



Accurate age classification using manual method and deep convolutional neural network based on orthopantomogram images

Yu-cheng Guo^{1,2,3} · Mengqi Han^{1,2} · Yuting Chi³ · Hong Long³ · Dong Zhang^{3,4} · Jing Yang⁴ · Yang Yang⁴ · Teng Chen⁵ · Shaoyi Du³

Received: 29 November 2020 / Accepted: 11 February 2021 / Published online: 4 March 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Age estimation is an important challenge in many fields, including immigrant identification, legal requirements, and clinical treatments. Deep learning techniques have been applied for age estimation recently but lacking performance comparison between manual and machine learning methods based on a large sample of dental orthopantomograms (OPGs). In total, we collected 10,257 orthopantomograms for the study. We derived logistic regression linear models for each legal age threshold (14, 16, and 18 years old) for manual method and developed the end-to-end convolutional neural network (CNN) which classified the dental age directly to compare with the manual method. Both methods are based on left mandibular eight permanent teeth or the third molar separately. Our results show that compared with the manual methods (92.5%, 91.3%, and 91.8% for age thresholds of 14, 16, and 18, respectively), the end-to-end CNN models perform better (95.9%, 95.4%, and 92.3% for age thresholds of 14, 16, and 18, respectively). This work proves that CNN models can surpass humans in age classification, and the features extracted by machines may be different from that defined by human.

Keywords Age estimation · Deep learning · Classification · Orthopantomography

Introduction

Age estimation of individuals plays an important role in the forensic context including identification of the deceased,

immigrants, legal and criminal prosecutions, and paediatrics for clinical diagnosis and treatment planning. The age classification is often strongly requested from authorities, different institutions, or courts when they need to determine whether an individual will be treated as a child, juvenile, or adult [1]. Hagen et al. [2] reviewed 597 expert reports to demonstrate the outcome of forensic age assessments of migrants with doubtful minority declaration and found that there were 37.8% minors in non-criminal cases and 74.5% unaccompanied minors have reached the age of majority. Misclassifications will result in the loss of minor protection or adult impunity [3]. The German Social Code VIII (SGB VIII) has governed that youth welfare offices should determine whether an unaccompanied foreign national is a minor through the professional visual inspection or medical examination [4]. Despite diverse relevant ages in different countries, the legal age for criminal liability is within the scope of 14 and 21 years [5–7]. China's criminal law considers children younger than 14 having no full criminal responsibility, teenagers 14–16 years of age having limited or partial criminal responsibility, and teenagers older than 16 having full criminal responsibility. In Chinese civil law, children over 18 years of age have full civil responsibility [8, 9]. In an ethically sensitive scenario, it is technically unacceptable if a minor is

✉ Shaoyi Du
dushaoyi@xjtu.edu.cn

¹ Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an Jiaotong University, 98 XiWu Road, Xi'an 710004, Shaanxi, People's Republic of China

² Department of Orthodontics, Stomatological Hospital of Xi'an Jiaotong University, 98 XiWu Road, Xi'an 710004, Shaanxi, People's Republic of China

³ Institute of Artificial Intelligence and Robotics, College of Artificial Intelligence, Xi'an Jiaotong University, 28 Xianning West Road, Xi'an 710049, Shaanxi, People's Republic of China

⁴ College of Automation Science and Engineering, Xi'an Jiaotong University, 28 Xianning West Road, Xi'an 710049, Shaanxi, People's Republic of China

⁵ College of Medicine and Forensics, Xi'an Jiaotong University Health Science Center, 76 West Yanta Road, Xi'an 710004, Shaanxi, People's Republic of China

misclassified as an adult. Since the development of bone completes early and is more influenced by environmental and nutritional factors compared with dental development, the dental age estimation is an alternative choice for determining whether an individual was a minor or an adult. Mincer et al. [10], Cameriere et al. [11], and Olze et al. [12] successively established dental age estimation methods through different systems, i.e. the Demirjian staging system, the third molar maturity (I_{3M}), and four-stage classification of the periodontal ligament. In terms of the classification of the legal age thresholds of 14, 16, or 18 years, Cameriere et al. [13, 14] also reported that the cutoff values of the I_{3M} and the second molar maturity (I_{2M}) were efficient indicators that need to be distinguished.

Although the methods mentioned above have been accurately applied in different populations, there are still some limitations with respect to the clinical applications, such as the subjectivity and reproducibility of the technique and measurement bias. Additionally, these methods are labour-intensive and time-consuming. Recently, deep learning has been increasingly popular in medicine, especially convolutional neural networks (CNNs). The need for these advanced technology methods in the field of dentistry is also on the rise, with applications including the detection and classification of individual teeth [15, 16], dental age estimation [17, 18] and classification [19], the detection of apical lesions [20] and periodontal disease [21], and the diagnosis of osteoporosis [22]. In 2021, Galibourg et al. [23] compared the Demirjian and Willems methods with machine learning regression models for age estimation, and found that all the machine learning models based on the developmental stages defined by Demirjian were more accurate in dental age estimation. As for the age classification between minors and adults, Stern et al. [19] first reported a classifier derived using a random forest algorithm and a deep CNN based on magnetic resonance imaging (MRI) of the third molar, with an accuracy of just 85.09%. Deep learning has achieved some success in many fields of medicine; however, previous studies have explored only the performance of deep learning models without a comparison with manual methods, and the sample size of those models was also small. In addition, the reasons for the better performance of deep learning techniques were not clarified. Therefore, it is still necessary to test whether deep learning techniques can replace humans in some circumstances and to explore whether deep learning techniques can find other regulations and features that have not been previously found.

