calculate dental age, which reduced the mean absolute error between the automatic and manual

Table 1 Average precisions of our germ detector and accuracy of the developmental stage classifier

Fold	Germ detection AP ₅₀	Developmental stage classification			
		Single-root		Multi-root	
		Top-1	Top-3	Top-1	Top-3
1	98.32	66.68	98.32	72.35	98.16
2	98.38	68.43	98.54	71.61	98.62
3	98.14	68.23	98.57	71.74	98.41
4	98.23	68.16	98.13	71.01	98.16
5	98.23	70.03	98.76	71.01	98.44
Average	98.26	68.30	98.46	71.54	98.36

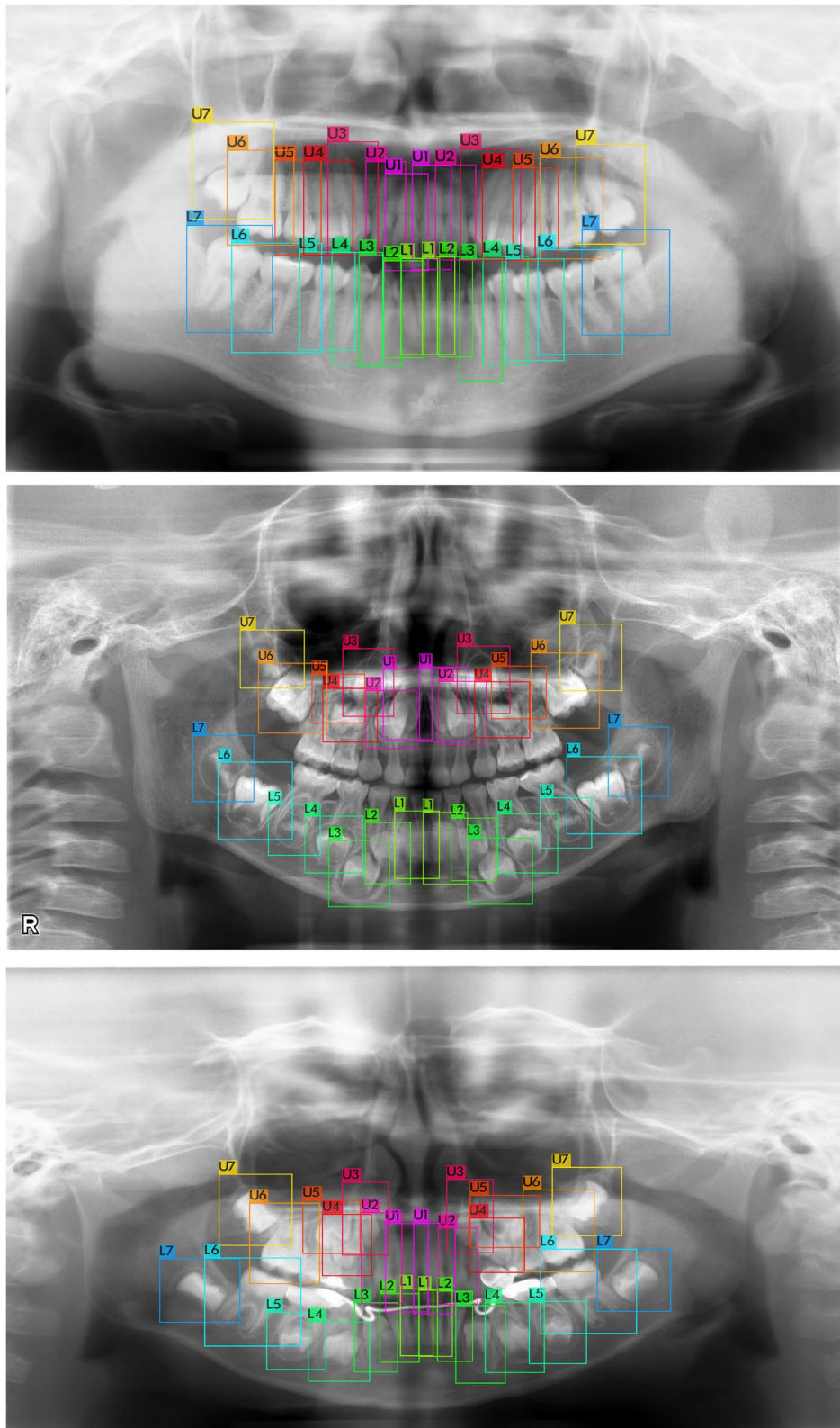


Fig. 2 Examples of germ detection without congenitally missing teeth. All germs were detected correctly

