Journal of Dental Sciences 20 (2025) 477–486

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jds.com

Original Article

Establishment and evaluation of a deep learning-based tooth wear severity grading system using intraoral photographs

Ya-Ning Pang [a†], Zhen Yang [b†], Ling-Xiao Zhang [a], Xiao-qiang Liu [b], Xin-Shu Dong [b], Xun Sheng [c], Jian-guo Tan [b], Xin-Yu Mao [d*], Ming-yue Liu [e**]

[a] *Institute of Applied Electronics, School of Electronics, Peking University, Beijing, China*
[b] *Department of Prosthodontics, Peking University School and Hospital of Stomatology & National Center for Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Research Center of Oral Biomaterials and Digital Medical Devices & Beijing Key Laboratory of Digital Stomatology & NHC Key Laboratory of Digital Stomatology & NMPA Key Laboratory for Dental Materials, Beijing, China*
[c] *Department of General Dentistry, Kunming Medical University School and Hospital of Stomatology & Yunnan Key Laboratory of Stomatology, Kunming, China*
[d] *School of Electronics Engineering and Computer Science, Peking University, Beijing, China*
[e] *First Clinical Division, Peking University School and Hospital of Stomatology & National Center for Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Research Center of Oral Biomaterials and Digital Medical Devices & Beijing Key Laboratory of Digital Stomatology & NHC Key Laboratory of Digital Stomatology & NMPA Key Laboratory for Dental Materials, Beijing, China*

Received 29 April 2024; Final revision received 11 May 2024
Available online 21 May 2024

KEYWORDS
Artificial intelligence;
Deep learning;
Diagnosis;
Tooth wear

Abstract *Background/purpose:* Artificial intelligence (AI) can assist in medical diagnosis owing to its high accuracy and efficiency. This study aimed to develop a diagnostic system for automatically determining the degree of tooth wear (TW) using intraoral photographs with deep learning.
Materials and methods: The study included 388 intraoral photographs. A tooth segmentation model was first established using the Mask R-CNN architecture, which incorporated U-Net and SGE attention mechanisms. Subsequently, 2774 individual tooth images output from the

* Corresponding author. School of Electronics Engineering and Computer Science, Peking University, No.5 Yiheyuan Road, Haidian District, Beijing, 100871, China.
** Corresponding author. First Clinical Division, Peking University School and Hospital of Stomatology, No.A37 XiShiKu Street, XiCheng District, Beijing, 100034, China.
E-mail addresses: xymao@pku.edu.cn (X.-Y. Mao), doctorlmy0616@yeah.net (M.-y. Liu).
† These authors contributed equally to this work.

https://doi.org/10.1016/j.jds.2024.05.013

segmentation model were included into the classification task, labeled and randomized into training, validation, and test sets with 6.0:2.0:2.0 ratio. A vision transformer model optimized using a mask mechanism was constructed for TW degree classification. The models were evaluated using the accuracy, precision, recall, and F1-score metrics. The time required for AI analysis was calculated.

Results: The accuracy of the tooth segmentation model was 0.95. The average accuracy, precision, recall, and F1-score in the classification task were 0.93, 0.91, 0.88, and 0.89, respectively. The F1-score differed in different grades (0.97 for grade 0, 0.90 for grade 1, 0.88 for grade 2, and 0.82 for grade 3). No significant difference was observed in the accuracy between different surfaces. The AI system reduced the time required to grade an individual tooth surface to 0.07 s, compared to the 2.67 s required by clinicians.

Conclusion: The developed system provides superior accuracy and efficiency in determining TW degree using intraoral photographs. It might assist clinicians in the decision-making for TW treatment and help patients perform self-assessments and disease follow-ups.

Introduction

As a common clinical non-caries dental hard tissue disease, tooth wear (TW) refers to the progressive loss of hard tissues of the teeth owing to mechanical friction or chemical acid erosion.[1] Severe TW can cause damage to the teeth, musculature of the oral and maxillofacial system, and temporomandibular joint, manifested as various complications such as dentin sensitivity, pulpitis, tooth fracture, occlusal disorders, masticatory dysfunction, and temporomandibular joint disorders.[2] The treatment of severely worn dentition involves multiple disciplines and requires the comprehensive application of sophisticated theories (e.g., occlusion and aesthetics) and restoration techniques, making it one of the biggest challenges in restorative dentistry. With increased population aging and changes in dietary habits, TW has been increasing, showing a trend toward affecting younger individuals.[3] Schlueter et al. suggested that the prevalence of erosive wear in deciduous and permanent teeth was 30%—50% and 20%—45%, respectively.[4] Therefore, research on TW has become increasingly significant in today's era, when increasing awareness and importance are attached to oral diseases.

The accurate diagnosis of TW severity is crucial for the development of treatment plans. Currently, the clinical assessment of TW severity is mainly conducted using semiquantitative index grading methods. Commonly used TW indices include the Tooth Wear Index (TWI),[5] Basic Erosive Wear Examination (BEWE),[6] and Tooth Wear Evaluation System (TWES).[7,8] These indices describe the extent and manifestation of hard tissue loss in teeth, including the loss of enamel morphological features, level of dentin exposure, loss of clinical crown height, and whether the pulp is exposed. Among them, TWI and its modified versions are the most widely used and representative index grading methods for measuring and monitoring multifactorial TW. The index is simple to use clinically with clear evaluation criteria and high reliability; additionally, it is suitable for both epidemiological surveys and the evaluation of individual patients' TW.[9] However, without computer assistance, using the index fully as a research tool is not feasible owing to the time required and the amount of data generated when applied to the entire dentition.

Traditionally, clinicians make a comprehensive judgment of TW severity through direct visual inspection based on intraoral inspection or observation from models/photographs.[10] This has the advantages of convenience and speed, making it particularly suitable for observing masses of patients in the clinic. However, it has the following defects: (1) The accuracy of the diagnosis relies on the clinician's experience, resulting in large subjective bias, and there are considerable differences in diagnostic efficiency and repeatability among different clinicians.[11] (2) For cases in the early stages of the disease where wear is difficult to distinguish with the naked eye, the likelihood of misdiagnosis and missed diagnosis is high. (3) Owing to operational and visual field limitations of the direct intraoral visual inspection, it is difficult to finely evaluate the wear details of tooth surface; additionally, it is difficult to comprehensively evaluate every tooth surface. (4) It is very time-consuming to comprehensively evaluate the entire dentition by counting each tooth surface of all the teeth.

