Submitted: 27 May, 2025 Accepted: 24 July, 2025 Published: 03 January, 2026 DOI:10.22514/jocpd.2026.011

Journal of Clinical Pediatric Dentistry

ORIGINAL RESEARCH

Automated processing of intraoral clinical photographs using deep learning techniques

Jungmin Eum[1,†], Hyejun Seo[2,†], Taesung Jeong[1,3], Eungyung Lee[1,3], Soyoung Park[1,3], Jonghyun Shin[1,3,*]

[1]Department of Pediatric Dentistry, School of Dentistry, Pusan National University, 50612 Yangsan, Republic of Korea
[2]Department of Dentistry, Ulsan University Hospital, University of Ulsan College of Medicine, 44033 Ulsan, Republic of Korea
[3]Dental and Life Science Institute & Dental Research Institute, School of Dentistry, Pusan National University, 50612 Yangsan, Republic of Korea

***Correspondence**
jonghyuns@pusan.ac.kr
(Jonghyun Shin)

[†] These authors contributed equally.

Abstract

Background: Intraoral clinical photographs are essential for diagnosis and treatment planning; however, capturing and managing high-quality images in pediatric dentistry is challenging. This study aimed to develop a deep-learning model for classifying and editing five types of intraoral clinical photographs (frontal, left buccal, right buccal, upper occlusal, and lower occlusal) of pediatric patients. **Methods**: A total of 3100 intraoral clinical photographs of 620 pediatric patients were used. The images were aligned to the normal occlusal plane and annotated into five categories. Two deep convolutional neural networks were implemented: Inception-ResNet-v2 for image orientation regression and Faster Region-based Convolutional Neural Network (R-CNN) for region detection. The models were trained and validated using five-fold cross-validation in Matrix Laboratory (MATLAB) 2024b, and their performance was assessed based on root mean squared error (RMSE), mean absolute error (MAE), classification accuracy, Intersection over Union (IoU), precision, recall, and average precision. **Results**: The image orientation correction network yielded a mean RMSE of 0.571° and an MAE of 0.407° in a five-fold cross-validation. The region detection network achieved a high classification accuracy across the five intraoral categories, with values ranging from 0.977 to 0.997. As the IoU threshold increased, average precision values decreased. **Conclusions**: The results show that our proposed method enables automated processing of intraoral photographs, particularly in terms of image rotation correction, cropping, and classification, with sufficiently high accuracy.

Keywords

Artificial intelligence; Convolutional neural network; Deep learning; Image processing; Pediatric dentistry; Photography; Dental

1. Introduction

Imaging data play a crucial role in most medical settings. In dentistry, imaging is essential for diagnosing conditions, planning treatment, monitoring progress, and facilitating communication between clinicians and patients [1]. Orthodontists rely on imaging data to make clinical decisions, monitor tooth movement, and plan treatments [1]. Intraoral clinical photographs are non-invasive, radiation-free, and can be archived for long-term storage, thus allowing clinicians to thoroughly examine a patient's oral condition over time [2]. These images provide detailed information on tooth morphology, alignment, and gingival health. Furthermore, clinical photographs can be used independently to assess clinical characteristics or combined with other diagnostic tools such as plaster models and radiographic images [3].

Digital photographs are an effective diagnostic tool in pediatric dentistry [4–8]. This advancement supports the relationship between clinicians and patients in terms of diagnosis, clinical documentation, and patient communication [9]. The use of photographs for diagnostic purposes in pediatric dentistry is a cost-effective and non-invasive method [10]. However, capturing high-quality photographs of pediatric patients is challenging because of technical difficulties. Factors such as oral cavity size, patient behavior, and limited clinical time hampers image acquisition, potentially compromising the clinical procedure and image quality. To maintain cooperation, clinicians minimize repeated captures and improve image quality through alignment, cropping, and other adjustments [9].

Traditionally, these images are labeled and stored manually based on their clinical characteristics. However, with advancements in digital dentistry, image data are now more frequently archived in digital clinical record systems, streamlining access for future evaluation and patient follow-up [11]. Efficient use of image data requires basic categorization and enhancements in quality and consistency through editing [12]. Therefore, developing a system integrating automated classification and editing would be beneficial for improving dental practice efficiency and reducing specialist workload.

As artificial intelligence (AI) becomes a new standard in dentistry and healthcare, deep learning, a specialized branch of AI that utilizes convolutional neural networks (CNNs), is now widely applied to the analysis of image data [3]. Currently, deep learning is used in dental image processing, demonstrating high accuracy in detecting plaque and caries, assessing the need for extraction-based orthodontics, and automatically classifying images [1–3, 13, 14].

Previous studies using deep learning for intraoral image classification have reported an accuracy of greater than 98% in distinguishing intraoral and extraoral clinical photographs [1, 3]. In addition, most commercially available deep learning-based diagnostic software systems—such as CellmatiQ (DentaliQ ortho, CellmatiQ GmbH, Hamburg, HH, Germany), ORCA AI (CephX v4.02, ORCA Dental AI Ltd., Herzliya, Israel), and WebCeph (version 1.0.0, AssembleCircle Corp., Gyeonggi-do, Korea)—primarily support tasks such as landmark detection and cephalometric analysis [15]. While some CNN-based models have been developed for classifying views or identifying dental findings and anatomical regions [3, 13], these approaches have primarily focused on single-purpose tasks. Moreover, to the best of our knowledge, additional preprocessing tasks, such as image orientation and cropping, require new deep learning approaches that have been rarely addressed in previous studies or integrated into existing systems [16].

Accordingly, this study implemented two CNNs, each specialized in image editing and classification. We aim to enhance the general efficiency and precision of the processing workflow by assigning each network a specific function [17]. This study focused on building a deep learning framework that can classify and edit five types of intraoral clinical photographs (frontal, left buccal, right buccal, upper occlusal, and lower occlusal) obtained from pediatric and adolescent patients.

2. Materials and methods

2.1 Image dataset

This retrospective study was conducted on 620 randomly selected patients who visited the Department of Pediatric Dentistry at Pusan National University Dental Hospital (Yangsan, South Korea) for orthodontic treatment between July 2022 and January 2024. A total of 323 male and 297 female participants were included in the study population, whose average age was 9.54 years, with ages ranging from 4 to 18 years. Dental situations such as missing teeth, orthodontic appliances, and prosthetic restorations were not excluded to reflect a wide range of real-world clinical conditions [3]. Images with significant acquisition errors were excluded, including the presence of foreign materials (*e.g.*, water droplets on mirrors), motion-related blurring, and insufficient retraction resulting in incomplete visualization of the teeth.