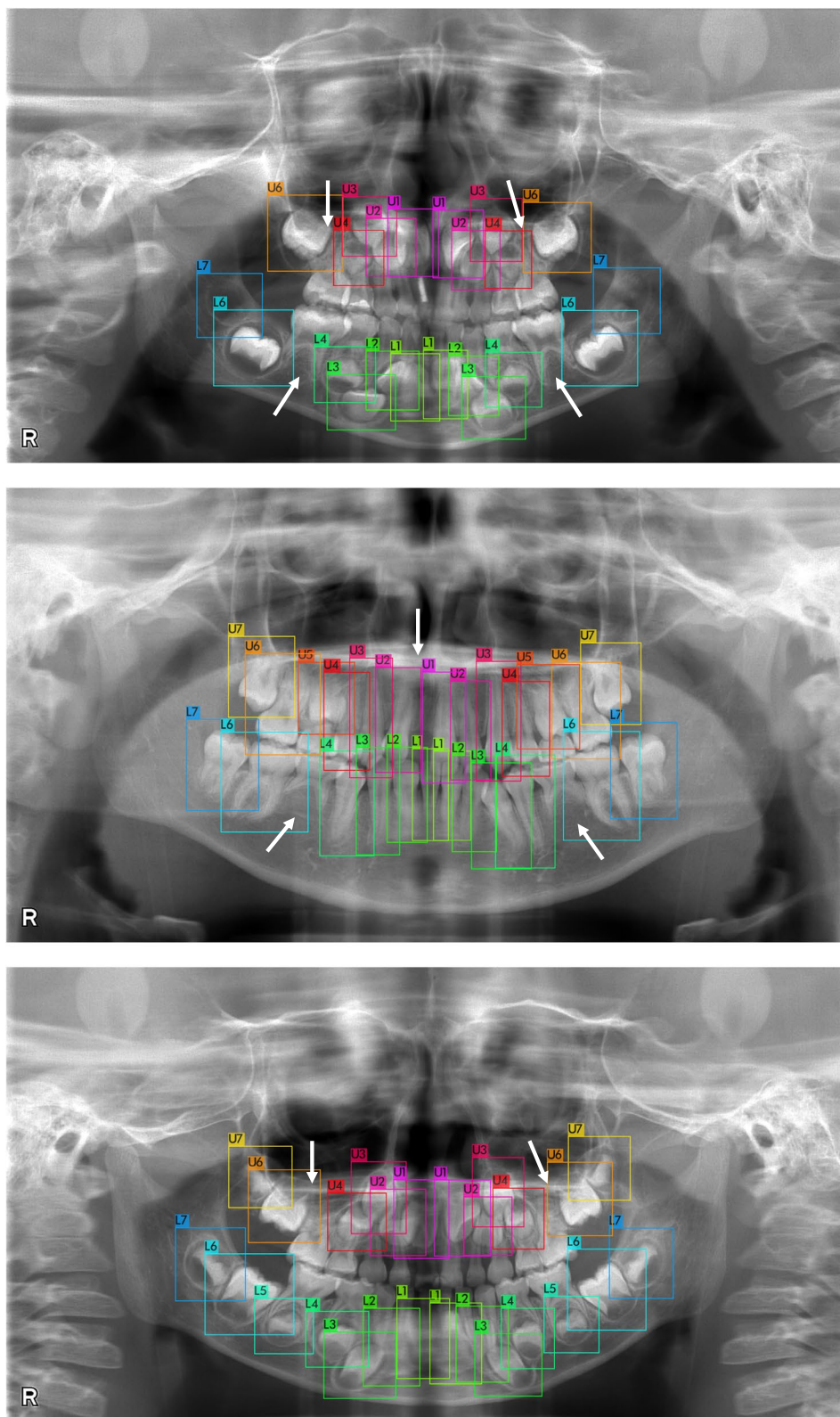


Fig. 3 Examples of germ detection with congenitally missing teeth. Arrows indicate the points where dental germs are absent and not detected

