Age Estimation using Panoramic Radiographs by Transfer Learning

Chuang Chuang MU[1], Gang LI[1]

Objective: *To assess the accuracy of transfer learning models for age estimation from panoramic photographs of permanent dentition of patients with an equal sex and age distribution and provide a new method of age estimation.*
Methods: *The panoramic photographs of 3000 patients with an equal sex and age distribution were divided into three groups: a training set (n = 2400), validation set (n = 300) and test set (n = 300). The ResNet, EffiecientNet, VggNet and DenseNet transfer learning models were trained with the training set. The models were subsequently tested using the data in the test set. The mean absolute errors were calculated and the different features extracted by the deep learning models in different age groups were visualixed.*
Results: *The mean absolute error (MAE) and root mean square error (RMSE) of the optimal transfer learning model EfficientNet-B5 in the test set were 2.83 and 4.59, respectively. The dentition, maxillary sinus, mandibular body and mandibular angle all played a role in age estimation.*
Conclusion: *Transfer learning models can extract different features in different age groups and can be used for age estimation in panoramic radiographs.*
Key words: *age estimation, deep learning, forensic odontology, panoramic radiograph, transfer learning*
Chin J Dent Res 2022;25(2):119–124; doi: 10.3290/j.cjdr.b3086341

Age estimation based on panoramic radiographs is essential in forensic science and anthropological research. Several methods have been proposed for age estimation in children and adolescents. Among these, those proposed by Demirjian et al[1] and Nolla et al[2] are adopted frequently. Their methods define the different stages of tooth development by the first appearance of calcified points to the closure of the tooth apex in panoramic radiographs and score them accordingly, and the scores are then converted to a dental age from 3 to 17 years. Willems et al[3] modified the scoring system proposed by Demirjian et al[1] and applied it to Belgian children, achieving relatively high accuracy in age estimation. Cameriere et al[4] established a regression model according to age with measurements of open apices in different teeth. The estimated age obtained from this model has a residual error of 0.035[4]. For adults, age estimation is based on the negative correlation between age and pulp/pulp chamber size. Mittal et al[5] measured the lengths and widths of six anterior teeth at different levels in 152 panoramic radiographs and performed regression analysis for age estimation. The standard error from this regression model was 7.97 between chronological and estimated age[5]. Cameriere et al[6] analysed the pulp/tooth

1 Department of Oral and Maxillofacial Radiology, Peking University School and Hospital of Stomatology, National Centre of Stomatology, National Clinical Research Centre for Oral Diseases, National Engineering Laboratory for Digital and Material Technology of Stomatology, Beijing Key Laboratory of Digital Stomatology, Research Centre of Engineering and Technology for Computerised Dentistry Ministry of Health, NMPA Key Laboratory for Dental Materials, Beijing, P.R. China.

Corresponding author: Dr Gang LI, Department of Oral and Maxillofacial Radiology, Peking University School and Hospital of Stomatology, #22 Zhongguancun South Avenue, Haidian District, Beijing,100081, P.R. China. Tel: 86-10-82195328; Fax: 86-10-82195328. Email: kqgang@bjmu.edu.cn

area ratios of the premolars in panoramic radiographs of 606 patients. The mean absolute errors (MAEs) obtained from the regression model established in the study ranged from 4.43 to 6.02 (at a 95% confidence interval)[6]. Manual extraction of tooth development features for age estimation is used mainly in traditional algorithms, but this is time-consuming and complicated.

Deep learning, an artificial intelligence technique, has been widely applied to the automatic extraction of image features. This has led to an improvement in traditional dental imaging tasks[7-10]. Vila-Blanco et al[9] and Guo et al[11] used end-to-end neural networks based on panoramic radiographs for age estimation, with promising results. The data sets in their studies consisted primarily of samples from patients aged under 25 years. Whether deep learning performs equally well for other dental images, however, is unknown. Transfer learning[12] is a common deep learning technique that is used for similar tasks in computer vision. In this process, one pretrained model can be used as the starting point for another model aimed at another similar task. Thus, the purpose of the present study was to use different transfer learning models for age estimation in panoramic radiographs and explore the extracted features in different age groups.