Several dentists took digital photographs, using multiple digital single-lens reflex (DSLR) cameras under various acquisition settings, including differences in aperture (F-number), International Organization for Standardization (ISO) sensitivity, and lighting conditions, and use of auto or manual focus. Metadata on the exact settings and operator identity were not recorded, and no reference was available regarding which clinician captured each image. This may have introduced random variability to the quality of photographs [3]. All photographs were taken from standardized positions to minimize variation in composition. One set of intraoral photographs comprised the frontal (in occlusion), left buccal (in occlusion), right buccal (in occlusion), upper occlusal (using a mirror), and lower occlusal (using a mirror) images. A retractor was also used during occlusion to provide a detailed examination of the oral cavity. The upper and lower occlusal photographs were captured using a mirror; therefore, the researcher performed mirror image reversal. A total of 3100 images were collected and archived at a resolution of 6000 × 4000 pixels in Joint Photographic Experts Group (JPEG) format. All intraoral images were anonymized prior to model development and analysis to comply with institutional review board requirements and data protection regulations. To ensure proper anonymization, each patient was assigned a non-identifiable numerical identifier (ID), and all corresponding intraoral images were labeled with sequential codes to maintain image grouping without including any personally identifiable information. Among the collected images, 3000 were used for model training and validation, while a separate test set of 100 images was prepared for performance comparison between the deep-learning model and human evaluators.

2.2 Data annotation and labelling

The collected data were processed in two steps to train the deep-learning model. First, the orientation was performed relative to the normal occlusal plane, and the angle of orientation was recorded by an expert pediatric dentist. PhotoScape version 3.7 software (MOOII Tech, Seoul, South Korea) was used for this process. To evaluate intra-rater reliability for the recorded orientation angles, a subset comprising 5% of the dataset (150 images) was randomly selected and remeasured by the same examiner after a two-week interval. The resulting intraclass correlation coefficient (ICC) was 0.998, suggesting a high degree of reproducibility. Next, the rotated clinical photographs were subjected to data labelling to process rectangular bounding boxes for the regions of interest using the Image Labeler in MATLAB 2024b software (MathWorks Inc., Natick, MA, USA). The labeling process, which was conducted by a different expert pediatric dentist, involved drawing a bounding box to exclude unnecessary information—such as cheek retractors, lips, and fingers—based on the cropping guidelines [18, 19]. Frontal, left buccal, right buccal, upper occlusal, and lower occlusal regions were annotated. To evaluate intra-rater reliability, 5% of the dataset was randomly selected and reannotated two weeks later, yielding an intersection over union (IoU) score of 0.944.

2.3 Deep convolutional neural networks

The proposed automatic method for processing intraoral photos consisted of two networks: Inception-ResNet-v2, used as a regression network for correcting intraoral photo orientation, and Faster R-CNN, a detection network for identifying regions of interest.

The orientation of the intraoral photographs was corrected using the Inception-ResNet-v2 model, which was pretrained with more than one million images from the ImageNet dataset. This model combines the strengths of the inception network and residual blocks of the ResNet backbone architecture, and has shown high accuracy in various deep learning studies in the dental and medical fields [18, 20]. To adapt to the task, the architecture was reconfigured from a classification network to a regression model through layer modifications. The convolutional layers of the original classification model were retained to extract the image features, whereas the final learnable and classification layers were modified to suit a regression task. Specifically, the original fully connected layer, softmax layer, and classification output were replaced with a single fully connected layer, producing one output value and enabling the network to perform regression instead of classification.

To detect regions of interest, this study utilized a Faster R-CNN model with Inception-ResNet-v2 as its backbone network. As an improved version of Fast R-CNN [21], Faster R-CNN addresses the limitations of its predecessor, which depends on a selective search [22] for generating region proposals, a step that often creates performance bottlenecks. Faster R-CNN employs a Region Proposal Network, which allows region proposals and feature extraction to be performed using a unified architecture. The network incorporates a softmax layer for classification and a bounding-box regression layer for precise localization. These two branches generate outputs that are compared against the ground truth annotations to compute the loss, and the model is trained end-to-end via backpropagation.

2.4 Training and five-fold cross-validation

All models were trained using MATLAB 2024b equipped with Deep Learning and Parallel Computing Toolboxes running on an NVIDIA RTX 4090 GPU (Santa Clara, CA, USA) (24 GB Random Access Memory (RAM)). The intraoral photo orientation correction network based on Inception-ResNet-v2 (Fig. 1a) was optimized using the Adam optimizer with a fixed initial learning rate of 1×10^{-4}. Training was conducted for a maximum of 100 epochs, using a minibatch size of 16. Validation was performed every 20 iterations, and early stopping was applied with a patience of 30 validation checks. The region of interest detection network based on Faster R-CNN (Fig. 1b) also used the Adam optimizer but employed a piecewise learning rate schedule, where the learning rate was reduced by a factor of 0.2 every 10 epochs. This network was likewise trained for up to 100 epochs with a minibatch size of 16, using the same validation strategy and early stopping criteria. In both cases, the training data were shuffled at the beginning of each epoch, and five-fold cross-validation was used to ensure robust performance evaluation.

For training and validation, all the images were resized to a format of 299 × 299 pixels to meet the input size requirements of the Inception-ResNet-v2 architecture, ensuring compatibility with the network and standardized preprocessing. To ensure stable convergence across cross-validation