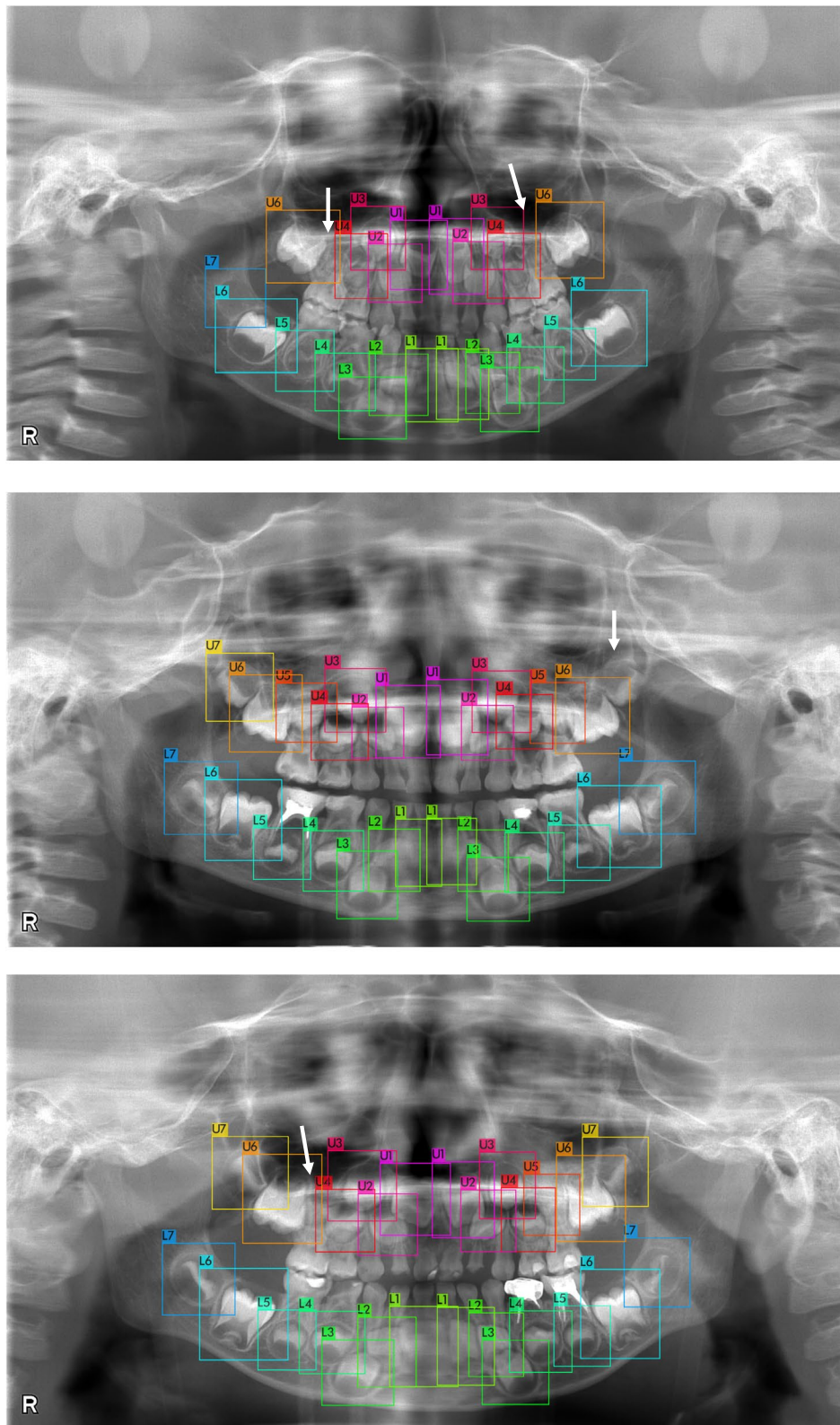