Materials and methods

A data set containing 3000 panoramic radiographs was collected retrospectively from the database of Peking University School and Hospital of Stomatology. The sample consisted of 1500 females and 1500 males, aged between 12 and 71 years. All the patients' ages were confirmed in the hospital's patient information system. Their age and sex distributions are shown in Table 1.

Image acquisition

The panoramic radiographs were all acquired using panoramic radiograph equipment (Hyperion X9, MyRay, Cefla, Imola, Italy) with exposure parameters of 60~85 kV, 1~10 mA and 13.0 seconds according to patient size.

The panoramic radiographs were reviewed by a professional oral and maxillofacial radiologist. The inclusion criteria were as follows:

- no fractures or large defects in either the maxilla or mandible;
- no space-occupying lesions;
- no primary teeth;
- no systemic diseases or developmental delays.

Transfer learning models

In this study, ResNet[13], EfficientNet[14], VggNet[15] and DenseNet[16] models pretrained on ImageNet[17] (Tampa, FL, USA) images were used to extract features in panoramic radiographs for automatic age estimation. ImageNet is a large-scale hierarchical image database.

Image processing and augmentation

All 3000 panoramic radiographs were divided into three sets: a training set (n = 2400, 80%), a validation set (n = 300, 10%) and a test set (n = 300, 10%). The division was conducted by keeping age and sex equal within the three sets. All the panoramic radiographs had a 24-bit colour depth and a height and width of 2500 × 1248 pixels. The images were padded while preserving the aspect ratio and resized to 224 × 224 pixels to satisfy the pretrained model input size. Random horizontal flip was the only image augmentation technique used due to the fixed image acquisition parameters for the panoramic radiographs used in the present study.

Model training and evaluation

The fully connected layers of pretrained models were modified to target directly presenting the estimated age. The softmax function was used to change the activation function in the last fully connected layer into a probability distribution, defined as follows:

$$p_i = \frac{\exp(x_i)}{\sum_{n=1}^{N} \exp(x_n)},$$

where ρ_i represents the probability.
The estimation age was then calculated as:

$$\hat{y} = \sum_{i=1}^{K} i\, p_i,$$

where ŷ represents the estimation age.
The loss function of the model was set to the sum of the softmax loss and the mean-variance loss[18], then the training set was fed into the model and the epoch was set as 200, with an early stop strategy if the loss result did not improve after 20 successive epochs. The network performance could be assessed with the validation set for hyperparameter adjustment and to determine the optimal deep learning model.

MAEs and RMSEs have been commonly used in age estimation studies[4,6,9,11]. Thus, this study applied MAE

and RMSE to evaluate the accuracy of age estimation with the transfer learning models in the test set. The estimated ages in different age and sex groups were recorded for further analysis.

The MAE is the mean difference between the estimated and chronological age:

$$MAE = \frac{1}{N}\sum_{n=1}^{N} |\hat{y}_n - y_n| ,$$

where ŷ and y represent the estimated and chronological age, respectively, and n is the counter for the images.

The RMSE is the standard deviation (SD) of the residuals (estimated errors between the estimated and chronological age):

$$\mathrm{RMSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2,$$

Statistical analysis and feature visualization

A Bland-Altman plot was used to compare the chronological and estimated age returned by the optimal transfer learning model. Class activation mapping[19] (CAM) was employed to visualise the attention regions for age estimation. Anatomically significant areas in the maps were marked and analysed.

Results

The MAEs and RMSEs of the transfer learning models for age estimation with the test set are displayed in Table 2. Different pretrained models had different accuracies in age estimation. The optimal model in this study was EfficientNet-B5, with an MAE of 2.83 and RMSE of 4.59, while the maximum MAE for Vgg19 and ResNet101 was 5.26 and the maximum RMSE for ResNet101 was 7.19. Further results from EfficientNet-B5 for different age groups and sexes with the test set are shown in Table 3.