With the advent of the big data era, artificial intelligence (AI), specifically deep learning (DL) technology, has achieved rapid development in the field of medical diagnosis.[12] Applying DL technology to medical diagnosis benefits from its advantages in image recognition. By constructing artificial neural networks, automatic analysis of large quantities of labeled medical image data is performed, and the constructed model is used to predict the target data and thus assist in disease diagnosis.[13] Currently, DL technology has made significant progress in assisting the diagnosis of various clinical diseases, such as cardiovascular diseases, tumors, retinal lesions, and Alzheimer's disease.[14] Compared with human intelligence, AI systems are good at extracting features from images that are imperceptible or unquantifiable to the human eye, thereby accurately and efficiently providing objective diagnostic opinions for clinicians. Moreover, AI systems can conduct a large amount of repetitive work with low replication costs,

thus serving as repeatable medical resources. Therefore, AI has significant prospects for clinical applications in the medical field.

With the vigorous development of AI technology in the field of clinical medicine, DL is gradually being used in various areas of dentistry; however, it is still in its infancy.[15] Its current applications in the diagnosis of oral diseases are as follows: in the field of operative dentistry and endodontics, AI can help to diagnose caries,[16] periapical lesions,[17] root fracture,[18] and identify C-type root canals[19] based on the data from apical radiographs, panoramic images, cone beam computed tomography (CBCT), near-infrared transillumination images, and digital photographs. In the field of periodontics, AI can identify dental plaque based on intraoral photos,[20] determine the degree of alveolar bone loss based on radiological examinations,[21] and diagnose periodontal diseases based on clinical information, X-rays, immune responses, and microbial composition.[22] In the field of oral and maxillofacial surgery, AI can help to detect cysts and tumors based on two-dimensional or three-dimensional radiology findings and cytopathological images.[23,24] In the field of orthodontics, AI can conduct automated cephalometric analysis on lateral cephalometric radiographs[25] and CBCT images,[26] evaluate cervical vertebral maturation,[27] and perform craniofacial skeletal classification[28] based on lateral cephalometry, thus assisting in the decision-making of treatment plans. However, the use of DL techniques to assist in the diagnosis of TW has not been reported.

This study aimed to construct a DL-based TW diagnostic model based on intraoral photographs that could intelligently grade the severity of TW. To achieve this goal, we collected intraoral photographs of patients, annotated the TW gradings, constructed a diagnostic model based on convolutional neural network (CNN) and vision transformer (ViT) algorithms, and evaluated its efficacy. The diagnostic model consists of two subtasks: tooth segmentation and TW severity classification. This system could help clinicians accurately and efficiently determine the extent of TW, design treatment plans, and assist patients with self-assessment and disease follow-ups.

Materials and methods

This study was approved by the Biomedical Ethics Committee of the Peking University School of Stomatology (Approval number: PKUSSIRB-202279122). All photographs were acquired and used with patient consent.

Data acquisition

Using plastic retractors and reflectors, intraoral photographs of the upper and lower permanent dental arches were taken by dental professionals using digital cameras (Nikon D810 or D610, Nikon Inc., Tokyo, Japan) with a sensitivity of ISO 100, an aperture value of F20-32, and a shutter speed of 1/200-1/100. Nine photographs were obtained for each patient (Fig. 1). The inclusion criteria for the photographs were as follows: (1) appropriate exposure without blurring, (2) photos taken at proper angles, and (3) absence of foreign objects obscuring or contaminating the tooth surface. Finally, 388 intraoral photographs from 61 patients who visited the Peking University School of

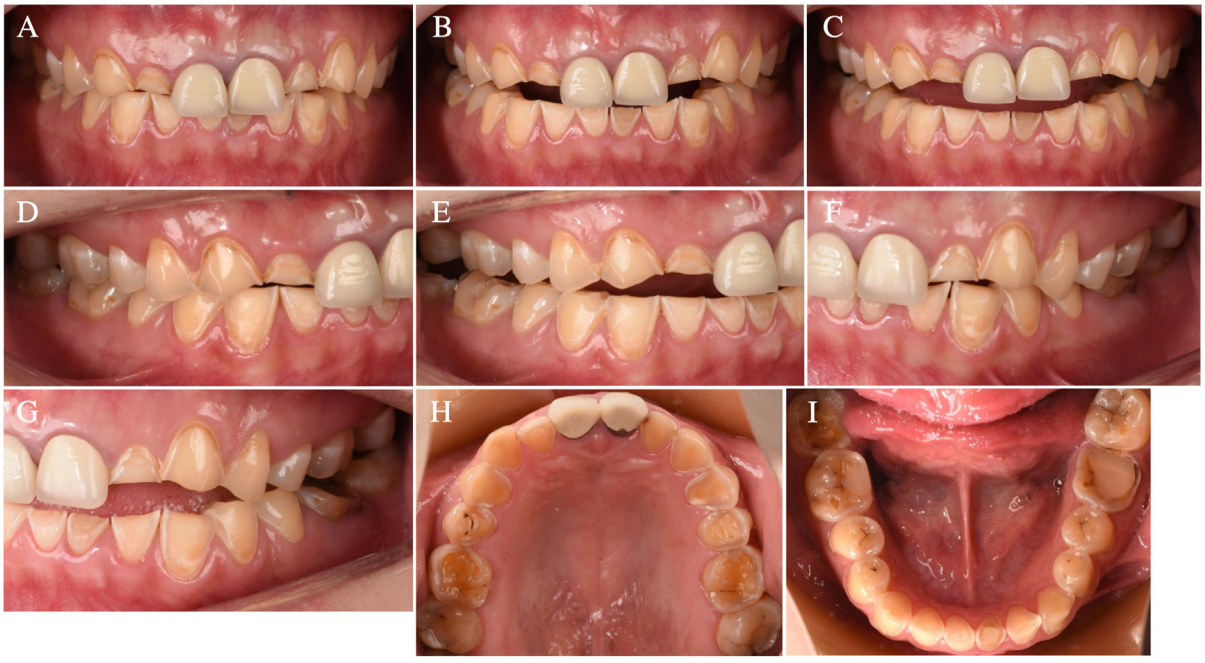

Figure 1 Clinical collection of nine recommended intraoral photographs. (A), (B), and (C) are the three frontal views with the teeth in maximum intercuspation, protrusive movements, and mouth opening; the anterior teeth of which were included in the classification model. (D), (E), (F), and (G) are the four lateral views with the teeth in maximum intercuspation and laterotrusive movements; the posterior teeth of which were included in the classification model. (H) and (I) are the maxillary and mandibular occlusal views.