This study constitutes the first report of CNNs for dental age classification based on a large sample of orthopantomograms (OPGs) (10257) and explores the difference between the features defined by humans and the features extracted by machines. We compared traditional manual method based on the Demirjian method with CNN models for legal age threshold classification. In addition, gradient-weighted class activation mapping (Grad-CAM) was used to assess the differences in the features of interest between human beings and machines.

Our results prove that CNN models can surpass humans in age classification and that the features extracted by machines may be different from those defined by human.

Materials and methods

Data acquisition

A total of 10,257 OPGs from 4579 males and 5678 females aged between 5.00 and 24.00 years were analysed. The study group was divided into 20 age groups by the year of age. As in age group 15, patients with ages ranging from 15.00 to 15.99 were involved. The OPGs (Sirona Dental Systems GmbH, Germany. Model-No.: D3352; Software-Version: 02.28; Hardware-Version: BA) were obtained from the records database of the Department of Oral Radiology, Stomatology Hospital of Xi'an Jiaotong University, between 2015 and 2019. All the OPGs were saved in the JPG format. Ethical approval was granted by the ethics committee of the Stomatology Hospital of Xi'an Jiaotong University (No. 2017-524). The inclusion criteria were as follows: (1) northern Chinese origin; (2) normal development and dental conditions; (3) no missing permanent teeth in the left mandible; (4) no pathologic condition affecting the maturity of teeth; and (5) no image distortion or lack of clarity. The distribution by age and gender is shown in Table 1. The records consisted of the gender, date of birth, date of exposure, and identification number.

In the data division of the three experiments, we randomly selected 10% of each category as the validation set, 10% of each category as the test set, and the remaining data as the training set. The training set was used for the establishment of manual methods and the training of pre-proposed deep learning models. The validation set was only used for determining the optimal deep learning models, while the test set was for determining the classification performance of both manual and machine methods.

Manual classification

All OPGs were most separately estimated by two experts with extensive clinical experience and forensic training after a sufficient discussion. To evaluate the dental age, the first step was to estimate the developmental stage of the left mandibular eight permanent teeth in each OPG based on the Demirjian method [24], and then, the second step was to transform stages A to H into numeric stages 1 to 8, respectively. We established a logistic regression linear model with dichotomous variable Y , with 0 representing an individual under the age threshold and 1 representing an individual equal to or over the age threshold. We derived logistic regression linear models for each age threshold to measure the relationship between the

Table 1 The distribution of the sample by age and gender

Age group	Male	Female	Total
5.00–5.99	276	225	501
6.00–6.99	283	222	505
7.00–7.99	303	205	508
8.00–8.99	347	162	509
9.00–9.99	284	232	516
10.00–10.99	266	237	503
11.00–11.99	223	295	518
12.00–12.99	229	273	502
13.00–13.99	213	299	512
14.00–14.99	208	304	512
15.00–15.99	213	297	510
16.00–16.99	235	286	521
17.00–17.99	236	303	539
18.00–18.99	195	310	505
19.00–19.99	180	329	509
20.00–20.99	170	331	501
21.00–21.99	193	353	546
22.00–22.99	166	335	501
23.00–23.99	187	346	533
24.00–24.99	172	334	506
Total	4579	5678	10,257

dependent variable (Y) and eight independent variables (S_x , $x = 1, 2, \dots, 8$) or one independent variable (S_8 , the numeric stage of the third molar) in the training set. In the test set, we test the classification performance of the logistic regression models to determine an individual was younger than legal age threshold ($Y = 0$) or equal to the legal age threshold or older ($Y = 1$).

Machine classification

To verify whether the deep learning model can replace humans to automatically extract feature information for age classification on OPGs without manually defined staging system interference, this paper built an end-to-end classification network. By using a composite model scaling algorithm to optimize the width, depth, and resolution of the network, we selected the EfficientNet network [25] and ensured the classification accuracy. Simultaneously, to avoid accidental results, we conducted another set of controlled experiments. By adding SE blocks to the original ResNet 101 [26] network to construct the SE-ResNet 101 network, the correlation between feature channels is modelled, and important features are enhanced to ensure the classification ability of the classifier. The framework of this network is shown in Fig. 1. The EfficientNet network is a compound model scaling algorithm, which helps our model to achieve a significant improvement

in the accuracy by comprehensively optimizing the network width, network depth, and resolution. At the same time, to ensure the inevitability of the final result, we also adopted SE-ResNet 101, which embeds the squeeze and excitation block (SE-block) into ResNet, and more nonlinear operations are added so that the model can better fit the complex correlation between channels. In addition, at the optimization algorithm level, the model uses the SGD optimization algorithm and only selects one sample at a time to update the model parameters, which promotes the faster convergence of our fusion network model. At the same time, to set a better learning rate, promote the model to converge, and avoid an “explosion” of the loss value of the objective function, we used an exponentially slowed learning rate optimization strategy, and the learning rate was decremented by the exponential difference every 30 rounds. This approach guarantees the results of learning. We also carried out data enhancement processing on these training data, such as random cropping and random horizontal flipping, which can urge the model to more fully learn the feature information of the OPGs. Figure 1 shows the whole framework of the legal age threshold classification.