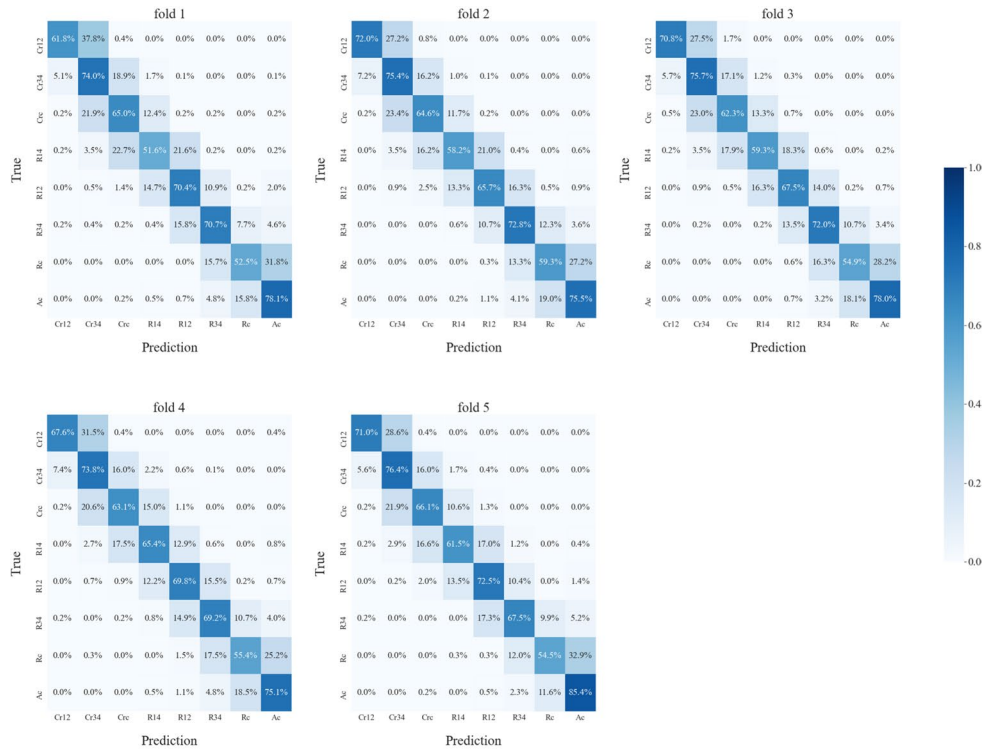