Stomatology from January 2020 to February 2024 were included. The patients' ages ranged from 21 to 69 years. Regarding the TW degree classification, only segmented images of tooth surfaces without any carious, direct or indirect restorations were included to avoid potential evaluation bias.

Data annotation

In the tooth segmentation task, LabelMe software was used for data annotation. LabelMe is an image annotation software implemented in Python that uses Qt as its graphical interface, supporting the labeling and annotation of images and videos. In the TW degree classification task, each tooth surface was annotated with one of the four labels (modified-TWI[5]): (1) TWI 0, no wear into dentin; (2) TWI 1, dentin exposure less than 1/3 of the surface; (3) TWI 2, dentin exposure greater than 1/3 of the surface; (4) TWI 3, pulp or secondary dentin exposure. The labeled images were divided into training, validation, and test datasets at a ratio of 6.0:2.0:2.0 using the random number table method (all datasets contained the same proportion of images with different degrees of TW).

To ensure accuracy, each case was annotated by two attending prosthodontic clinicians, followed by a consistency check. After annotation labeling was completed, a random spot check was conducted by a chief clinician with 25 years of clinical experience to ensure the accuracy of the annotations for quality control.

Model and training

The TW severity grading system (TWGS) comprises two components: tooth segmentation and TW degree classification.

The tooth segmentation model was based on the mask region-based convolutional neural network (Mask R-CNN),[29] with the ResNet-50 in the backbone structure and the fully convolutional network (FCN) layers being improved and replaced (Fig. 2). An attention mechanism (spatial group-wise enhance, SGE) was introduced into ResNet, and the FCN structure was replaced with U-Net to improve the accuracy of the model in identifying fine structures and edges. The U-Net structure contains three downsampling layers and three upsampling layers, with a learning rate of 10^{-5}. The SGE self-attention mechanism was incorporated into the skip connection of the U-Net network to reinforce the effective features and improve the accuracy of the mask output. The number of splits in the SGE module was set to eight. The loss function of this model consists mainly of classification, regression box, and mask losses.

$$L = L_{cls} + L_{box} + L_{mask}$$

In the TW degree classification task, the TW severity of the individual surface of the target tooth was determined by classifying the output images from the tooth segmentation task. Considering the significant differences in features between different tooth surfaces, the classification task categorized the input single-tooth data into occlusal/incisal surface and non-occlusal/non-incisal surface groups. The TW degree of the individual tooth surface was then classified using the ViT[30] model that incorporates a mask mechanism (Fig. 3).

The two models constitute the intelligent TWGS, and its specific workflow is as follows. First, the images are input into the segmentation model, and the model outputs the segmented individual tooth images and the corresponding bounding box information. Thereafter, the individual tooth images are input into the classification model to obtain the TW degree classification results for individual tooth surfaces. Finally, by combining these classification results with the bounding box information, the prediction results for each tooth surface are aggregated back into the original input image to obtain the TW grading results for each tooth surface.

Model evaluation

The accuracy and efficiency of the TWGS were evaluated. In the tooth segmentation task, mean pixel accuracy (mAP)

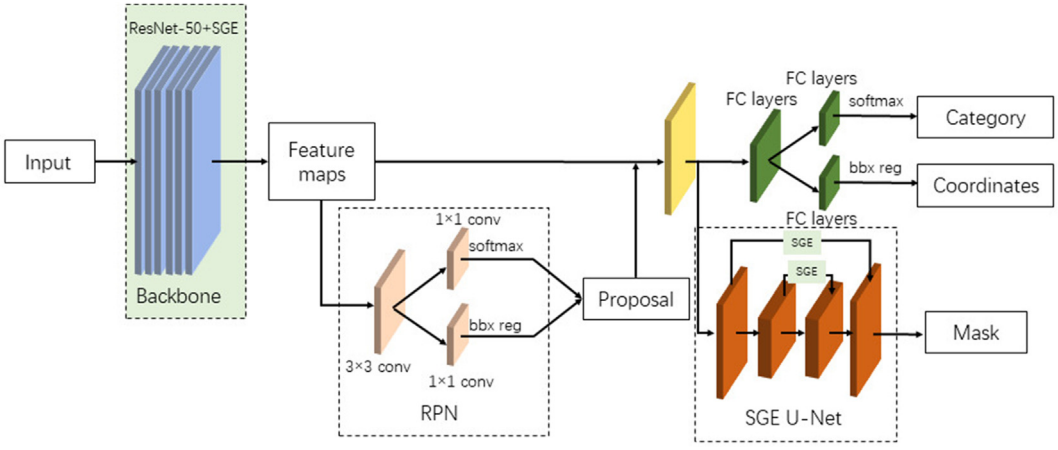

Figure 2 Overall architecture of the tooth segmentation model, which is based on the Mask R-CNN network and improved by the combination of an attention mechanism and the U-Net algorithm. Abbreviations: SGE, Spatial group-wise enhance; RPN, Region proposal network; FC layers, Fully connected layers; Mask R-CNN, Mask region-based convolutional neural network.