The plots for chronological age versus estimated age by EfficientNet-B5 with the test set (n = 300) are shown in Fig 1. The points are close to the diagonal line, especially those representing ages under 41 years. Points falling directly on the diagonal line indicate a perfect match between chronological and estimated age.

The Bland-Altman plot was created using SPSS Statistics 24.0 (SPSS, Chicago, IL, USA). The differences between chronological age and the age estimated using EfficientNet-B5 are shown in Fig 2. The mean difference, chronological age minus estimated age, was 0.0 years and the SD was 4.59. Thus, the lower 95% limit was 0.00 − 1.96 × 4.59 = −9.00 years and the upper 95% limit was 0.00 + 1.96 × 4.59 = 9.00 years. A total of 285 points out of 300 were within the 95% limitsof agreement. Thus, this model provides reliable estimates of agreement in this study.

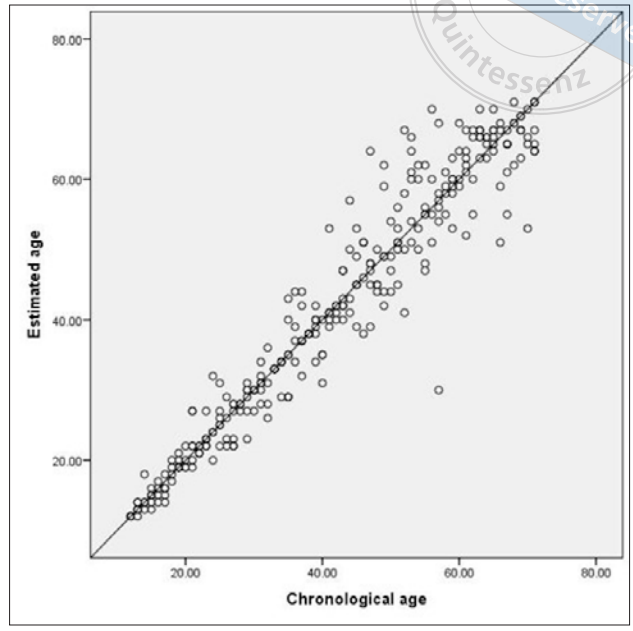

Fig 1 Chronological age versus estimated age.

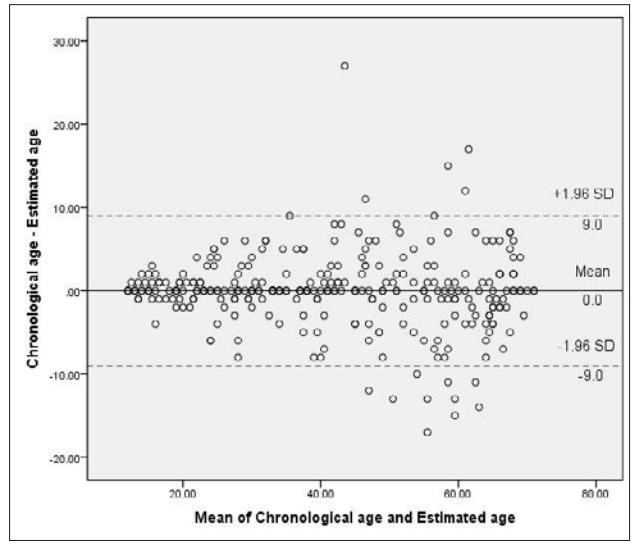

Fig 2 Differences between chronological and estimated ages versus mean of chronological and estimated ages, with 95% limits of agreement (broken lines).