Fig. 4 Examples of germ detection with some failure. Some germs were not detected. Arrows indicate points where dental germs are present but not detected

Single root



Multi root

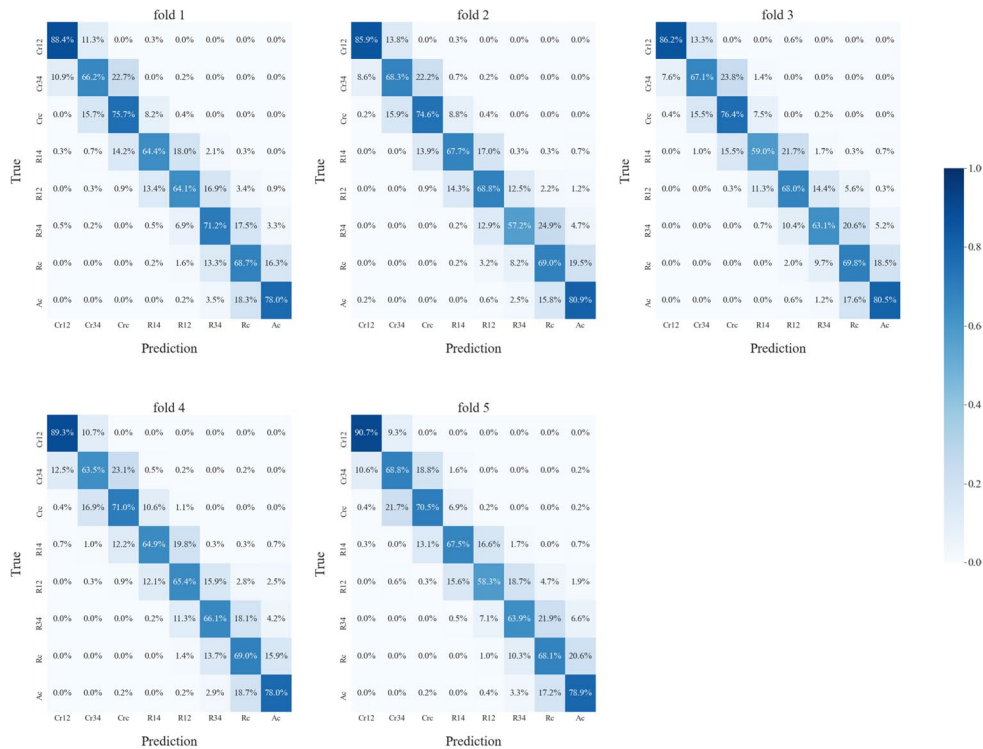


Fig. 5 Results of germ classifier for single- or multi-root dental germ images obtained for each fold of cross-validation training. Those confusion matrices are normalized by the number of elements in each class to reveal each class's accuracy

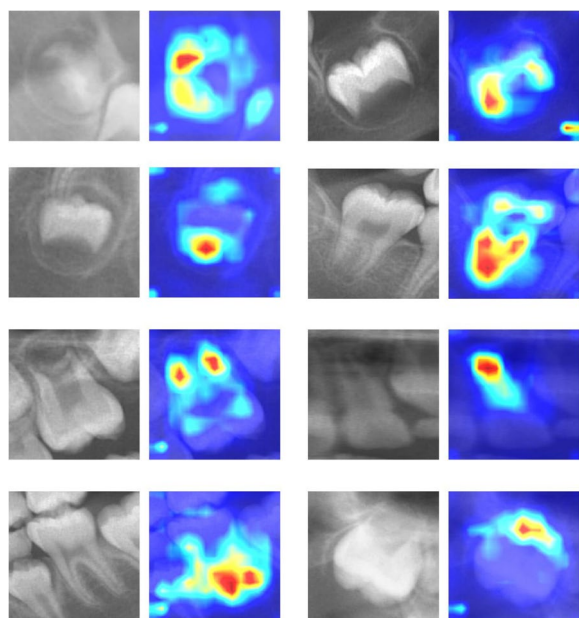


Fig. 6 Examples of representative dental germ images and corresponding Grad-CAM of germ classifier for interpretability and explainability. Germ classifier can recognize the shape or form of the dental germ

Table 2 The mean absolute errors between the automatic and manual overall dental age calculations

Calculation method	Mean absolute error	Standard deviation
Single selection	0.274	0.342
Weighted average	0.261	0.274
Expected value	0.396	0.588

calculations by experts, as presented in Table 2. The single selection showed a similar mean absolute error, but the standard deviation was worse than the weighted average, indicating that the calculated value may spread out over a wide range and may be far from the actual dental age. The expected value showed the worst result, which suggested that using all data may be noisier than using Top-3 accurate values.

Our germ detection model achieved a high AP_{50} of 98%; nonetheless, a few dental germs were sometimes not detected, as illustrated in Fig. 4. However, this may not be critical as regards dental age calculation because we can still use over 20 dental germs and average them for calculation despite several germ detection failures. Thus, our automatic calculation method is robust against detection failures.