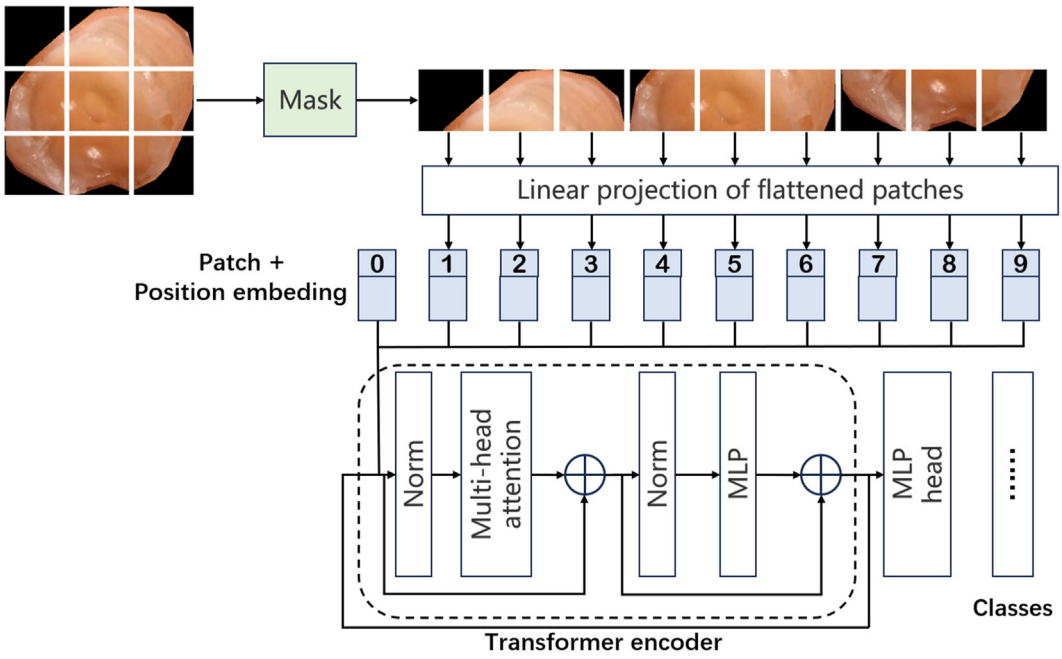

Figure 3 Overall architecture of the tooth wear degree classification model, which is based on the vision transformer model and improved with a mask mechanism. Abbreviation: MLP, Multilayer perceptron.

was selected to evaluate the accuracy of the model. In the TW degree classification task, precision, recall, F1-score, accuracy, macro-average, and weighted-average were used as the evaluation indicators for the accuracy of the constructed model. These indicators were calculated as follows: precision $= \frac{TP}{TP+FP}$, recall $= \frac{TP}{TP+FN}$, F1-score $= \frac{2TP}{2TP+FP+FN}$ (also known as the dice), and accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$. Here, TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. The accuracy and Cohen's Kappa coefficient of the binary AI model (defined by the presence or absence of dentin exposure) were calculated. The time required for the AI system to complete the segmentation and classification tasks was recorded and compared with the performance of clinicians. Statistical analysis was conducted using IBM SPSS Statistics version 26.0 (IBM Corp., Armonk, NY, USA).

Results

The tooth segmentation task involved 388 intraoral photographs satisfying the inclusion criteria. The segmentation output results of the different algorithms for the same image in the test set are illustrated in Fig. 4, and a comparison of the accuracies of the different algorithms is presented in Table 1. After replacing the FCN structure of the Mask R-CNN architecture with U-Net (improved by the incorporation of the SGE self-attention mechanism), the model performed better in the recognition of tooth edges and became more sensitive to the recognition of mask areas, achieving a better recognition of noise in the images because of lighting. After incorporating the attention mechanism module (SGE) into the ResNet of the backbone structure, the model achieved better recognition of fine tooth structures, leading to higher accuracy. The introduction of these two structures enhanced the segmentation accuracy (mPA) of the model to 0.95.

The classification task was performed based on the output from the tooth segmentation task, and the segmented individual tooth images were used as the training dataset. A total of 2774 individual tooth images were annotated and categorized into the occlusal/incisal (778, 28.05%) and non-occlusal/non-incisal (1996, 71.95%) groups before training. The distribution of TWI grades of the annotated images in the occlusal/incisal group was as follows: 108 annotated images with TWI grade 0 (13.88%), 216 images with TWI grade 1 (27.76%), 349 images with TWI grade 2 (44.86%), and 105 images with TWI grade 3 (13.50%). In the non-occlusal/non-incisal group, there were 1312 annotated teeth with TWI grade 0 (65.73%), 454 teeth with TWI grade 1 (22.75%), 197 teeth with TWI grade 2 (9.87%), and 33 teeth with TWI grade 3 (1.65%).

The classification output results are presented in Fig. 5. The performance indicators of the classification model on the test set (553 cases, 19.94% of all labeled data) for each TW grade on the different tooth surfaces are listed in Table 2. The trained diagnostic system achieved an accuracy of 0.93, a macro-averaged F1-score of 0.89, a recall of 0.88, and a precision of 0.91. The F1-score decreased from TWI grade 0 to 3, with values of 0.97 for grade 0, 0.90 for grade 1, 0.88 for grade 2, and 0.82 for grade 3. The accuracy of the model in classifying the occlusal/incisal and non-occlusal/non-incisal surfaces was 0.92 and 0.93, respectively, and there was no significant difference between the two groups.