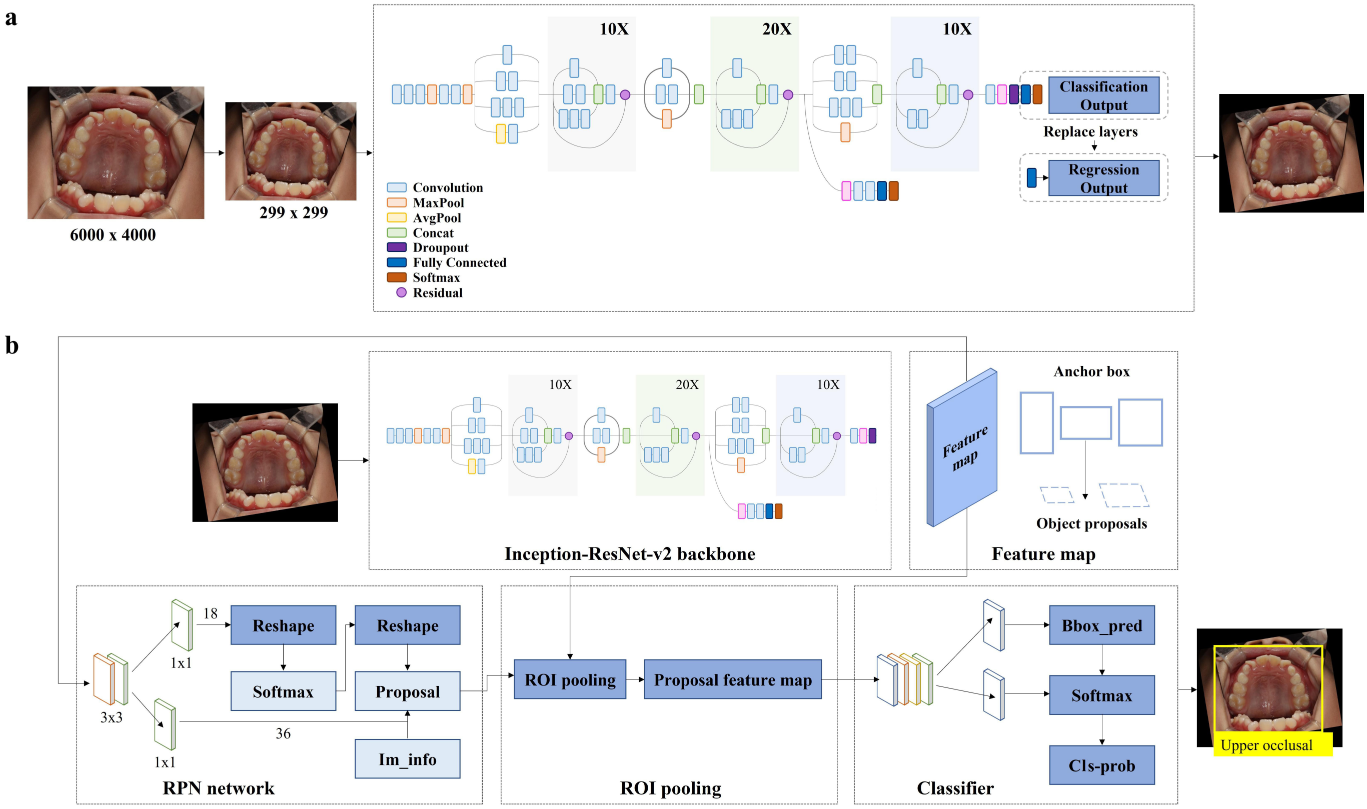


FIGURE 1. Schematic of the dual deep learning architecture. (a) Intraoral photo orientation correction network based on Inception-ResNet-v2. (b) ROI detection network based on Faster R-CNN. MaxPool: Max pooling; AvgPool: Average pooling; Concat: Concatenation; ROI: Region of interest; RPN: Region proposal network; Bbox_pred: Bounding box prediction; Cls-prob: Class probability; Im_info: Image information.

folds, the target labels (rotation angles) were standardized using z-score normalization, which centers the data around a mean of zero with unit variance. This approach was chosen to normalize the continuous angle values, thereby improving numerical stability during backpropagation and facilitating efficient optimization in the regression network. The parameters (mean and standard deviation) used for normalization were calculated from the combined training and validation sets, and the predicted outputs were denormalized back to degree values for evaluation and interpretation. Data augmentation was applied during training using the imageDataAugmenter function of MATLAB. The augmentation pipeline included random rotations (±5°), scaling (0.8–1.2×), and horizontal/vertical translations (±10 pixels), which were applied on-the-fly to introduce variability while preserving anatomical integrity. A five-fold cross-validation approach was adopted to improve generalizability and mitigate data bias considering the limited sample size. The dataset was randomly partitioned into five subsets, of which four subsets were allocated for training and one for validation in each iteration. This procedure was iterated five times, with each subset being used once as the validation set, thereby ensuring efficient and comprehensive utilization of the entire dataset.

2.5 Performance metrics

To assess the regression performance for intraoral photo orientation correction, we employed two standard metrics: root mean squared error (RMSE) and mean absolute error (MAE). These are defined in Eqns. 1,2, where y, $\hat{y}$, and $\bar{y}$ denote the ground truth, predicted value, and mean of the actual values, respectively, and n denotes the number of samples.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

For classification performance, the accuracy was computed by comparing the predicted and actual labels for each detected region. The label with the highest confidence score is considered the final prediction, as shown in Eqn. 3. Additionally, a confusion matrix and receiver operating characteristic (ROC) curve were evaluated to assess classification performance.

$$Classificaton\ Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{3}$$

The IoU metric, the ratio of the overlapping area to the union of the predicted and ground truth bounding boxes (Eqn. 4), was used to evaluate the localization performance.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{4}$$

The precision and recall were calculated using true positives (TP), false positives (FP), and false negatives (FN), with IoU thresholds set at 0.5, 0.75, and 0.5, respectively 0.95 in 0.05 increments (Eqns. 5,6).

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

Finally, the average precision (AP), a widely adopted metric in object detection, was computed as the area under the precision-recall (PR) curve (Eqn. 7).

$$AP = \int_0^1 P(r)\,dr \tag{7}$$

The overall workflow is illustrated in Fig. 2.

2.6 Statistical analysis

The investigated data were analyzed using SPSS 26.0 (SPSS Inc., IBM, Chicago, IL, USA). The Wilcoxon signed-rank test was used to verify the statistical significance of the image processing performance of deep-learning models and human evaluators.

3. Results

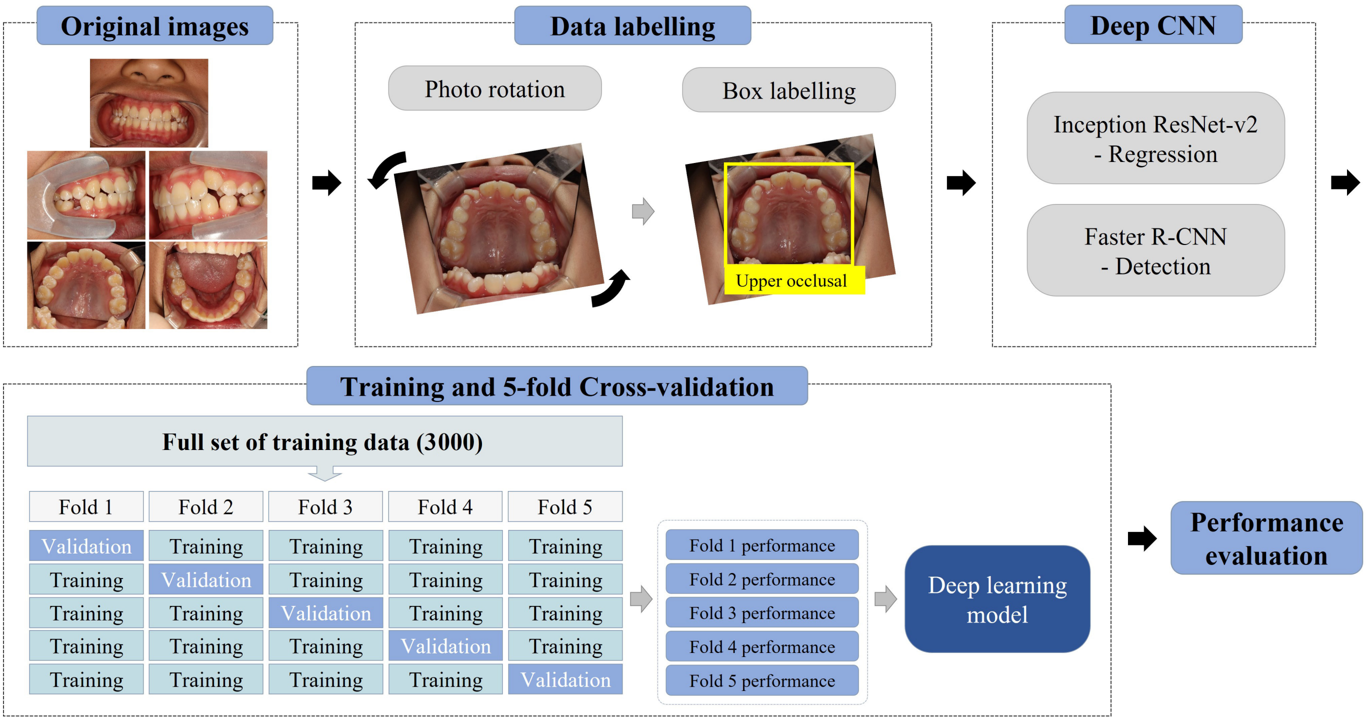
Both the intraoral photo orientation correction network based on Inception-ResNet-v2 and the region of interest detection network employing Faster R-CNN were evaluated through a five-fold cross-validation procedure to ensure robust performance assessment. The training curves for both networks exhibited consistent and gradual convergence of training and validation losses across all folds (Fig. 3). Although minor differences between training and validation losses were observed—reflecting typical generalization dynamics—the validation losses steadily approached values close to those of the training losses. This behavior indicates effective learning with minimal overfitting throughout the cross-validation process.