Statistical analyses

All statistical analyses were performed using the statistical software IBM SPSS 18.0 (IBM® SPSS® Statistics, Armonk, NY). The chronological age of each subject was calculated by subtracting the date of birth from the date of the radiograph. Weighted Cohen’s kappa was used to evaluate the intra-rater and inter-rater agreement of the developmental stages of each tooth using the Demirjian method. Two experienced dentists assessed 100 randomly selected OPGs and re-examined them after an interval of 2 weeks. The age of each subject was blinded to the experts obtaining the estimation.

A two-by-two contingency table was used to evaluate the performance of each method to classify whether an individual was under the legal age threshold (14, 16, or 18 years) or not. The accuracy (the proportion of cases that are correctly classified), the sensitivity (the proportion of individuals who are truly under the age threshold who are classified as being under the threshold), and the specificity (the proportion of individuals who are truly equal to or over the age threshold who are classified as being equal to or over the age threshold) of the test were evaluated.

Results

Manual classification

For the reproducibility tests, the weighted Cohen kappa for the average intra-rater agreement was 0.921, whereas the average inter-rater agreement was 0.834.

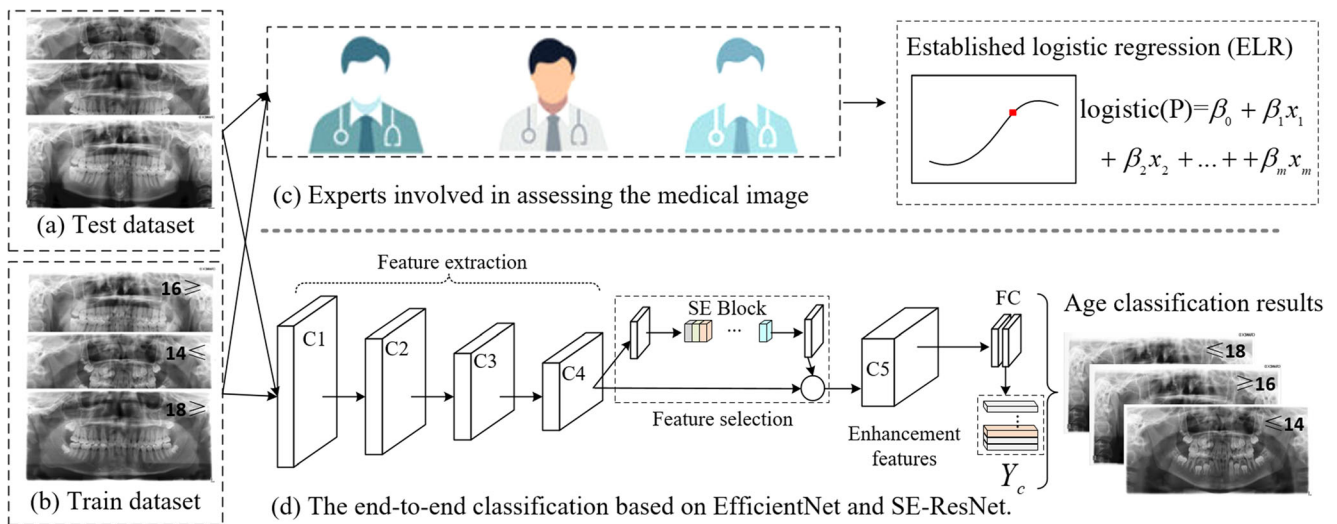


Fig. 1 The framework of legal age threshold classification. **a** Test dataset. **b** Training dataset. **c** Experts involved in assessing the medical image. **d** The end-to-end classification based on EfficientNet and SE-ResNet. C1–C5, convolution layers 1–5; SE, squeeze and excitation; FC, fully connected layer

By comparing the legal age threshold classifications of 14, 16, and 18 years, the results of this study indicate that the accuracy of the manual method is stable. As shown in Table 2, among the age thresholds at ages of 14, 16, and 18 years, the differences between the accuracy (ACC) are relatively small, with differences mostly lower than 2%. Most of Se are above 90% except for the Se of age threshold of 18 years old. When considering the age thresholds of 14 and 16 years, both the ACC and Sp are higher in the 31–38 (left mandibular eight teeth in the FDI numbering system) models, while the Se is higher in the 38 (left mandibular third molar in the FDI numbering system) models. However, at the age threshold of 18 years, the ACC, Se, and Sp are similar in both

regression models except the Se, with 94.9% for 31–38 model and 87.7% for 38 model.

Regardless of stable accuracy, the manual method is still unsatisfactory, which is attributed to the subjectivity and reproducibility of the technique as well as measurement bias. Besides, the traditional manual age classification method is labour-intensive and time-consuming.