The confusion matrix of the classification model on the test set is shown in Fig. 6. Experts diagnosed 281 test images (50.81%) as TWI grade 0, 127 images (22.97%) as TWI grade 1, 115 images (20.80%) as TWI grade 2, and 30 images

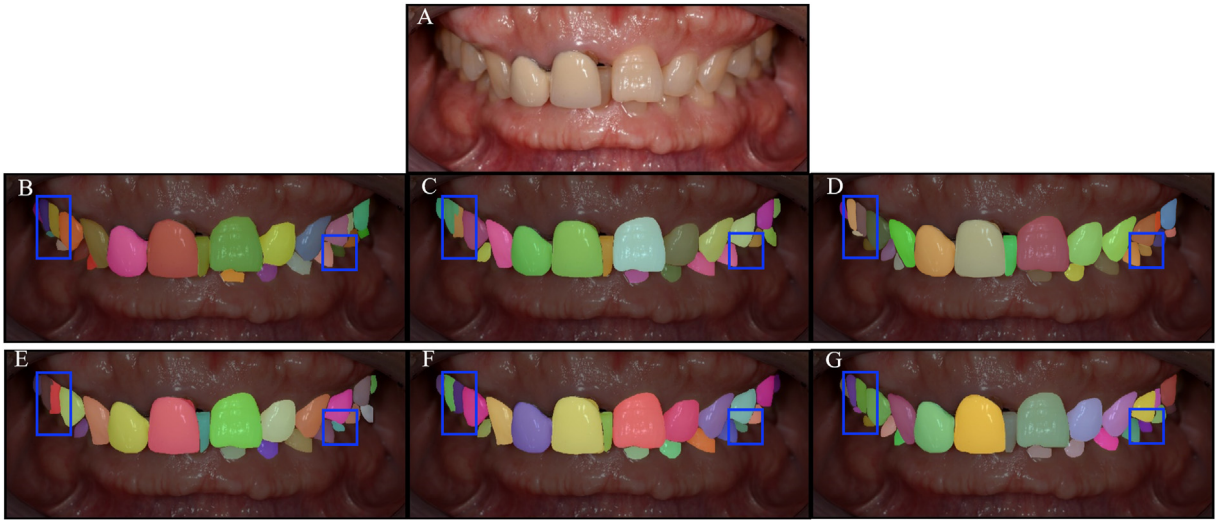

Figure 4 Example of input clinical image and the corresponding test results by different segmentation algorithms. (A) The clinical image input for the tooth segmentation task. (B) The expected output results of the tooth segmentation model. (C) The tooth segmentation results of the Mask R-CNN model. (D) The tooth segmentation results after replacing the FCN structure with the U-Net module. (E) The tooth segmentation results after incorporating the self-attention mechanism into the ResNet of the backbone structure. (F) The tooth segmentation results after incorporating U-Net and self-attention mechanism into the Mask R-CNN. (G) The tooth segmentation results after incorporating SGE U-Net and self-attention mechanism into the Mask R-CNN. Abbreviations: Mask R-CNN, Mask region-based convolutional neural network; FCN, Fully convolutional network; SGE, Spatial group-wise enhance.

(5.42%) as TWI grade 3. The AI-assisted diagnosis results showed 289 cases (52.26%) of TWI grade 0, 127 cases (22.97%) of TWI grade 1, 111 cases (20.07%) of TWI grade 2, and 26 cases (4.70%) of TWI grade 3. Among the 281 cases diagnosed by experts as TWI grade 0, 5 cases (1.78%) were diagnosed by the AI system as grade 1; among the 127 cases diagnosed by experts as TWI grade 1, 9 (7.09%) and 4 (3.15%) cases were diagnosed by the AI system as grades 0 and 2, respectively; among the 115 cases diagnosed by experts as TWI grade 2, 4 (3.48%), 8 (6.96%), and 3 (2.61%) cases were diagnosed by the AI system as grades 0, 1, and 3, respectively; among the 30 cases diagnosed by experts as TWI grade 3, 7 cases (23.33%) were diagnosed by the AI system as grade 2.

The accuracy of the binary AI model in the preliminary TW screening was 0.97. The Cohen's Kappa test result indicated that the AI model had a high level of consistency with experts in determining the existence of TW ($k = 0.93$, $P < 0.01$).

Table 1 Comparison of accuracy of different tooth segmentation algorithms.

Model	mPA
Mask R-CNN	0.78
Mask R-CNN + U-Net	0.80
Mask R-CNN + self-attention	0.83
Mask R-CNN + self-attention + U-Net	0.93
Mask R-CNN + self-attention + SGE U-Net	0.95

Abbreviations: mPA, Mean pixel accuracy; Mask R-CNN, Mask region-based convolutional neural network; SGE, Spatial group-wise enhance.

A comparison of diagnosis time between the AI model and human diagnosis is presented in Table 3. For manual reading, junior, midlevel, and senior clinicians take 3.32 s, 2.60 s, and 2.08 s, respectively, to determine the degree of TW on each surface, whereas the AI model takes only 0.07 s (0.02 s on average for tooth segmentation and 0.05 s on average for TW degree classification).

Discussion

TW is highly prevalent and exhibits a trend toward younger ages.[2] Severe TW can seriously affect both the aesthetics and functionality of a patient's teeth. The decision-making process for treating TW is affected by various factors and depends mainly on the severity and effects of the wear and

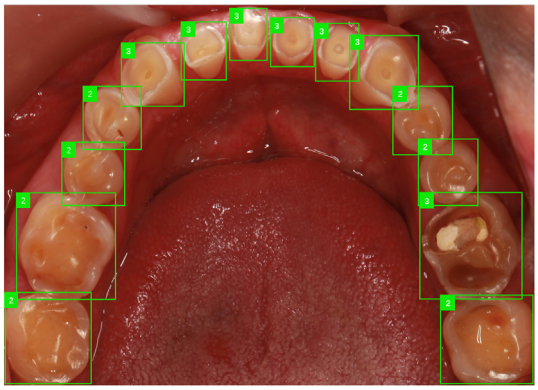

Figure 5 Results of deep learning-based tooth wear degree classification on intraoral photographs.

Table 2 Performance of the TW degree classification model.

	Total data			Occlusal/Incisal surface			Non-occlusal/Non-incisal surface		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
TWI 0	0.96	0.98	0.97	1.00	0.94	0.97	0.95	0.98	0.97
TWI 1	0.90	0.90	0.90	0.95	0.95	0.95	0.87	0.87	0.87
TWI 2	0.90	0.87	0.88	0.91	0.95	0.93	0.88	0.73	0.79
TWI 3	0.88	0.77	0.82	0.86	0.78	0.82	1.00	0.71	0.83
Accuracy	0.93	0.93	0.93	0.92	0.92	0.92	0.93	0.93	0.93
Macro-averaged	0.91	0.88	0.89	0.93	0.91	0.92	0.93	0.82	0.87
Weighted-averaged	0.93	0.93	0.93	0.92	0.92	0.92	0.93	0.93	0.93

Abbreviations: TW, Tooth wear; TWI, Tooth wear index.