3.1 Intraoral photo orientation correction network

Intraoral photo orientation correction achieved RMSE and MAE values of 0.571° and 0.407°, respectively, based on the mean values from the five-fold cross-validation (Table 1). These results indicate that the Inception-ResNet-v2 model accurately estimated and corrected the orientation of intraoral photographs, aligning them with the standard occlusal plane.

3.2 Region of interest detection network

The outcomes of five-fold cross-validation, including the classification accuracy and AP scores under various IoU thresholds, are listed in Table 2. The mean classification accuracy



Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Validation	Training	Training	Training	Training
Training	Validation	Training	Training	Training
Training	Training	Validation	Training	Training
Training	Training	Training	Validation	Training
Training	Training	Training	Training	Validation

FIGURE 2. Schematic of the workflow. Original intraoral images were rotated and labeled. A regression network (Inception-ResNet-v2) was used to correct image orientation, and a detection network (Faster R-CNN) was used to classify image types. A 5-fold cross-validation was conducted to train and evaluate the deep learning model using 3000 images. CNN: convolutional neural network.

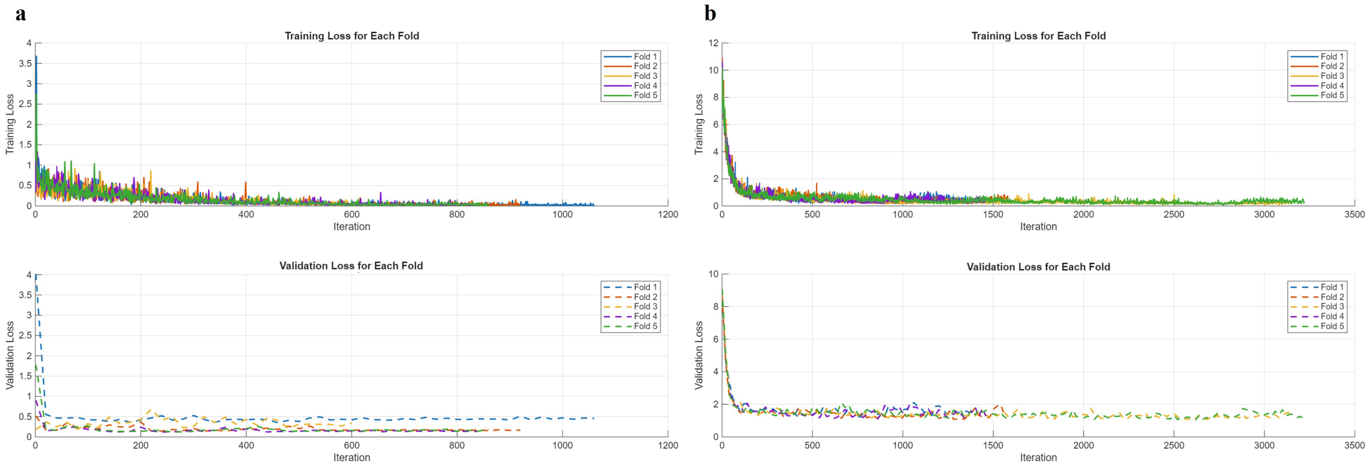


FIGURE 3. Training and validation loss curves for the two deep learning models evaluated using five-fold cross-validation. (a) Inception-ResNet-v2. (b) Faster R-CNN.

TABLE 1. Regression performance for photo orientation correction using five-fold cross-validation.

Fold	RMSE (degrees)	MAE (degrees)
1	0.392	0.285
2	0.897	0.571
3	0.497	0.374
4	0.579	0.442
5	0.488	0.364
Average	0.571	0.407

RMSE: root mean squared error; MAE: mean absolute error.

TABLE 2. Performance of classification and region of interest detection using 5-fold cross-validation.

Fold	Regions	Classification accuracy	AP (IoU threshold 0.5)	AP (IoU threshold 0.75)	AP (IoU threshold 0.5–0.95)
1					
	Frontal	0.992	0.911	0.820	0.609
	Left buccal	0.975	0.938	0.641	0.421
	Right buccal	0.992	0.956	0.696	0.473
	Upper	0.992	0.980	0.798	0.661
	Lower	0.983	0.942	0.872	0.707
2					
	Frontal	0.992	0.956	0.823	0.443
	Left buccal	0.983	0.946	0.648	0.425
	Right buccal	0.992	0.945	0.744	0.578
	Upper	0.992	0.988	0.863	0.672
	Lower	0.983	0.965	0.886	0.777
3					
	Frontal	1.000	0.941	0.841	0.616
	Left buccal	0.975	0.961	0.651	0.440
	Right buccal	0.992	0.954	0.705	0.472
	Upper	0.992	0.948	0.782	0.635
	Lower	0.983	0.937	0.861	0.684
4					
	Frontal	1.000	0.964	0.795	0.447
	Left buccal	0.975	0.960	0.702	0.480
	Right buccal	1.000	0.957	0.770	0.620
	Upper	0.992	0.966	0.869	0.666
	Lower	0.992	0.971	0.856	0.772
5					
	Frontal	1.000	0.960	0.817	0.467
	Left buccal	0.975	0.954	0.680	0.460
	Right buccal	1.000	0.952	0.775	0.609
	Upper	1.000	0.962	0.855	0.690
	Lower	1.000	0.969	0.898	0.801
Average					
	Frontal	0.997	0.946	0.819	0.517
	Left buccal	0.977	0.952	0.664	0.445
	Right buccal	0.995	0.962	0.738	0.550
	Upper	0.993	0.969	0.833	0.665
	Lower	0.988	0.965	0.875	0.748

AP: average precision; IoU: Intersection over Union.

for the frontal, left buccal, right buccal, upper occlusal, and lower occlusal images were 0.997, 0.977, 0.995, 0.993 and 0.988, respectively. The model's performance was evaluated using a confusion matrix and demonstrated excellent results in the classification of clinical images (Fig. 4a). Additionally, an ROC Curve was employed to assess the model's performance across different thresholds, and the high Area Under the Curve (AUC) further confirmed the model's robust overall performance (Fig. 4b).

To compute the AP values, PR curves were derived and differed depending on the applied IoU threshold. For our validation dataset, PR curves were generated using IoU thresholds of 0.5, 0.75, and a range spanning from 0.5 to 0.95 in increments of 0.05. As the threshold increased from 0.5 to 0.75 and then to 0.95, a decline in model performance was observed.