The class activation mapping results for different age groups are shown in Fig 3. Differently coloured areas represent the weights according to the colour bar. The colour maps for the groups aged 12 to 21 years (Figs 3a and b) and 22 to 31 years (Figs 3c and d) lie more in the dentition than in the other parts of the panoramic

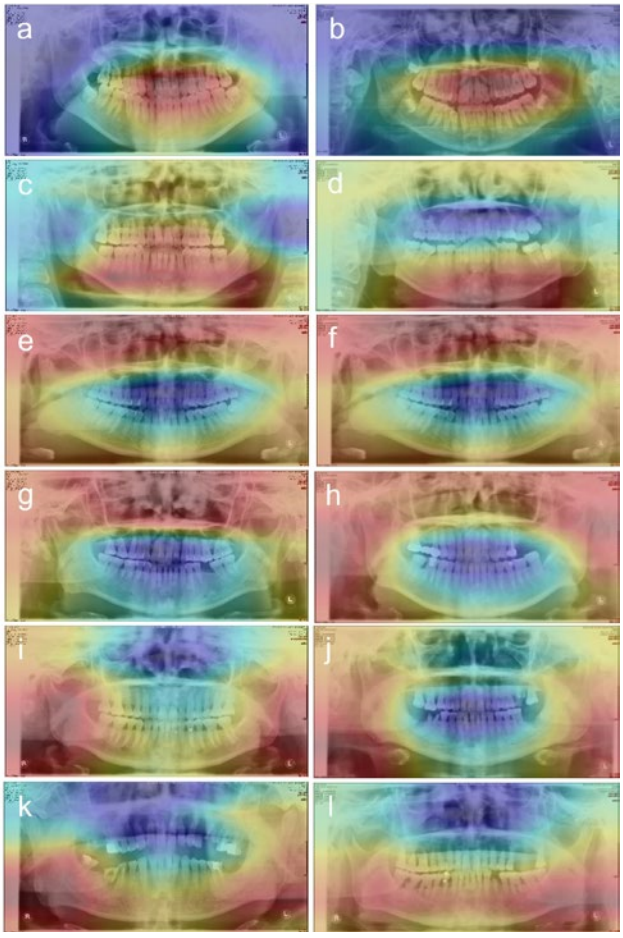

Fig 3 Class activation mapping results for the different age groups.

radiographs. The maxillary sinus was of greater concern for the groups aged 32 to 41 (Figs 3e and f) and 42 to 51 years (Figs 3g and h). The mandibular body and mandibular angle were more important in the groups aged 52 to 61 (Figs 3i and j) and 62 to 71 years (Figs 3k and l).

Discussion

In this study, transfer learning models based on ResNet, EfficientNet, VggNet and DenseNet were established using panoramic radiographs for two reasons. First, these models have been verified on ImageNet and have demonstrated good classification performance, and second, they have been applied in similar related studies[11,20,21].

The minimum and maximum MAEs and RMSEs of the transfer learning models used for age estimation with the test set were 2.83 (EfficientNet-B5) and 5.26 (Vgg19 and ResNet101), 4.59 (EfficientNet-B5) and 7.19 (ResNet101), respectively. Other transfer learning models also showed different levels of accuracy in age estimation. These models are trained, validated and tested on the same data; thus, the differences in the models' architecture cause the differences in the final accuracies. Deep learning is a black box, and the principal differences between different models' architectures need to be further studied by computer vision, and customised architecture for tooth age may perform better.

Table 1 Age and sex distribution of the data set.

Age (y)	Male	Female	Total
12–21	250	250	500
22–31	250	250	500
32–41	250	250	500
42–51	250	250	500
52–61	250	250	500
62–71	250	250	500
Total	1500	1500	3000

Using the optimal EfficientNet-B5 model, women in the group aged 22 to 31 years had the smallest estimation error (MAE 0.96, RMSE 1.52), whereas men in the group aged 52 to 61 years had the largest (MAE 5.12, RMSE 7.03). In general, the estimation error between chronological age and estimated age increased as age increased, as shown in Table 3. The Bland-Altman plot (Fig 2) shows that the differences increased with the mean, which is consistent with Table 3 and Fig 1.