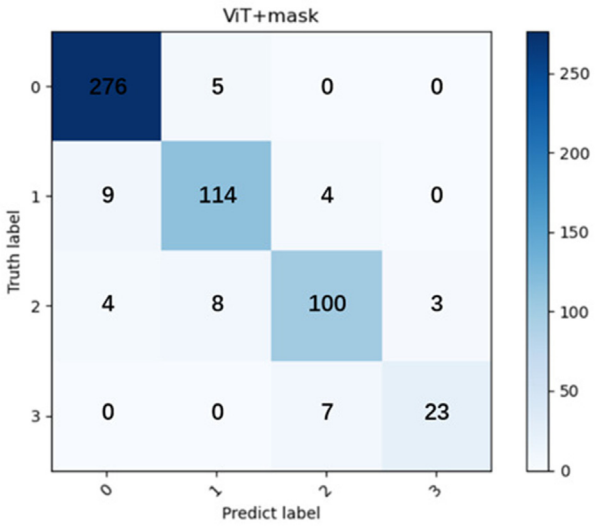

Figure 6 Confusion matrix of the classification model. Abbreviation: ViT, Vision transformer.

the wishes of the patient.[1] The severity of TW is a crucial clinical indicator. Traditionally, the assessment of TW severity was mainly conducted manually. However, the examiners' subjectivity and experience often lead to inconsistency, affecting the accuracy and repeatability of the evaluation. Moreover, conducting a detailed index evaluation of every tooth surface of the dentition is time-consuming. This study developed a DL-based system for determining the degree of TW from intraoral photographs, providing accurate and efficient diagnostic opinions for clinicians; the system is also applicable to the routine monitoring of TW.

The intelligent TWGS developed in this study consists of two sections: tooth segmentation and TW degree classification. Large quantities of studies have applied DL models to tooth segmentation, and typical models include U-Net, Mask R-CNN, and Segmentation Transformer.[31] U-Net is a typical encoder and decoder structure proposed in 2015.[32] The network can successfully realize the multilevel feature representation and fusion of medical images, and can incomplete high-accuracy segmentation tasks on small samples. Although U-Net performs well in semantic segmentation, it cannot perform instance segmentation. Mask R-CNN was proposed in 2017; it is an object detection algorithm based on Faster R-CNN.[29] It can separate objects in images at the pixel level based on object localization, achieving high precision and efficiency in object detection and instance segmentation. Considering the difficulty in recognizing occluded tooth structures, the variables in image quality, and the relatively small number of cases, the tooth segmentation task of this study combined Mask R-CNN with U-Net and introduced the SGE attention mechanism to identify fine structures and the rough boundaries between teeth and the background, as well as prevent overfitting caused by the small datasets. The results indicated that the accuracy of the segmentation model was 0.95. Segmentation errors mainly occur in regions with fine tooth structures in the dark part of the image and segmentation edges. However, these errors have less impact on the results of the subsequent classification task.

In the classification task, the TW severity was discriminated by classifying the TW degree of the individual tooth surfaces output by the segmentation task. A ViT model incorporating a mask mechanism was constructed for classification. The emergence of the ViT algorithm in 2020[30] was a breakthrough in image processing. It slices images

Table 3 Comparison of diagnostic time (mean [SD]) for each tooth surface between clinicians and the AI model (n = 100).

	TWI 0	TWI 1	TWI 2	TWI 3	Average
Junior clinician (min)	2.79 (0.67)	3.96 (1.41)	3.41 (1.07)	3.11 (1.07)	3.32 (0.99)
Midlevel clinician (min)	2.36 (0.45)	2.67 (0.61)	2.60 (0.68)	2.78 (0.85)	2.60 (0.61)
Senior clinician (min)	2.04 (0.49)	2.28 (0.48)	2.11 (0.41)	1.88 (0.35)	2.08 (0.42)
Average	2.39 (0.65)	2.97 (1.21)	2.71 (0.97)	2.59 (0.99)	2.67 (0.91)
AI model (s)	0.07 = 0.02 (segmentatin model) + 0.05 (classification model)				

Abbreviations: SD, Standard deviation; AI, Artificial intelligence; TWI, Tooth wear index.

into patches and adopts the multi-head attention mechanism to extract global features and find relationships among different image regions, showing a performance improvement over many previous DL models based on CNN.[33] In this study, considering the diversity of input image categories and the limitation of the sample size, the model introduced a mask mechanism into ViT. During the model training, the pixels of the input images were replaced at a certain probability with a mask generated based on the average of the surrounding pixels, thereby alleviating the overfitting problem. In addition, the dataset was expanded through data augmentation to improve the generalizability of the model and avoid overfitting. The results suggested that the accuracy of this classification model was 0.93, and no significant difference was detected between functional and nonfunctional surfaces.

The accuracy of the TWGS varied when discriminating between different degrees of TW. When TW is relatively mild, it is difficult to determine with the naked eye whether the dentin surface is covered with a thin layer of enamel, making accurate judgment challenging. Consequently, there were 14 cases of misjudgment between TWI grades 0 and 1 in the test set, accounting for 35% of all misjudgments. Moreover, the results demonstrated that the higher the TW degree, the lower the accuracy of the system's recognition, specifically for TWI 3, with an F1-score of only 0.82, indicating that the TWGS has a low recognition ability for secondary dentin and pulp exposure. This could be owing to the following reasons: first, the relatively complex morphology of secondary dentin makes it more difficult for the network to learn the features. Second, the study included only 138 annotated images for TWI 3, which is slightly insufficient. Future improvements to the TWGS will focus on collecting and analyzing photographs of teeth with severe wear to enhance the efficiency of data collection and effectiveness of the training process, thereby improving the accuracy of the system's judgments on severe TW. Additionally, the F1-score of the TWGS for the recognition of TWI grades 1 and 2 was 0.90 and 0.88, respectively. There were 12 cases of misclassification between TWI grades 1 and 2, accounting for 30% of all misclassifications. This could be owing to the following reasons: the input images of the classification model were annotated as semi-quantitative TWI grading, the grades 1 and 2 of which were differentiated based on a threshold of one-third of the tooth surface in terms of dentin exposure proportion. Evaluation is relatively subjective and it is difficult to determine situations that approach a critical point. Our future work will further optimize the research method by segmenting the dentin-exposed area using DL algorithms and then calculating its proportion on the corresponding surface by computer, thereby improving the precision and objectivity of judgments.