AP scores were calculated individually for all five intraoral regions. At the IoU threshold of 0.5, the regions with the highest to lowest AP were the upper occlusal, lower occlusal, right buccal, left buccal, and frontal regions. At a 0.75 threshold, the order changed to lower occlusal, upper occlusal, frontal, right buccal, and left buccal. When evaluated over the full IoU range (0.5–0.95 with 0.05 increments), lower occlusal again yielded the highest average AP, followed by upper occlusal, right buccal, frontal, and left buccal, respectively (Fig. 5).

Fig. 6 shows a sample input image fed into the CNN model, model output, and the final cropped region of interest.

3.3 Comparative evaluation of deep-learning model and human performance

The comparative results between human evaluators and the deep-learning model on the test set are presented in Table 3. Five dentists (D), each with 1 to 3 years of clinical experience, independently assessed a test set of 100 intraoral photographs comprising 20 images for each of the five standard intraoral views. The Wilcoxon signed-rank test revealed significant differences ($p < 0.001$) in single image processing time, MAE, RMSE, and IoU (threshold = 0.5) between the two groups.

4. Discussion

In this study, we developed a fully automated system for classifying and editing intraoral clinical photographs, and the evaluation results demonstrated high accuracy.

The Inception-ResNet-v2-based regression model used in this study achieved a remarkably low orientation error, with a mean RMSE of 0.571° and an MAE of 0.407°. Although both metrics quantify the discrepancy between predicted and actual orientation angles, they are complementary: RMSE penalizes larger errors more heavily and is thus sensitive to outliers, while MAE provides a direct and more interpretable measure of the average error magnitude [23, 24]. These values are clinically significant because RMSE and MAE fall well below the perceptual thresholds for angular detection reported in the literature—approximately 2° for dental professionals and 4° for laypersons [25]. This sub-degree level of precision suggests that the orientation correction achieved by our model is unlikely to be perceived by clinicians, thereby reinforcing its practical utility in the standardization of intraoral photographs and the automation of clinical workflows.

Notably, we developed an automatic orientation alignment model that considers various tooth eruption patterns in pediatric patients [26]. Although previous studies have attempted to identify the regions of interest in intraoral photographs [1, 3], research focusing exclusively on pediatric patients is scarce. Furthermore, to the best of our knowledge, no study has specifically addressed image-orientation alignment in pediatric intraoral photography. The proposed model was designed to achieve consistent rotational alignment even in pediatric patients with diverse tooth eruption stages and limited behavioral cooperation. By replacing the repetitive and time-consuming task of manual orientation alignment, the proposed model is expected to contribute to the standardization of image analysis

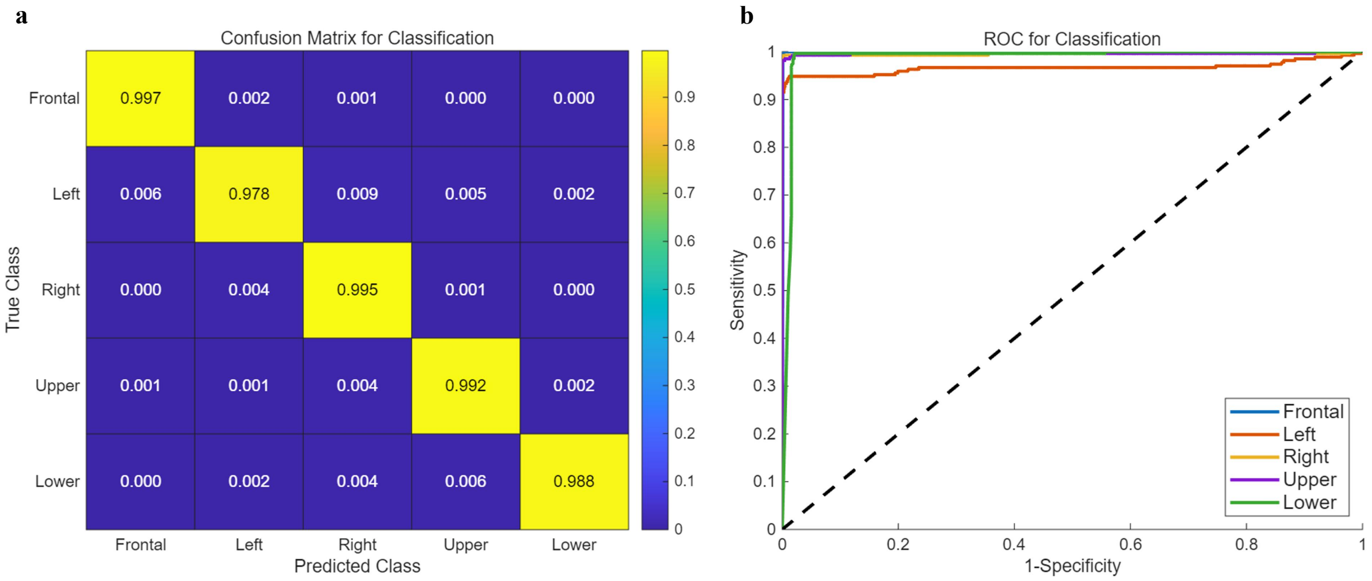


FIGURE 4. Classification performance of the Inception-ResNet-v2 model for intraoral image classification. (a) Confusion matrix. (b) ROC curve. ROC: Receiver operating characteristic.