Our automatic dental age calculation achieved a mean absolute error of 0.261 years (about 3 months) compared with human experts, raising the concern of this difference

being clinically acceptable. Most previous studies have focused on chronological age estimation [20–23, 45], whereas our research aimed to evaluate dental age calculation. Therefore, our results are not directly comparable to those of previous studies. One potential metric to evaluate our results can be the difference in years between the developmental stages of the teeth. For each teeth, the minimum difference between each developmental stage and its adjacent stage is 0.4 years [8]. Our model has achieved better result of 0.261 years, indicating that our automatic dental age calculation is accurate by less than one developmental stage error and thus is acceptable for supporting dentists. Moreover, the automatic calculation can be performed in a few minutes, which is significantly faster than manual calculation [12] and is useful not only for pediatric or orthodontic dentists but also for general dentists and even students. We believe that our results will serve as a new benchmark for further research in dental age calculation.

Our method can easily be applied to other dental age calculation methods based on developmental stage assessment [9–11]. For those methods, first, determine the dental germs should be determined. Then the developmental stages should be classified to obtain dental age, using the procedure described in our model. If another method is to be used, the model should be modified to change the calculation algorithm, including the volume assessment of teeth, pulp-to-tooth ratio method, coronal pulp cavity measurement, and open apices method [4, 5, 7].

Our proposed model can be useful not only for dental age calculation using various methods but also other clinically supportive applications. When there were congenitally missing teeth on the panoramic radiograph, the germ detector did not respond to the missing teeth’s location, as shown in Fig. 3. This behavior can inform dentists about missing teeth, which is a crucial factor in treatment planning. Moreover, our germ classifier can help human experts improve their diagnostic skills for developmental stage classification by receiving feedback from the decision-making process illustrated in Fig. 6. In the future, human and AI collaboration in dentistry will be expected in academic education and clinical practice [13, 46, 47].

This study has some limitations. Although the number of training datasets in this study is much larger compared to that in previous studies in the field of dentistry, it is still small compared to that in other fields. For example, ImageNet consists of 14 million natural images [48], MS-Celeb-1M has 10 million face images [49], and RadImageNet provides 1.35 million medical images [50]. There may be some room for further improvement of automatic calculation performance by training with a larger dataset. Also, our datasets contain a relatively large number

of healthy images. To reduce this bias and to overcome imbalanced datasets, adding the public datasets such as Tufts Dental Database [51] or federation learning across multiple medical institutions [52] could be solutions to be consider.

Another limitation is that our datasets lack metadata like chronological age or sex because of ethical reasons. In particular, since the metadata of sex and race are important factors for dental age estimation, they are necessary to evaluate the difference between our results and further studies in which metadata are available. Additionally, if the metadata of age are available, our model can be modified to calculate not only dental age but also chronological age, which is useful in forensic science [45]. Thus, a large-scale dental image dataset, which has metadata and is annotated by experts, is expected to help in developing successful AI models in dentistry.

Conclusion

In this study, we achieved automatic dental age calculation with a clinically acceptable error compared to manual calculation by human experts using two-stage deep learning with high accuracy in dental germ detection and developmental stage classification. Dental age is crucial for treatment planning in pediatric and orthodontic dentistry, and our method supports faster dental treatment planning than that with manual calculation.

Acknowledgements

Not applicable.

Authors' contributions

K.K. wrote the main manuscript text and contributed to the study design, data acquisition, analysis, and interpretation. K.K., R.K., A.M., R.O., and K.No. contributed to the data acquisition, clinical analysis, and interpretation. K.K. and K.No. contributed to network analysis and interpretation. All authors reviewed and approved the final manuscript.

Funding

This research was partially supported by ASAHIROENTGEN IND. CO., LTD. (<https://asahi-xray.co.jp/>) and partially supported by JSPS KAKENHI Grant Number JP21K12725.

Availability of data and materials

The datasets generated or analyzed during the current study are not publicly available to protect patient privacy but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The Ethics Committee of the Osaka University Graduate School of Dentistry approved this study (approval: R3-E27). The requirement for informed consent was waived, due to the retrospective nature of the study. All methods used in this study were performed in accordance with the Act on the Protection of Personal Information and Ethical Guidelines for Medical and Health Research Involving Human Subjects. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Division for Medical Informatics, Osaka University Dental Hospital, 1-8 Yamada-oka, Suita, Osaka 565-0871, Japan. ²Department of Pediatric Dentistry, Osaka University Graduate School of Dentistry, 1-8 Yamada-oka, Suita, Osaka 565-0871, Japan.