Machine classification

To confirm whether the procedure and feature determined by human are probable for age classification, we developed the end-to-end CNN models without human interference. The

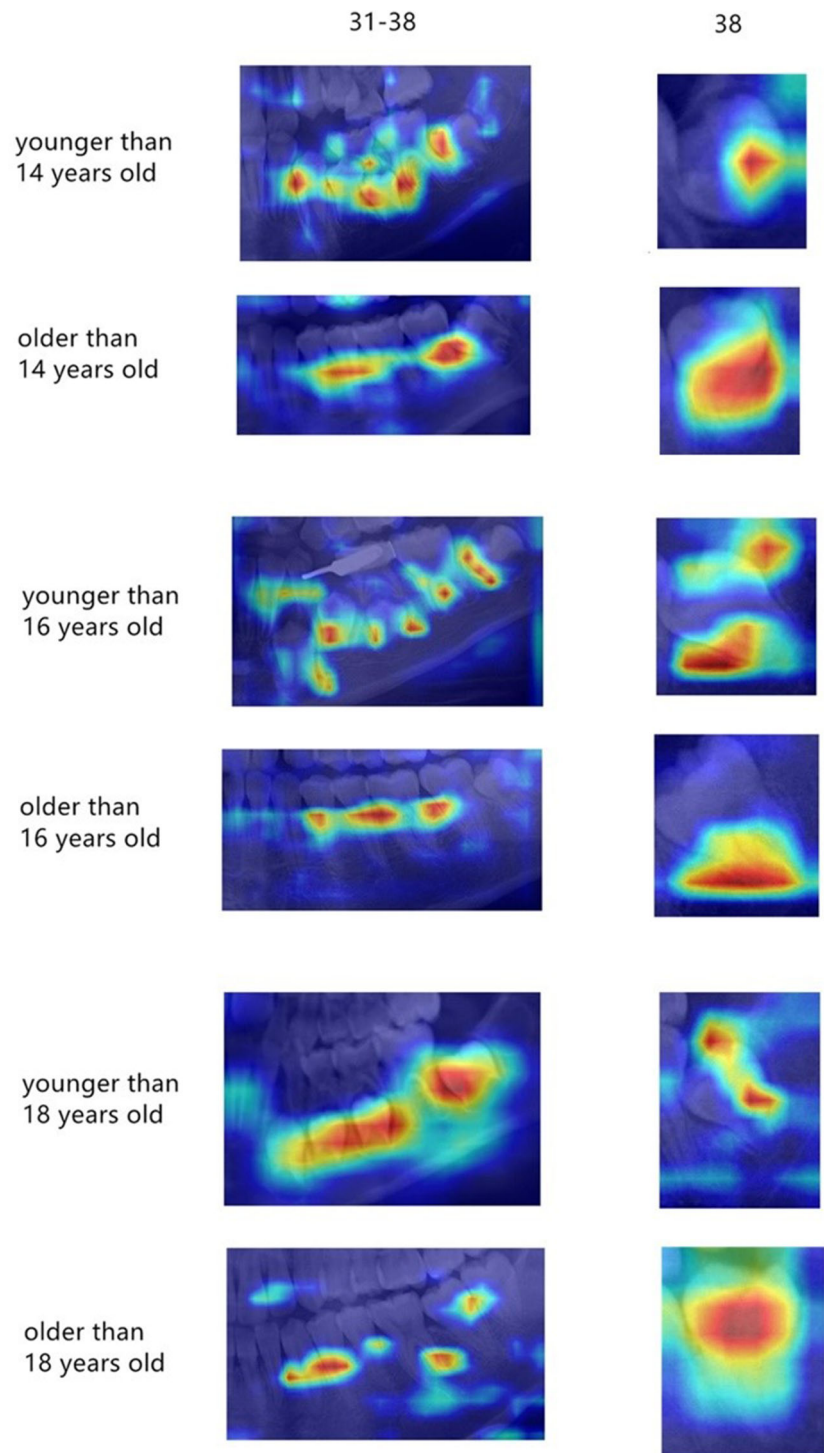
Table 2 The classification performance of the manual method and machine method based on 31–38 or 38

Threshold age	Methods	Accuracy		Sensitive		Specificity		
		31–38	38	31–38	38	31–38	38	
14	Manual	92.5%	90.0%	91.5%	96.2%	93.3%	85.3%	
	Machine	EfficientNet-B5	92.0%	92.0%	96.0%	93.3%	88.8%	91.5%
		EfficientNet-B3	95.9%	92.2%	95.4%	88.2%	96.3%	93.9%
		EfficientNet-B1	95.4%	93.4%	93.4%	82.4%	97.0%	98.2%
	Se-ResNet101	95.6%	93.3%	95.6%	89.5%	95.6%	95.0%	
16	Manual	91.3%	91.5%	96.0%	95.8%	85.9%	86.6%	
	Machine	EfficientNet-B5	90.9%	94.2%	97.5%	94.9%	82.9%	93.6%
		EfficientNet-B3	95.4%	95.7%	94.8%	96.4%	96.1%	95.2%
		EfficientNet-B1	95.1%	93.3%	94.8%	94.6%	95.5%	92.3%
	Se-ResNet101	95.0%	94.2%	94.6%	94.9%	95.5%	93.6%	
18	Manual	91.8%	91.7%	94.9%	87.7%	95.5%	95.5%	
	Machine	EfficientNet-B5	91.2%	90.5%	94.2%	94.6%	86.3%	86.6%
		EfficientNet-B3	92.3%	92.6%	92.6%	91.1%	91.9%	94.4%
		EfficientNet-B1	91.2%	93.3%	90.3%	92.7%	92.7%	94.1%
	Se-ResNet101	92.2%	90.9%	92.7%	89.2%	91.3%	93.0%	

end-to-end CNN models of automated age classification took 0.003 s per image on average. From Table 2, the performance of the end-to-end CNN models is better than that of the manual method, with the ACC, Se, and Sp increased except for the Se of the age threshold of 14 years old for 31–38 model and 16 years old for 38 model as well as Sp of the age threshold of 18 years old. The ACC of age threshold of 14 and 16 years reaches over 95%, and that of 18 years old reaches 93.3%.