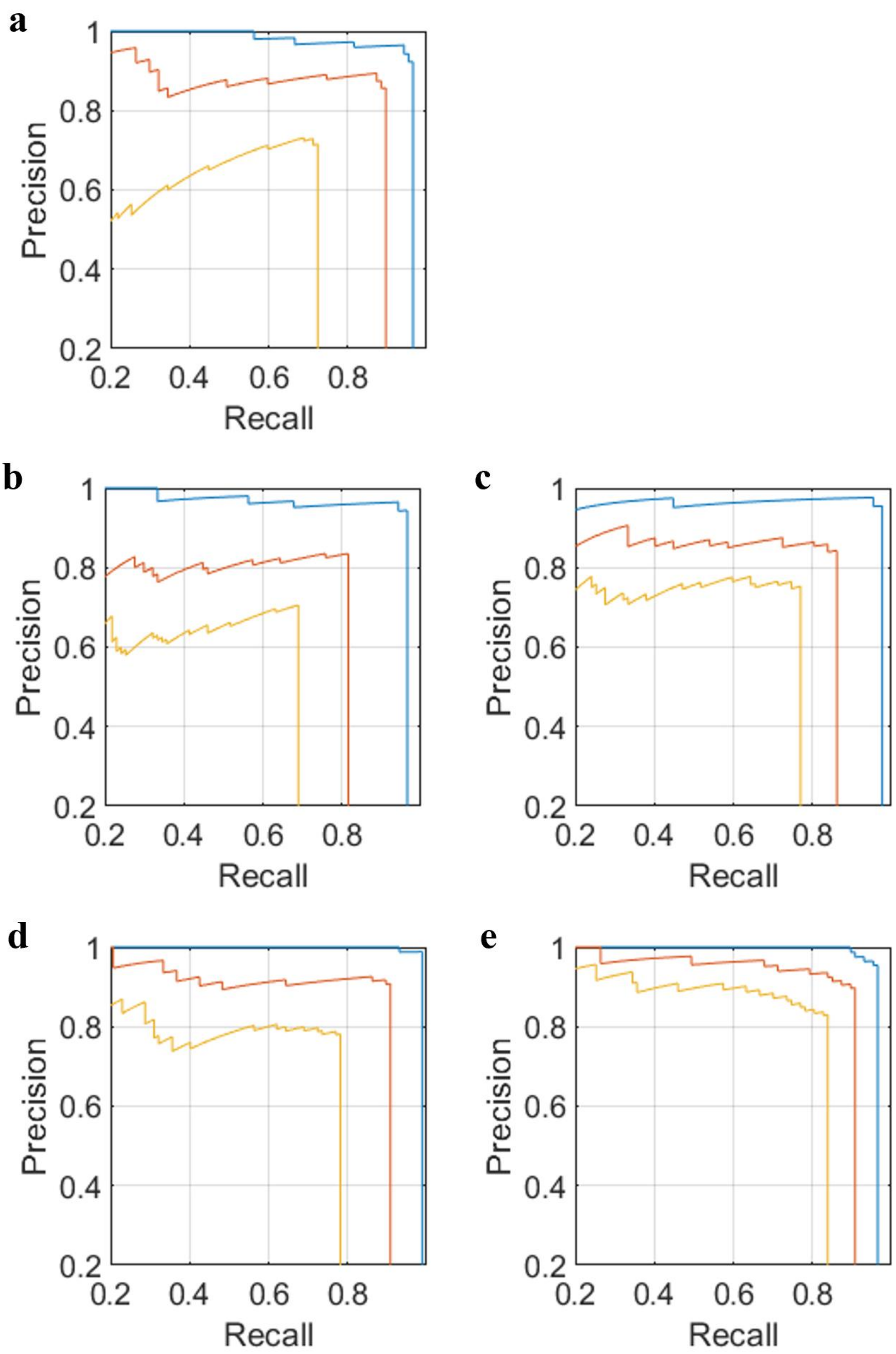


FIGURE 5. Variation in precision–recall (PR) curves across different intersection over union (IoU) thresholds. (a) Frontal. (b) Left buccal. (c) Right buccal. (d) Upper occlusal. (e) Lower occlusal. PR curves were generated at IoU thresholds of 0.50 (blue), 0.75 (orange), and 0.50–0.95 in 0.05 increments (yellow).

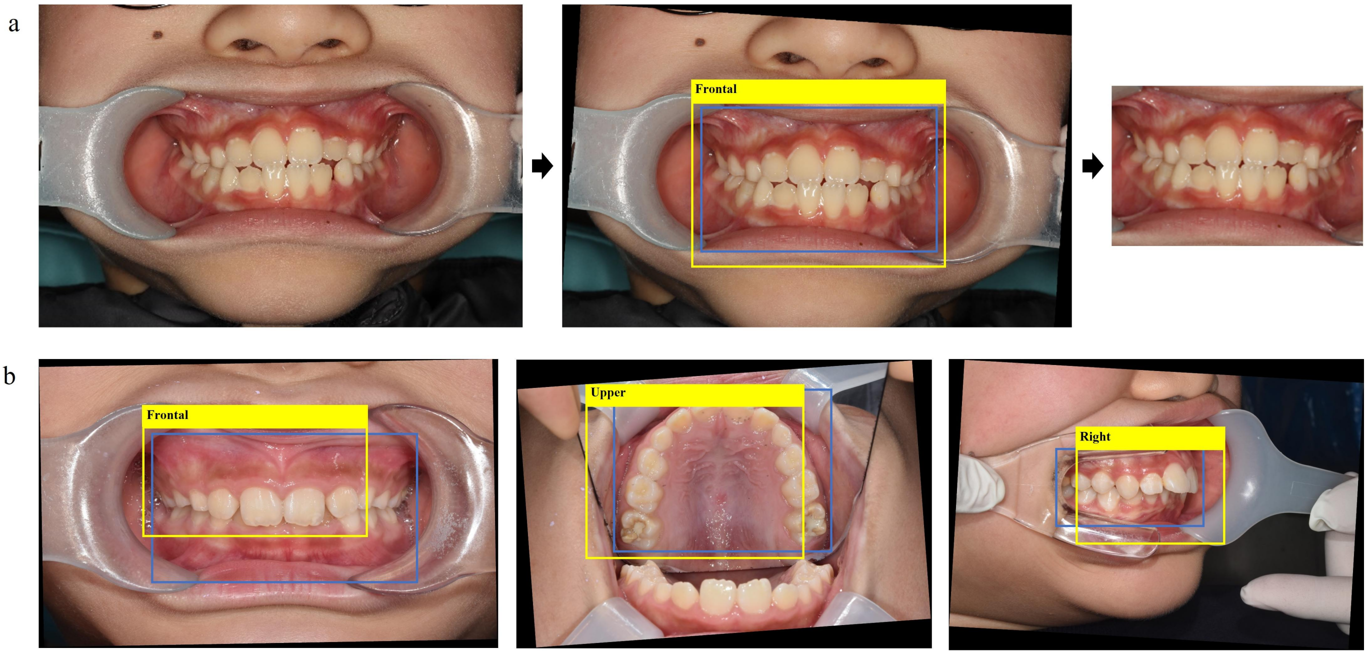


FIGURE 6. Example output showing orientation correction and region classification by the model. (a) The blue bounding box indicates the ground truth region of interest, annotated using the Image Labeler in MATLAB 2024b (MathWorks Inc., Natick, MA, USA), while yellow bounding box represents the model-predicted region of interest. (b) Representative failure cases demonstrating the model's limitations. From left to right, the images show a poorly localized region, an image with fogging on the intraoral mirror, and soft tissue interference caused by a retractor.

TABLE 3. Comparison of image processing performance between dentists and the deep-learning model.

	Single image processing time (s)*	Rotation angle prediction		Region of interest detection
		MAE (degrees)*	RMSE (degrees)*	IoU (at threshold 0.5)*
D (n = 5)	20.67 ± 4.95	0.720	1.181	0.907
DL	0.11	1.227	2.007	0.843

*Processing time includes both orientation correction and region classification tasks. Values for the dentist group (D) represent the mean performance of five evaluators. MAE: mean absolute error; RMSE: root mean square error; IoU: intersection over union; D: dentists; DL: deep-learning model. *Statistically significant differences between groups (Wilcoxon signed-rank test, $p < 0.001$).*

and improve clinical efficiency in pediatric dental practice.

The region of interest detection model based on Faster R-CNN demonstrated high performance, achieving a mean AP of over 0.946 at an IoU threshold of 0.5. However, the detection performance exhibited a declining trend as the IoU threshold increased (AP = 0.75, AP = 0.5–0.95). This decrease may reflect the inherent difficulty of identifying precise locations in clinical intraoral photographs, particularly in regions with complex soft tissue boundaries or variable imaging conditions, as higher IoU thresholds demand pixel-level accuracy. Among the various views, the highest detection performance was observed in the upper and lower occlusal surfaces, whereas a relatively lower performance was observed in the left and right buccal views. The superior performance of occlusal views may be attributed to their relatively consistent anatomical structures [27, 28] and the lower likelihood of interference from soft tissues during image acquisition, such as the tongue or buccal mucosa, which allows for more consistent imaging. However, buccal images are more prone to occlusion by intraoral structures and are more sensitive to variations in image quality due to differing capture conditions [28]. These factors may hinder the model's capacity to learn consistent features, which in turn may reduce detection performance. Therefore, further research is needed to improve the detection performance in buccal views. We hypothesize that expanding the dataset with additional intraoral images would further improve model performance. In addition, strategies such as standardizing image acquisition protocols and incorporating additional angled buccal views may enhance feature consistency and improve detection accuracy in future models [29].