Received: 26 June 2023 Accepted: 23 January 2024

Published online: 31 January 2024

References

1. Bagherian A, Sadeghi M. Assessment of dental maturity of children aged 3.5 to 13.5 years using the Demirjian method in an Iranian population. *J Oral Sci.* 2011;53(1):37–42.
2. Arciniaga Ramos NA. Comparative analysis between dental, skeletal and chronological age. *Rev Mex Ortodon.* 2013;1(1)
3. Mutiara Sukma S, Ira A, Lucy P. The differences of chronological age with dental age based on the alqahtani method aged 6-12 years. *J Med Dent Sci.* 2021;1(1):61–71.
4. Puranik M, Priyadarshini C, Uma SR. Dental age estimation methods: a review. *Int J Adv Health Sc Tech.* 2015;1:19–25.
5. Stavrianos C, et al. Dental age estimation of adults: a review of methods and principals. *Res J Med Sci.* 2008;2:258–68.
6. Ohtani S, et al. Racemization of aspartic acid in human cementum with age. *Arch Oral Biol.* 1995;40(2):91–5.
7. Panchbhai AS. Dental radiographic indicators, a key to age estimation. *Dentomaxillofac Radiol.* 2011;40(4):199–212.
8. Kuremoto K, et al. Estimation of dental age based on the developmental stages of permanent teeth in Japanese children and adolescents. *Sci Rep.* 2022;12(1):3345.
9. Haavikko K. The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantomographic study. *Suom Hammaslaak Toim.* 1970;66(3):103–70.
10. Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. *Hum Biol.* 1973;45(2):211–27.
11. Nolla CM. The development of the permanent teeth. *J Dent Child.* 1960;27:254–66.
12. Kapoor P, Jain V. Comprehensive chart for dental age estimation (DAecc8) based on Demirjian 8-teeth method: simplified for operator ease. *J Forensic Legal Med.* 2018;59:45–9.
13. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res.* 2020;99(7):769–74.
14. Hwang JJ, et al. An overview of deep learning in the field of dentistry. *Imaging Sci Dent.* 2019;49(1):1–7.
15. Khanagar SB, et al. Developments, application, and performance of artificial intelligence in dentistry - a systematic review. *J Dent Sci.* 2021;16(1):508–22.
16. Başaran M, et al. Diagnostic charting of panoramic radiography using deep-learning artificial intelligence system. *Oral Radiol.* 2022;38(3):363–9.
17. Vinayahalingam S, et al. Automated chart filing on panoramic radiographs using deep learning. *J Dent.* 2021;115:103864.
18. Kim J, et al. Deep learning-based identification of mesiodens using automatic maxillary anterior region estimation in panoramic radiography of children. *Dentomaxillofac Radiol.* 2022;51(7):20210528.
19. Yang J, et al. Automated Dental Image Analysis by Deep Learning on Small Dataset. In: 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), vol. 01. Tokyo, Japan: IEEE; 2018. p. 492–7.
20. Wallraff S, et al. Age estimation on panoramic dental X-ray images using deep learning. Wiesbaden: Springer Fachmedien Wiesbaden; 2021.
21. Milošević D, et al. Automated estimation of chronological age from panoramic dental X-ray images using deep learning. *Expert Syst Appl.* 2022;189:116038.
22. Parlak Baydoğan M, Coşgun Baybars S, Arslan Tuncer S. Age detection by deep learning from dental panoramic radiographs. *Artif Intell Theory Appl.* 2022;2(2):51–8.