The results of the class activation mapping suggest that in different age groups, different anatomical structures are considered. In the younger age groups (12 to 21 and 22 to 31 years), the extracted features were primarily in the dentition, which is consistent with traditional methods. In the middle age groups (32 to 41 and 42 to 51 years), the feature areas moved to the maxillary sinus. Jun et al[22] analysed the volume of the maxillary sinus using high-resolution computed tomography (CT) and found that maximum growth was reached in the fourth decade of life in men and in the third decade of life in women. Aktuna et al[23] found that maxillary sinus volume decreases as age increases, and the present results confirm these findings. The mandibular body and mandibular angle were emphasised in the older age groups (52 to 61 and 62 to 71 years). This coincides with the findings from the study by Upadhyay et al[24], in which the investigators used physicoforensic anthropometry and lateral cephalometric methods to measure 185 subjects and obtained results showing a decrease in the mandibular angle as age increased. The finding that the deep learning models extracted these features shows that these anatomical structures also have the potential to aid age estimation.

Table 2 MAE and RMSE for different pretrained models.

Model	Both sexes		Female		Male	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
ResNet18	2.95	4.69	2.99	4.96	2.92	4.42
ResNet50	3.78	5.98	3.93	6.21	3.68	5.75
ResNet101	5.26	7.19	5.10	7.11	5.43	7.28
Vgg16	4.51	5.96	4.40	5.96	4.63	5.97
Vgg19	5.26	6.85	5.48	7.29	5.04	6.36
EfficientNet-B1	4.59	6.23	4.60	6.38	4.59	6.07
EfficientNet-B3	3.87	5.44	3.85	5.70	3.89	5.16
EfficientNet-B5	2.83	4.59	2.83	4.79	2.83	4.38
DenseNet121	3.15	4.81	3.00	4.79	3.30	4.84

Table 3 MAE and RMSE with different age groups and sexes for EfficientNet-B5.

Age group	Both sexes		Female		Male	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
12–21 y	1.06	1.70	1.08	1.78	1.04	1.62
22–31 y	1.64	2.56	0.96	1.52	2.32	3.28
32–41 y	2.42	3.87	2.76	3.96	2.08	3.77
42–51 y	3.86	5.22	3.68	5.56	4.04	4.85
52–61 y	4.78	7.01	4.44	6.99	5.12	7.03
62–71 y	3.52	5.06	4.08	6.05	2.96	3.84
Total	2.83	4.59	2.83	4.79	2.83	4.38

Vila-Blanco et al[9] used 2289 panoramic radiographs to establish two convolutional neural network models for age estimation. The MAEs obtained from the two models were 3.19 ± 4.32 and 2.84 ± 3.75, respectively. The data set was unbalanced and, most importantly, the subjects involved were aged mostly under 20 years (1381/2289)[9]. This may have caused bias in the age estimation for adults. Guo et al[11] collected 10,257 panoramic radiographs from patients aged 5 to 24 years and developed end-to-end neural networks to compare their predictions with the manual method (by Demirjian et al[1]). Their results proved that conventional neural network models can surpass the manual method in age classification; however, in their study, the subjects were all aged under 25 years, and they did not provide the MAE or explore the ability of the neural networks to extract features at other ages, as was done in this study.

Although we collected panoramic radiographs from patients with equal distributions of age and sex to avoid confounding factors that might impact the results of such a study as far as possible, limitations still exist. First, only a few pretrained deep learning models were used, and other neural network architectures and pretrained models were not compared, thus the differences across different model architectures need to be explored further. Second, deep learning usually requires a large-scale data set. Although the data set used in the present study was notably large and similar in size to other studies, it was still smaller than the typical datasets used for facial tasks[20].