When comparing the classification accuracy of intraoral clinical photographs with those of previous studies [1, 3], the present study achieved an accuracy of 99.0%, whereas Ryu *et al.* [3] and Li *et al.* [1] reported accuracies of 98.0% and 99.4%, respectively. Although slight differences in performance were observed due to variations in the model architecture and validation methods, the classification performance in this study was similarly high, which is consistent with previous findings. This high accuracy may be partially explained by the use of cross validation, which provides more stability and, in some cases, higher performance estimates than single-train–test split validation, particularly in studies with limited sample sizes [30].

These findings have important clinical implications. Compared with manual image processing, which is time-consuming and prone to interoperator variability [31], this level of automation enables a more efficient and consistent workflow. AI-based approaches can match or even surpass clinician performance in various diagnostic imaging tasks [11]. In addition, such systems have been shown to reduce the burden of repetitive and labor-intensive procedures in clinical dental practice [11, 12, 31]. Building on these findings, the present study represents a step toward the integration of deep learning tools into pediatric dental care, contributing to improved standardization, reliability, and timeliness in image-based assessments.

When compared directly with human evaluators using the test set, the deep-learning (DL) model demonstrated a substantial reduction in evaluation time (0.11 seconds per image) relative to dentists (20.67 ± 4.95 seconds), emphasizing its potential value in time-sensitive clinical workflows. However, it should be noted that the model exhibited slightly lower accuracy in orientation correction and region detection than human evaluators. This discrepancy may indicate a degree of overfitting or sample bias, particularly given the relatively small dataset. As with most deep learning approaches, our model is subject to general limitations such as overfitting, data-specific bias, and limited interpretability. These challenges underscore the importance of integrating explainable AI techniques and validating model performance across external datasets [32].

This study had several limitations. First, all photographs were obtained from a single institution. Although they were captured by multiple dentists using various cameras and acquisition settings, this internal variability may not sufficiently mitigate potential biases arising from institution-specific imaging protocols and patient populations. Second, despite implementing data augmentation and five-fold cross-validation to enhance model performance and mitigate overfitting, the relatively limited dataset size remains a significant limitation. To improve the robustness and generalizability of the proposed model, future studies should incorporate larger datasets, including external or multicenter validation. In particular, incorporating more demographically, clinically, and technically diverse data could broaden the model's applicability beyond pediatric orthodontics to adult populations and other dental specialties, thereby enhancing its clinical utility. Third, the retrospective nature of this study limited our ability to perform a stratified analysis of model performance based on device type or image acquisition parameters, as metadata regarding camera models, specific settings, and operator identity were unavailable. Variability in medical image data—arising from differences in clinical protocols, imaging devices, acquisition settings (such as resolution, lighting, and focus), and operator technique—can lead to inconsistencies in image quality. These inconsistencies can significantly affect the performance and generalizability of deep learning models [33–35]. In particular, low-quality images—such as those affected by blur, low resolution, or noise—have been shown to reduce the accuracy of CNN-based models, especially when such cases are underrepresented in the training dataset [36]. Therefore, future studies should prospectively collect detailed acquisition metadata and incorporate objective image quality metrics to enable subgroup analyses. Moreover, incorporating a wider range of image qualities—including suboptimal images commonly encountered in clinical practice—and applying data augmentation techniques such as color and contrast adjustments, as well as degradations like noise and blur, may collectively enhance the model's robustness to real-world variability [37].

In addition to addressing these limitations, future studies evaluating clinician acceptance and workflow integration will be essential to ensure the successful use of AI-based tools in pediatric dentistry. To maximize clinical utility, future research could also examine the performance of the model across diverse patient populations and imaging conditions. The dual deep learning model adopted in this study also holds promise for future research. Particularly, the classified and edited intraoral clinical photographs may be further used for simultaneous tasks, such as identifying restoration types and detecting dental caries. Moreover, the system could be integrated into broader diagnostic workflows, such as orthodontic assessments and patient education tools.

5. Conclusions

In this study, we developed a deep learning-based model for the automated processing of five standard types of intraoral clinical photographs—including orientation correction, cropping, and classification—in pediatric dental patients. The model yielded highly accurate regression and detection results, indicating its potential to reduce inconsistencies and repetitive tasks among clinicians during image processing. By incorporating images from pediatric and adolescent patients under various clinical conditions, the model demonstrates strong potential for integration into pediatric dental workflows. It may serve as a valuable tool in pediatric dentistry by streamlining image processing and supporting standardized diagnosis and documentation in clinical settings.

ABBREVIATIONS

AI, Artificial intelligence; AP, Average precision; CNN, Convolutional neural networks; FN, False negatives; FP, False positives; IoU, Intersection over Union; MAE, Mean absolute error; PR, Precision-recall; RMSE, Root mean squared error; TP, True positives; DSLR, digital single-lens reflex; ICC, intraclass correlation coefficient; MaxPool, Max pooling; AvgPool, Average pooling; Concat, Concatenation; ROI, Region of interest; RPN, Region proposal network; Bbox_pred, Bounding box prediction; Cls_prob, Class probability; Im_info, Image information; ROC, Receiver operating characteristic; AUC, Area Under the Curve; DL, deep-learning; R-CNN, Region-based Convolutional Neural Network; ISO, International Organization for Standardization; JPEG, Joint Photographic Experts Group; ID, Identifier; RAM, Random Access Memory.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are available on reasonable request from the corresponding author.

AUTHOR CONTRIBUTIONS

HS, TJ, EL and JS—designed the study. JE and HS—performed the research. JE, HS and SP—analyzed the data. JE—wrote the first draft of the manuscript. JS—provided critical feedback and approved the final manuscript. All authors contributed to this article. All authors commented on the previous versions of the manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Institutional Review Boards (IRB) of the Pusan National University Dental Hospital (IRB approval number: PNUDH 2025-02-010). Given the retrospective nature of the study, the Institutional Review Board of the Pusan National University Dental Hospital waived the requirement for written informed consent.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Li S, Guo Z, Lin J, Ying S. Artificial intelligence for classifying and archiving orthodontic images. BioMed Research International. 2022; 2022: 1473977.

[2] Ryu J, Kim YH, Kim TW, Jung SK. Evaluation of artificial intelligence model for crowding categorization and extraction diagnosis using intraoral photographs. Scientific Reports. 2023; 13: 5177.

[3] Ryu J, Lee YS, Mo SP, Lim K, Jung SK, Kim TW. Application of deep learning artificial intelligence technique to the classification of clinical orthodontic photos. BMC Oral Health. 2022; 22: 454.