Another advantage of TWGS lies in the time required for diagnosis. The manual image reading for assessing TW degree takes an average of 2.67 s per tooth surface, whereas the TWGS only requires an average running time of 0.07 s (0.02 s for segmentation and 0.05 s for classification). To evaluate the time consumption of TW diagnosis accurately, this study recorded the time required for junior, midlevel, and senior clinicians to manually determine different degrees of TW and compared with that of TWGS. The results indicated that the manual diagnostic time was related to the clinician's experience, with a minimum time consumption of 2.08 s for senior clinicians, which is still far higher than that of TWGS. In addition, different TW levels required different manual reading times, with a range of 2.39 s (TWI 0) to 2.97 s (TWI 1), whereas the AI diagnosis can be completed within 0.1 s irrespective of TW degree. Therefore, the TWGS not only ensures high accuracy but also significantly improves image-reading efficiency.

For the evaluation of TW in individual patients, because the TWGS diagnoses the degree of wear based on individual tooth surfaces, it can determine the amount of TW (grading) for each tooth. Additionally, it can determine the tooth surfaces that are affected and the number of teeth involved (localized or extensive) from the perspective of the entire dentition by combining intraoral photographs from different angles. With the TWGS, precise assessment and longitudinal monitoring of TW can be achieved by comparing diagnostic data at different time points, thereby assisting doctors in making decisions on treatment plans and timing. However, the nine photographs included did not cover the lingual/palatal side of the posterior and lower anterior teeth, which is not conducive for making a comprehensive judgment of an individual patient's dentition. Additionally, the photographs acquired by different devices and photographers differ in color, resolution, and angle, which can affect the judgment accuracy of the TWGS. Recently, the rapid development of digital technology has provided new research approaches for TW diagnosis. Three-dimensional tooth images obtained through digital scanning technology can clearly and completely record the TW characteristics of the entire dentition, thereby helping doctors observe the TW and determine the severity more accurately and comprehensively.[34] In the future, three-dimensional images obtained from intraoral scans could be used as an alternative to intraoral photograph data for the TWGS. Based on the three-dimensional data obtained using built-in light sources and variable camera angles of intraoral scanners, the image quality can be standardized and the angles can be flexibly adjusted to recognize every tooth surface, enabling a comprehensive judgment of the TW degree of all tooth surfaces in the entire dentition.

From a preventive medicine perspective, early detection and intervention are crucial for TW treatment. The binary model can be applied for the preliminary evaluation of TW (defined as the presence or absence of dentin exposure). The accuracy of the binary model was 0.97, and an excellent agreement was achieved between TWGS and experts with a Cohen's Kappa score of 0.93. Therefore, the model might also be used as an effective initial screening method for TW as well as for daily self-monitoring of oral health. Currently, with the development of 5G technology, mobile application softwares based on smartphones and AI technology (such as iHome and OralCam) provide the possibility of maintaining users' oral health. By uploading tooth images with the help of specific intelligent dental devices or via smartphones, the system can conduct a preliminary screening of oral diseases, such as caries, periodontal disease, and tooth cracks.[35] However, the current software

does not yet have the functionality to recognize TW. The TWGS developed in this study can be further integrated with personal mobile devices to become an affordable and daily use TW recognition system for the public, providing a convenient, low-cost means of early warning, monitoring, and personalized prevention guidance for TW.

The limitations and future research directions of this study are as follows: (1) The output results of the current AI system were obtained based on learning from the intraoral photographs. However, the sample size of the training and validation sets for AI learning is relatively small with an underbalanced sample size between groups. These deficiencies will be rectified and optimized in the future research with the increase of sample size, especially for the teeth with severe wear. (2) In the training process of the classification network, teeth with caries, fillings, and restorations that affect the judgment of TW degree were excluded from the dataset. As a result, the model cannot adequately separate the aforementioned conditions from wear in actual applications. The TWGS could be optimized further by automatically identifying worn teeth for inclusion before TW degree determination. (3) This study only used the modified-TWI as the evaluation standard for TW degree. As the most widely used index grading method, it can grade and quantify the severity of different types of TW; however, this index still has defects such as low scoring of the degree of lesion[9] and failure to integrate TW grading with clinical treatment. In future work, we will include more comprehensive TW evaluation systems (such as BEWE and TWES) to better assist clinicians in making treatment decisions. The corresponding guidelines will also be established and confirmed further. (4) Currently, there are various algorithms for AI to perform image recognition, and the identification results differ under different algorithm frameworks. Future research also needs to continuously refine and optimize the algorithm of the TWGS to improve the accuracy of the system further.

In conclusion, our study demonstrated that the intelligent TWGS based on the CNN and ViT algorithms, using intraoral photographs, can perform TW degree grading accurately and efficiently. The system has the potential to help clinicians identify the severity of TW, thereby providing a basis for designing treatment plans and formulating preventive strategies against TW. Meanwhile, it might benefit patients in self-assessment and disease follow-ups. This study lays a solid foundation for the future development of more precise and convenient intelligent diagnostic systems for TW.

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Acknowledgements

The study received funding supported by the Beijing Natural Science Foundation (No. 7232223 and L222023), the Program for New Clinical Techniques and Therapies of the Peking University School and Hospital of Stomatology (PKUSSNCT-22B08), and the Academician Expert Workstation of Yunnan Province (202305AF150174).