Conclusion

Transfer learning models can be used for age estimation with panoramic radiographs. Differences between the different sexes and age groups were also observed and presented. Class activation mapping showed that different anatomical features were used for age estimation in different age groups. The role of these features in age estimation needs to be studied further.

Conflicts of interest

The authors declare no conflicts of interest related to this study.

Author contribution

Dr Chuang Chuang MU contributed to data preparation, established models and drafted the manuscript; Dr Gang LI contributed to design, supervised the study and revised the manuscript.

(Received May 19, 2021; accepted July 15, 2021)

References

1. Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. Hum Biol 1973;45:211–227.
2. Nolla CM. The development of permanent teeth. J Dent Children 1960;27:254–266.
3. Willems G, Van Olmen A, Spiessens B, Carels C. Dental age estimation in Belgian children: Demirjian's technique revisited. J Forensic Sci 2001;46:893–895.
4. Cameriere R, Ferrante L, Cingolani M. Age estimation in children by measurement of open apices in teeth. Int J Legal Med 2006;120: 49–52.
5. Mittal S, Nagendrareddy SG, Sharma ML, Agnihotri P, Chaudhary S, Dhillon M. Age estimation based on Kvaal's technique using digital panoramic radiographs. J Forensic Dent Sci 2016;8:115.
6. Cameriere R, De Luca S, Alemán I, Ferrante L, Cingolani M. Age estimation by pulp/tooth ratio in lower premolars by orthopantomography. Forensic Sci Int 2012;214:105–112.
7. Tuzoff DV, Tuzova LN, Bornstein MM, et al. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofac Radiol 2019;48:20180051.
8. Poedjiastoeti W, Suebnukarn S. Application of convolutional neural network in the diagnosis of jaw tumors. Healthc Inform Res 2018;24:236–241.
9. Vila-Blanco N, Carreira MJ, Varas-Quintana P, Balsa-Castro C, Tomas I. Deep neural networks for chronological age estimation from OPG images. IEEE Trans Med Imaging 2020;39:2374–2384.
10. Farhadian M, Salemi F, Saati S, Nafisi N. Dental age estimation using the pulp-to-tooth ratio in canines by neural networks. Imaging Sci Dent 2019;49:19–26.
11. Guo YC, Han M, Chi Y, et al. Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images. Int J Legal Med 2021;135:1589–1597.
12. Pan SJ, Qiang Y. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 2010;22:1345–1359.
13. He K, Zhang X , Ren S , Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
14. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning, 2019. arXiv:1905.11946
15. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
16. Huang G, Liu Z, Van Der Maaten L, et al. Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:4700–4708.
17. Jia D, Wei D, Socher R, Li L, Kai L, Li F. ImageNet: A Large-scale Hierarchical Image Database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009:248–255.
18. Pan H, Hu H, Shan S, Chen X. Mean-Variance Loss for Deep Age Estimation from a Face. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018:5285–5294.
19. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:2921–2929.
20. Yi D, Lei Z, Li S Z. Age estimation by multi-scale convolutional network. Asian Conference on Computer Vision, 2014:144–158.
21. Gao BB, Xing C, Xie CW, Wu J, Geng X. Deep label distribution learning with label ambiguity. IEEE Trans Image Process 2016;26:2825–2838.
22. Jun BC, Song SW, Park CS, Lee DH, Cho KJ, Cho JH. The analysis of maxillary sinus aeration according to aging process; volume assessment by 3-dimensional reconstruction by high-resolutional CT scanning. Otolaryngol Head Neck Surg 2005;132:429–434.
23. Aktuna Belgin C, Colak M, Adiguzel O, Akkus Z, Orhan K. Three-dimensional evaluation of maxillary sinus volume in different age and sex groups using CBCT. Eur Arch Otorhinolaryngol 2019;276: 1493–1499.
24. Upadhyay RB, Upadhyay J, Agrawal P, Rao NN. Analysis of gonial angle in relation to age, gender, and dentition status by radiological and anthropometric methods. J Forensic Dent Sci 2012;4:29–33.