23. Vila-Blanco N, et al. Deep neural networks for chronological age estimation from OPG images. *IEEE Trans Med Imaging*. 2020;39(7):2374–84.
24. Wang CY, Bochkovskiy A, Liao HYM. Scaled-YOLOv4: Scaling Cross Stage Partial Network. In: *IEEE/Cvf conference on computer vision and pattern recognition 2021*. *Cvpr*; 2021. p. 13024–33.
25. Bochkovskiy, A., C.-Y. Wang, and H.-Y.M. Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection. 2020:arXiv:2004.10934.
26. Harris EF. Tooth-coding Systems in the Clinical Dental Setting. *Dent Anthropol J*. 2018;18(2):43–9.
27. Redmon, J. and A. Farhadi, YOLOv3: An Incremental Improvement. 2018: arXiv:1804.02767.
28. Lin TY, et al. Focal loss for dense object detection. In: *IEEE International Conference on Computer Vision (Iccv), 2017; 2017*. p. 2999–3007.
29. Li X, et al. Generalized Focal Loss V2: Learning Reliable Localization Quality Estimation for Dense Object Detection. In: *IEEE/Cvf conference on computer vision and pattern recognition, Cvpr 2021; 2021*. p. 11627–36.
30. Liu W, et al. SSD: single shot MultiBox detector. *Computer vision - Eccv 2016, Pt I*, vol. 9905; 2016. p. 21–37.
31. Tan M, Pang R, Le QV. EfficientDet: Scalable and Efficient Object Detection. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020; 2020*.
32. Berrar D. Cross-validation. In: Ranganathan S, et al., editors. *Encyclopedia of bioinformatics and computational biology*. Oxford: Academic Press; 2019. p. 542–5.
33. Tan MX, Le QV. EfficientNetV2: smaller models and faster training. In: *International Conference on Machine Learning, Vol 139; 2021*. p. 7102–10.
34. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: *36th International Conference on Machine Learning, ICML 2019; 2019*. p. 10691–700.
35. He KM, et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr)*. p. 770–8.
36. Huang G, Liu Z, Weinberger KQ. Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016*. p. 2261–9.
37. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. In: *30th IEEE Conference on Computer Vision and Pattern Recognition (Cvpr 2017); 2017*. p. 1800–7.
38. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. *Entropy*. 2021;23(1):18.
39. Tjoa, E. and C.T. Guan, A survey on explainable artificial intelligence (XAI): toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 2021. 32(11): 4793–4813.
40. Selvaraju RR, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis*. 2020;128(2):336–59.
41. Okawa R, Kokomoto K, Nakano K. Dental effects of enzyme replacement therapy in case of childhood-type hypophosphatasia. *BMC Oral Health*. 2021;21(1):323.
42. Różyło-Kalinowska I. Panoramic radiography in dentistry. *Clin Dent Rev*. 2021;5(1):26.
43. Wang, C.-Y., A. Bochkovskiy, and H.-Y.M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 2022: arXiv:2207.02696.
44. Mohammad N, et al. Accuracy of advanced deep learning with tensorflow and keras for classifying teeth developmental stages in digital panoramic imaging. *BMC Med Imaging*. 2022;22(1):66.
45. Chaillet N, Nyström M, Demirjian A. Comparison of dental maturity in children of different ethnic origins: international maturity curves for clinicians. *J Forensic Sci*. 2005;50(5):1164–74.
46. Hong X, et al. Can AI Teach Humans? Humans AI Collaboration for Lifelong Machine Learning. In: *2021 4th International Conference on Data Science and Information Technology; 2021*. p. 427–32.
47. Kokomoto K, et al. Intraoral image generation by progressive growing of generative adversarial network and evaluation of generated image quality by dentists. *Sci Rep*. 2021;11(1):18517.
48. Russakovsky O, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211–52.
49. Guo Y, et al. MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition; 2016. p. 87–102.
50. Mei X, et al. RadImageNet: an open radiologic deep learning research dataset for effective transfer learning. *Radiol Artif Intell*. 2022;4(5):e210315.
51. Panetta K, et al. Tufts dental database: a multimodal panoramic X-ray dataset for benchmarking diagnostic systems. *IEEE J Biomed Health Inform*. 2022;26(4):1650–9.
52. Sheller MJ, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep*. 2020;10(1):12598.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.