References

1. Loomans B, Opdam N, Attin T, et al. Severe tooth wear: European consensus statement on management guidelines. *J Adhesive Dent* 2017;19:111—9.
2. Oudkerk J, Grenade C, Davarpanah A, Vanheusden A, Vandenput S, Mainjot AK. Risk factors of tooth wear in permanent dentition: a scoping review. *J Oral Rehabil* 2023;50:1110—65.
3. Entezami S, Peres KG, Li HH, Albarki Z, Hijazi M, Ahmed KE. Tooth wear and socioeconomic status in childhood and adulthood: findings from a systematic review and meta-analysis of observational studies. *J Dent* 2021;115:103827.
4. Schlueter N, Luka B. Erosive tooth wear-a review on global prevalence and on its prevalence in risk groups. *Br Dent J* 2018;224:364—70.
5. Smith BG, Knight JK. An index for measuring the wear of teeth. *Br Dent J* 1984;156:435—8.
6. Bartlett D, Ganss C, Lussi A. Basic Erosive Wear Examination (BEWE): a new scoring system for scientific and clinical needs. *Clin Oral Invest* 2008;12(Suppl 1):S65—8.
7. Wetselaar P, Lobbezoo F. The tooth wear evaluation system: a modular clinical guideline for the diagnosis and management planning of worn dentitions. *J Oral Rehabil* 2016;43:69—80.
8. Wetselaar P, Wetselaar-Glas MJM, Katzer LD, Ahlers MO. Diagnosing tooth wear, a new taxonomy based on the revised version of the Tooth Wear Evaluation System (TWES 2.0). *J Oral Rehabil* 2020;47:703—12.
9. Bardsley PF. The evolution of tooth wear indices. *Clin Oral Invest* 2008;12(Suppl 1):S15—9.
10. Tan JG. Etiology and differential diagnosis of severely worn dentition. *Zhonghua Kou Qiang Yi Xue Za Zhi* 2020;55:599—602.
11. Mehta SB, Loomans BAC, Bronkhorst EM, Banerji S, Bartlett DW. The impact of e-training on tooth wear assessments using the BEWE. *J Dent* 2020;100:103427.
12. Kulkarni PA, Singh H. Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. *JAMA* 2023;330:317—8.
13. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436—44.
14. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med* 2019;25:24—9.
15. Khanagar SB, Al-Ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry - a systematic review. *J Dent Sci* 2021;16:508—22.
16. Kühnisch J, Meyer O, Hesenius M, Hickel R, Gruhn V. Caries detection on intraoral images using artificial intelligence. *J Dent Res* 2022;101:158—65.
17. Setzer FC, Shi KJ, Zhang ZY, et al. Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images. *J Endod* 2020;46:987—93.
18. Johari M, Esmaeili F, Andalib A, Garjani S, Saberkari H. Detection of vertical root fractures in intact and endodontically treated premolar teeth by designing a probabilistic neural network: an ex vivo study. *Dentomaxillofacial Radiol* 2017;46:20160107.
19. Jeon SJ, Yun JP, Yeom HG, et al. Deep-learning for predicting C-shaped canals in mandibular second molars on panoramic radiographs. *Dentomaxillofacial Radiol* 2021;50:20200513.
20. You WZ, Hao AM, Li S, et al. Deep learning-based dental plaque detection on permanent teeth and the influenced factors. *Zhonghua Kou Qiang Yi Xue Za Zhi* 2021;56:665—71.
21. Nguyen KCT, Duong DQ, Almeida FT, et al. Alveolar bone segmentation in intraoral ultrasonographs with machine learning. *J Dent Res* 2020;99:1054—61.

22. Feres M, Louzoun Y, Haber S, Faveri M, Figueiredo LC, Levin L. Support vector machine-based differentiation between aggressive and chronic periodontitis using microbial profiles. *Int Dent J* 2018;68:39—46.
23. Ver Berne J, Saadi SB, Politis C, Jacobs R. A deep learning approach for radiological detection and classification of radicular cysts and periapical granulomas. *J Dent* 2023;135:104581.
24. Yang SY, Li SH, Liu JL, et al. Histopathology-based diagnosis of oral squamous cell carcinoma using deep learning. *J Dent Res* 2022;101:1321—7.
25. Weingart JV, Schlager S, Metzger MC, et al. Automated detection of cephalometric landmarks using deep neural patchworks. *Dentomaxillofacial Radiol* 2023;52:20230059.
26. Gupta A, Kharbanda OP, Sardana V, Balachandran R, Sardana HK. A knowledge-based algorithm for automatic detection of cephalometric land- marks on CBCT images. *Int J Comput Assist Radiol Surg* 2015;10:1737—52.
27. Amasya H, Yildirim D, Aydogan T, Kemaloglu N, Orhan K. Cervical vertebral maturation assessment on lateral cephalometric radiographs using artificial intelligence: comparison of machine learning classifier models. *Dentomaxillofacial Radiol* 2020;49:20190441.
28. Yu HJ, Cho SR, Kim MJ, Kim WH, Kim JW, Choi J. Automated skeletal classification with lateral cephalometry based on artificial intelligence. *J Dent Res* 2020;99:249—56.
29. He KM, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. Venice, Italy: ICCV), 2017. October 22-29.
30. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. In: *Proceedings of the 9th international conference on learning representations*. Vienna: OpenReview.net, 2021. May 3-7.
31. Chen XK, Ma N, Xu TK, Xu C. Deep learning-based tooth segmentation methods in medical imaging: a review. *Proc Inst Mech Eng H* 2024;238:115—31.
32. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *MICCAI 2015: 18th international conference*. Cham: Springer, 2015.
33. Al-Hammuri K, Gebali F, Kanan A, Chelvan IT. Vision transformer architecture and applications in digital health: a tutorial and survey. *Vis Comput Ind Biomed Art* 2023;6:14.
34. Kumar S, Keeling A, Osnes C, Bartlett D, O'Toole S. The sensitivity of digital intraoral scanners at measuring early erosive wear. *J Dent* 2019;81:39—42.
35. Liu LZ, Xu JW, Huan YX, Zou Z, Yeh SC, Zheng LR. A smart dental health-iot platform based on intelligent hardware, deep learning, and mobile terminal. *IEEE J Biomed Health Inform* 2020;24:898—906.