To explore the reason of better classification performance for machine method, we drew the Grad-CAM. Through analysing the heatmap images of the test group (Fig. 2), we find that instead of focusing on the tooth morphology, which exhibits a high density in the X-ray images, the deep convolutional neural network (DCNN) pays more attention to the low density of the images, such as the dental pulp cavity, periodontal membrane, the area between adjacent

Fig. 2 The Grad-CAMs for different age thresholds based on 31–38 or 38



teeth, and the area between the deciduous teeth and permanent teeth. The results indicate that without the intervention of manual factors, machine learning can independently extract features in OPGs highly related to the age, and establish the complex comprehensive correlation between multiple features and chronological age.

Similar results are obtained in manual and machine methods (Table 2). It is found that the age classifications of different age thresholds need different scales of dental development information. The ACC, Se and Sp of the 31–38 models are mostly higher than 38 models for age classification of 14 and 16 years old, while the age classification of 18 years old gets opposite results. For lower age thresholds, most individuals have different teeth under development around the age threshold, so the classification based on more teeth can obtain higher accuracy. However, for a high age threshold, most individuals only have third molars under development, so too many teeth taking into account may result in an inverse interference, which decreases the accuracy. In addition, the low sensitivity of the 38 models among the two methods in the age threshold classification of 14 years may be caused by the slow development speed of third molars when children are approximately 14 years old. Since the third molar develops slowly, the images of the third molar in the OPGs are similar, which makes it difficult to distinguish an individual under or over 14 years old.

Discussion

By comparing the legal age threshold (14, 16, and 18 years old) classification based on the traditional manual method and deep learning model, we find that the deep learning model established for age threshold classification can replace or even surpass the manual method. It is indicated that although the

performance of manual method is stable, the end-to-end deep learning technique without human interference can effectively overcome the limitation of the manual method and greatly improve the classification performance. Moreover, for the Grad-CAM method, we find that instead of using dental morphological traits that are easy for a human to identify, the machine method focuses on the low-density features around the teeth, which may be the reason for the better performance of the end-to-end deep learning model.

The stability of the manual method may contribute to the ability of humans to avoid the feature interaction problem and noise interference based on their professional knowledge and previous experience to distinguish normal images from noisy images. There are various features in the OPGs correlated to age, but some of them do not change regularly with age and are not easy for humans to identify. As shown in Fig. 3, the distribution of dental development stages is different in different age groups, and the older the subject is, the more teeth finish development (stage 8), which indicates that the age classification of the manual method which only focuses on dental development is highly related to age, which makes the classification performance stable. However, there are still several limitations in age classification. First, traditional manual methods are likely to incorporate a relatively higher degree of intra- and inter-observer errors due to the subjectivity in the stage evaluation, which can lead to an increase in the prediction error [27]. Second, the relationship between dental development and age is a nonlinear function which is hard to specify, so we can only use the linear regression model to establish a local approximation. Third, the development of teeth is a continuous process. Demirjian et al. [24] divided tooth development into eight discontinuous stages, which cannot perfectly fit the complex relationship between dental development and

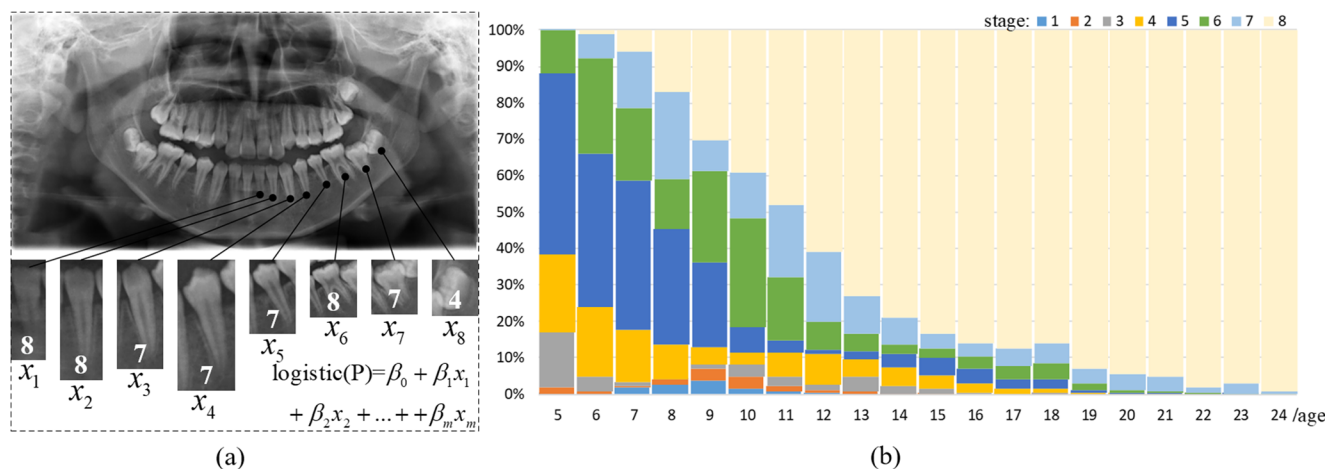


Fig. 3 The verification of the Demirjian method for age classification. **a** Estimating the dental development stages of eight left mandibular permanent teeth for age classification in OPGs. **b** The distribution of different dental stages of all eight teeth for 5–24 years old groups

age. Moreover, manual methods are labour-intensive and time-consuming, which makes them inconvenient to apply in routine clinical activity [28, 29].

In recent years, deep learning techniques have made breakthroughs in the fields of computer vision [30, 31], speech recognition, natural language processing, and bioinformatics [32, 33] achieved good results in many fields [34, 35]. Previous studies on age classification mainly focused on either the performance of traditional manual methods or deep learning models. This is the first study to verify whether the deep learning model can replace humans on dental age classification based on a large sample. We used an end-to-end deep learning technique to determine whether it can surpass the performance of manual method for age classification. By using this method, we conducted related experiments and concluded that the classification CNN methods can obtain the best prediction results in age threshold classification. Without the intervention of manual factors, machine learning can analyse the correlation between the age and the features of OPGs and establish the complex comprehensive correlation between multiple features and age. In addition, this learning model is motivated by data, which means that when the data are large enough, high-precision identification can be realized.

The Grad-CAM method can provide a heatmap of the adequately visualized interest regions of the deep learning models that contribute the most to increasing or decreasing the output. In the medical field, some studies [36–39] used the Grad-CAM method to identify the region of interest and validate the DCNN performance. In our study, we found that the DCNN pays more attention to the low density of the images, including the dental pulp cavity, periodontal membrane, the area between adjacent teeth, and the area between the deciduous teeth and permanent teeth. Since the dental pulp cavity experiences age-related pathological and physiological changes [40], Cameriere et al. [41] and Kvaal et al. [42] both established age estimation methods based on the age-related changes in the dental pulp cavity, which have been verified in some studies [18, 43, 44]. For the periodontal membrane, several methods have been proposed by Olze et al. [12, 45, 46] to classify an individual below or above the 18-year threshold, especially once root growth is completed. In our previous study [47], we also estimated the dental age based on the visibility of the periodontal ligament of the third molar in the northern Chinese population. Both the area between the adjacent teeth and the area between the deciduous teeth and permanent teeth change with age during the replacement of the permanent teeth, which can be supplementary features for dental age estimation. Human beings can only clearly identify the features with high density in X-ray images, while a machine can synthetically assess each feature in the images and select highly related regions for classification, combine these features, and fit the complicated relation of these features, which helps to increase the accuracy of the age classification.

Although our study established a satisfactory DCNN model for age classification performed better than traditional manual method, there are still two limitations. Firstly, this study only focuses on age classification, which is the foundation of age estimation, while whether the DCNN model is suitable for age estimation has not been confirmed. Secondly, the dental age estimation methods perform differently among various ethnic populations, but we only established the DCNN model and compared with manual method in a Chinese population. It is still not sure whether similar results can be obtained from other different populations.

In conclusion, comparing the age classification performance of the traditional manual method and the machine method, the end-to-end deep learning without human interference can surpass the performance of the manual method, and the features extracted in the CNN model are different from those defined by a human. This approach holds great promise for future forensic practice.

Funding This work was supported by the National Key Research and Development Program of China under Grant No. 2017YFA0700800, the National Natural Science Foundation of China under Grant Nos. 81701869, 61971343 and 61627811, and the Key Research and Development Program of Shaanxi Province of China under Grant No. 2020GXLH-Y-008.

Declarations Ethical approval was granted by the ethics committee of Stomatological Hospital of Xi'an Jiaotong University.

Conflict of interest The authors declare no competing interests.

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study, formal consent is not required.

Consent to participate For this type of study, formal consent is not required.

References

- Schmeling A, Olze A, Reisinger W, Geserick G (2001) Age estimation of living people undergoing criminal proceedings. *Lancet* 358:89–90. [https://doi.org/10.1016/S0140-6736\(01\)05379-X](https://doi.org/10.1016/S0140-6736(01)05379-X)
- Hagen M, Schmidt S, Schulz R, Vieth V, Ottow C, Olze A, Pfeiffer H, Schmeling A (2020) Forensic age assessment of living adolescents and young adults at the Institute of Legal Medicine, Munster, from 2009 to 2018. *Int J Legal Med* 134:745–751. <https://doi.org/10.1007/s00414-019-02239-2>
- Thevissen PW, Kvaal SI, Willems G (2012) Ethics in age estimation of unaccompanied minors. *J Forensic Odontostomatol* 30(Suppl 1):84–102
- Befurt L, Kirchhoff G, Rudolf E, Schmeling A (2020) Legal aspects of forensic age diagnostics on the basis of 42f German Social Code

- (SGB) VIII. Rechtsmedizin 30:241–248. <https://doi.org/10.1007/s00194-020-00392-2>
5. Olze A, Peschke C, Schulz R, Schmeling A (2008) Studies of the chronological course of wisdom tooth eruption in a German population. *J Forensic Leg Med* 15:426–429. <https://doi.org/10.1016/j.jflm.2008.02.008>
 6. Bassed RB, Briggs C, Drummer OH (2011) Age estimation and the developing third molar tooth: an analysis of an Australian population using computed tomography. *J Forensic Sci* 56:1185–1191. <https://doi.org/10.1111/j.1556-4029.2011.01769.x>
 7. Kasper KA, Austin D, Kvanli AH, Rios TR, Senn DR (2009) Reliability of third molar development for age estimation in a Texas Hispanic population: a comparison study. *J Forensic Sci* 54:651–657. <https://doi.org/10.1111/j.1556-4029.2009.01031.x>
 8. Lengfu Z (2006) On minors' civil liability ability. *J Shangluo Teach Coll* 20:59–63
 9. Jun L (2005) Investigation on disputed issues about criminal responsibility of criminal minor in China. *J Liaoning Univ (Phil Soc Sci)* 33:34–39. <https://doi.org/10.1016/j.forsciint.2006.02.019>
 10. Mincer HH, Harris EF, Berryman HE (1993) The A.B.F.O. study of third molar development and its use as an estimator of chronological age. *J Forensic Sci* 38:379–390
 11. Cameriere R, Ferrante L, De Angelis D, Scarpino F, Galli F (2008) The comparison between measurement of open apices of third molars and Demirjian stages to test chronological age of over 18 year olds in living subjects. *Int J Legal Med* 122:493–497. <https://doi.org/10.1007/s00414-008-0279-6>
 12. Olze A, Solheim T, Schulz R, Kupfer M, Pfeiffer H, Schmeling A (2010) Assessment of the radiographic visibility of the periodontal ligament in the lower third molars for the purpose of forensic age estimation in living individuals. *Int J Legal Med* 124:445–448. <https://doi.org/10.1007/s00414-010-0488-7>
 13. Cameriere R, Brkic H, Ermenc B, Ferrante L, Ovsenik M, Cingolani M (2008) The measurement of open apices of teeth to test chronological age of over 14-year olds in living subjects. *Forensic Sci Int* 174:217–221. <https://doi.org/10.1016/j.forsciint.2007.04.220>
 14. Cameriere R, Velandia Palacio LA, Pinares J, Bestetti F, Paba R, Coccia E, Ferrante L (2018) Assessment of second (I2M) and third (I3M) molar indices for establishing 14 and 16 legal ages and validation of the Cameriere's I3M cut-off for 18 years old in Chilean population. *Forensic Sci Int* 285(205):e1–e5. <https://doi.org/10.1016/j.forsciint.2017.12.043>
 15. Lee JH, Han SS, Kim YH, Lee C, Kim I (2019) Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 129:635–642. <https://doi.org/10.1016/j.oooo.2019.11.007>
 16. Banar N, Bertels J, Laurent F, Boedi RM, de Tobel J, Thevissen P, Vandermeulen D (2020) Towards fully automated third molar development staging in panoramic radiographs. *Int J Legal Med* 134: 1831–1841. <https://doi.org/10.1007/s00414-020-02283-3>
 17. Merdjetio Boedi R, Banar N, De Tobel J, Bertels J, Vandermeulen D, Thevissen PW (2020) Effect of lower third molar segmentations on automated tooth development staging using a convolutional neural network. *J Forensic Sci* 65:481–486. <https://doi.org/10.1111/1556-4029.14182>
 18. Zheng Q, Ge Z, Du H, Li G (2021) Age estimation based on 3D pulp chamber segmentation of first molars from cone-beam-computed tomography by integrated deep learning and level set. *Int J Legal Med* 135:365–373. <https://doi.org/10.1007/s00414-020-02459-x>
 19. Stern D, Payer C, Giuliani N, Urschler M (2019) Automatic age estimation and majority age classification from multi-factorial MRI data. *IEEE J Biomed Health Inform* 23:1392–1403. <https://doi.org/10.1109/JBHI.2018.2869606>
 20. Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, Schwendicke F (2019) Deep learning for the radiographic detection of apical Lesions. *J Endod* 45:917–22 e5. <https://doi.org/10.1016/j.joen.2019.03.016>
 21. Lee JH, Kim DH, Jeong SN, Choi SH (2018) Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontal Implant Sci* 48:114–123. <https://doi.org/10.5051/jpis.2018.48.2.114>
 22. Chu P, Bo C, Liang X et al (2018) Using octuplet siamese network for osteoporosis analysis on dental panoramic radiographs. *Conf Proc IEEE Eng Med Biol Soc* 2018:2579–2582. <https://doi.org/10.1109/EMBC.2018.8512755>
 23. Galibourg A, Cussat-Blanc S, Dumoncel J, Telmon N, Monsarrat P, Maret D (2021) Comparison of different machine learning approaches to predict dental age using Demirjian's staging approach. *Int J Legal Med*. 135:665–675. <https://doi.org/10.1007/s00414-020-02489-5>
 24. Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. *Hum Biol* 45:211–227
 25. Tan M, Le QV (2019) Efficientnet: rethinking model scaling for convolutional neural networks. arXiv:1905.11946
 26. Hu J, Shen L, Sun G (2020) Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42(8):2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
 27. Franklin D, Karkhanis S, Flavel A, Collini F, DeLuca S, Cameriere R (2016) Accuracy of a cut-off value based on the third molar index: Validation in an Australian population. *Forensic Sci Int* 266(575):e1–e6. <https://doi.org/10.1016/j.forsciint.2016.06.032>
 28. Galic I, Vodanovic M, Cameriere R et al (2011) Accuracy of Cameriere, Haavikko, and Willems radiographic methods on age estimation on Bosnian-Herzegovian children age groups 6–13. *Int J Legal Med* 125:315–321. <https://doi.org/10.1007/s00414-010-0515-8>
 29. Melo M, Ata-Ali J (2017) Accuracy of the estimation of dental age in comparison with chronological age in a Spanish sample of 2641 living subjects using the Demirjian and Nolla methods. *Forensic Sci Int* 270(276):e1–e7. <https://doi.org/10.1016/j.forsciint.2016.10.001>
 30. Athanasiadou E, Geradts Z, Van Eijk E (2018) Camera recognition with deep learning. *Forensic Sci Res* 3:210–218. <https://doi.org/10.1080/20961790.2018.1485198>
 31. Bedeli M, Geradts Z, van Eijk E (2018) Clothing identification via deep learning: forensic applications. *Forensic Sci Res* 3:219–229. <https://doi.org/10.1080/20961790.2018.1526251>
 32. Yin W, Kann K, Yu M, Schutze H (2017) Comparative study of CNN and RNN for natural language processing. arXiv:1702.01923
 33. Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X (2019) Deep learning in bioinformatics: introduction, application, and perspective in the big data era. *Methods* 166:4–21. <https://doi.org/10.1016/j.ymeth.2019.04.008>
 34. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, Real-time object detection. *Conf Proc IEEE Comput Vis Pattern Recongnit (CVPR)* 2016:779–788. <https://doi.org/10.1109/CVPR.2016.91>
 35. Liu W, Anguelov D, Erhan D, Szegedy C, Berg AC (2016) SSD: single shot multibox detector. *Conf Proc Springer Cham European Conf on Comput Vis (ECCV)* 2016:21–37. https://doi.org/10.1007/978-3-319-46448-0_2
 36. Cheng CT, Ho TY, Lee TY, Chang CC, Chou CC, Chen CC, Chung IF, Liao CH (2019) Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol* 29:5469–5477. <https://doi.org/10.1007/s00330-019-06167-y>
 37. He T, Guo J, Chen N, Xu X, Wang Z, Fu K, Liu L, Yi Z (2020) MediMLP: using Grad-CAM to extract crucial variables for lung cancer postoperative complication prediction. *IEEE J Biomed*

- Health Inform 24:1762–1771. <https://doi.org/10.1109/JBHI.2019.2949601>
38. Iizuka T, Fukasawa M, Kameyama M (2019) Deep-learning-based imaging-classification identified cingulate island sign in dementia with Lewy bodies. *Sci Rep* 9:8944. <https://doi.org/10.1038/s41598-019-45415-5>
 39. Jonas S, Rossetti AO, Oddo M, Jenni S, Favaro P, Zubler F (2019) EEG-based outcome prediction after cardiac arrest with convolutional neural networks: performance and visualization of discriminative features. *Hum Brain Mapp* 40:4606–4617. <https://doi.org/10.1002/hbm.24724>
 40. Biuki N, Razi T, Faramarzi M (2017) Relationship between pulp-tooth volume ratios and chronological age in different anterior teeth on CBCT. *J Clin Exp Dent* 9:e688–ee93. <https://doi.org/10.4317/jced.53654>
 41. Cameriere R, Ferrante L, Belcastro MG, Bonfiglioli B, Rastelli E, Cingolani M (2007) Age estimation by pulp/tooth ratio in canines by peri-apical X-rays. *J Forensic Sci* 52:166–170. <https://doi.org/10.1111/j.1556-4029.2006.00336.x>
 42. Kvaal SI, Kolltveit KM, Thomsen IO, Solheim T (1995) Age estimation of adults from dental radiographs. *Forensic Sci Int* 74:175–185. [https://doi.org/10.1016/0379-0738\(95\)01760-g](https://doi.org/10.1016/0379-0738(95)01760-g)
 43. Timme M, Borkert J, Nagelmann N, Schmeling A (2020) Evaluation of secondary dentin formation for forensic age assessment by means of semi-automatic segmented ultrahigh field 9.4 T UTE MRI datasets. *Int J Legal Med* 134:2283–2288. <https://doi.org/10.1007/s00414-020-02425-7>
 44. Molina A, Bravo M, Fonseca GM, Marquez-Grant N, Martin-de-Las-Heras S (2021) Dental age estimation based on pulp chamber/crown volume ratio measured on CBCT images in a Spanish population. *Int J Legal Med* 135:359–364. <https://doi.org/10.1007/s00414-020-02377-y>
 45. Olze A, Solheim T, Schulz R, Kupfer M, Schmeling A (2010) Evaluation of the radiographic visibility of the root pulp in the lower third molars for the purpose of forensic age estimation in living individuals. *Int J Legal Med* 124:183–186. <https://doi.org/10.1007/s00414-009-0415-y>
 46. Olze A, Hertel J, Schulz R, Wierer T, Schmeling A (2012) Radiographic evaluation of Gustafson's criteria for the purpose of forensic age diagnostics. *Int J Legal Med* 126:615–621. <https://doi.org/10.1007/s00414-012-0701-y>
 47. Guo YC, Li MJ, Olze A, Schmidt S, Schulz R, Zhou H, Pfeiffer H, Chen T, Schmeling A (2018) Studies on the radiographic visibility of the periodontal ligament in lower third molars: can the Olze method be used in the Chinese population? *Int J Legal Med* 132:617–622. <https://doi.org/10.1007/s00414-017-1664-9>
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.