[4] Schulz-Weidner N, Gruber M, Schraml EM, Wöstmann B, Krämer N, Schlenz MA. Improving the communication of dental findings in pediatric dentistry by using intraoral scans as a visual aid: a randomized clinical trial. Dentistry Journal. 2024; 12: 15.

[5] Alqadi A, O'Connell AC. Dental photography for children: a global survey. International Journal of Paediatric Dentistry. 2024; 34: 790–798.

[6] Selvakumar BG, Elavarasu PK, Vinodhini V, Ponnusamy P, Venkatachalapathy S, Abinaya R. Third eye of dentistry—digital dental photography: a literature review (part 1). Journal of Primary Care Dentistry and Oral Health. 2024; 5: 101–103.

[7] Al Omer H, Al Firdous RA, Al Shamrani SH, Al Anazi SS, Natto ZS, Al Dayel O. Assessment of dental photography imaging as a diagnostic tool for incipient pits and fissures caries in permanent posterior teeth. Edelweiss Applied Science and Technology. 2025; 9: 638–655.

[8] Ashtiani GH, Sabbagh S, Moradi S, Azimi S, Ravaghi V. Diagnostic accuracy of tele-dentistry in screening children for dental caries by community health workers in a lower-middle-income country. International Journal of Paediatric Dentistry. 2024; 34: 567–575.

[9] Werle SB, Piva F, Assunção C, Guimarães LF, de Araújo FB, Coelho-de-Souza FH. Photography in pediatric dentistry: basis and applications. Revista Odonto Ciencia. 2015; 30: 60–64.

[10] Sakr L, Abbas H, Thabet N, Abdelgawad F. Reliability of teledentistry mobile photos versus conventional clinical examination for dental caries diagnosis on occlusal surfaces in a group of school children: a diagnostic accuracy study. BMC Oral Health. 2025; 25: 545.

[11] Liu J, Chen Y, Li S, Zhao Z, Wu Z. Machine learning in orthodontics: challenges and perspectives. Advances in Clinical and Experimental Medicine. 2021; 30: 1065–1074.

[12] Kühnisch J, Meyer O, Hesenius M, Hickel R, Gruhn V. Caries detection on intraoral images using artificial intelligence. Journal of Dental Research. 2022; 101: 158–165.

[13] You W, Hao A, Li S, Wang Y, Xia B. Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. BMC Oral Health. 2020; 20: 141.

[14] Zhang X, Liang Y, Li W, Liu C, Gu D, Sun W, *et al*. Development and evaluation of deep learning for screening dental caries from oral photographs. Oral Diseases. 2022; 28: 173–181.

[15] Schwendicke F, Chaurasia A, Arsiwala L, Lee JH, Elhennawy K, Jost-Brinkmann PG, *et al*. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. Clinical Oral Investigations. 2021; 25: 4299–4309.

[16] Rodriguez R, Dokladalova E, Dokládal P. 'Rotation invariant CNN using scattering transform for image classification', 2019 IEEE international conference on image processing. Taipei, Taiwan, 22–25 September 2019. IEEE: Piscataway, USA. 2019.

[17] Liu S, Johns E, Davison AJ. 'End-to-end multi-task learning with attention', proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach, USA, 16–20 June 2019. IEEE: Piscataway, USA. 2019.

[18] Desai V, Bumb D. Digital dental photography: a contemporary revolution. International Journal of Clinical Pediatric Dentistry. 2013; 6: 193–196.

[19] Marcato L, Sandler J. The best choice of equipment to obtain high quality standardised results in intra-oral photography—a comparison between the common practice in the UK and the gold standard set by the literature. Journal of Visual Communication in Medicine. 2018; 41: 90–96.

[20] Kang S, Shon B, Park EY, Jeong S, Kim EK. Diagnostic accuracy of dental caries detection using ensemble techniques in deep learning with intraoral camera images. PLOS ONE. 2024; 19: e0310004.

[21] Girshick R. Fast R-CNN. 2015 IEEE international conference on computer vision. Santiago, Chile, 7–13 December 2015. IEEE: Piscataway, USA. 2015.

[22] Sumit SB, Joshi S, Rana U. Comprehensive review of R-CNN and its variant architectures. International Research Journal on Advanced Engineering Hub. 2024; 2: 959–966.

[23] Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. Geoscientific Model Development. 2014; 7: 1247–1250.

[24] Hodson TO. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development Discussions. 2022; 15: 5481–5487.

[25] Alhuwaish HA, Almoammar KA. Development and validation of an occlusal cant index. BMC Oral Health. 2022; 22: 127.

[26] Khan AS, Nagar P, Singh P, Bharti M. Changes in the sequence of eruption of permanent teeth; correlation between chronological and dental age and effects of body mass index of 5–15-year-old schoolchildren. International Journal of Clinical Pediatirc Dentistry. 2020; 13: 368–380.

[27] Pandey K, Ali M, Kumar Verma A, Ahmad N, Chaturvedi S, Deo K. Relating mandibular incisor to the lingual frenum in dentulous and edentulous (complete denture wearers) subjects: an *in vitro* study. British Journal of Medicine & Medical Research. 2016; 12: 1–8.

[28] Beuer F, Schweiger J, Edelhoff D. Digital dentistry: an overview of recent developments for CAD/CAM generated restorations. British Dental Journal. 2008; 204: 505–511.

[29] Alharbi SS, Alhasson HF. Exploring the applications of artificial

intelligence in dental image detection: a systematic review. Diagnostics. 2024; 14: 2442.

[30] Singh V, Pencina M, Einstein AJ, Liang JX, Berman DS, Slomka P. Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging. Scientific Reports. 2021; 11: 14490.

[31] Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. Journal of Dental Research. 2020; 99: 769–774.

[32] Hartman H, Nurdin D, Akbar S, Cahyanto A, Setiawan AS. Exploring the potential of artificial intelligence in paediatric dentistry: a systematic review on deep learning algorithms for dental anomaly detection. International Journal of Paediatric Dentistry. 2024; 34: 639–652.

[33] Badano A, Revie C, Casertano A, Cheng WC, Green P, Kimpe T, *et al*. Consistency and standardization of color in medical imaging: a consensus report. Journal of Digital Imaging. 2015; 28: 41–52.

[34] Elyan E, Vuttipittayamongkol P, Johnston P, Martin K, McPherson K, Moreno-García CF, *et al*. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. Artificial Intelligence Surgery. 2022; 2: 24–45.

[35] Habib AR, Xu Y, Bock K, Mohanty S, Sederholm T, Weeks WB, *et al*. Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy. Scientific Reports. 2023; 13: 5368.

[36] Zhang JW, Fan J, Zhao FB, Ma BE, Shen XQ, Geng YM. Diagnostic accuracy of artificial intelligence-assisted caries detection: a clinical evaluation. BMC Oral Health. 2024; 24: 1095.

[37] Cejudo Grano de Oro JE, Koch PJ, Krois J, Garcia Cantu Ros A, Patel J, Meyer-Lueckel H, *et al*. Hyperparameter tuning and automatic image augmentation for deep learning-based angle classification on intraoral photographs—a retrospective study. Diagnostics. 2022; 12: 1526.

How to cite this article: Jungmin Eum, Hyejun Seo, Taesung Jeong, Eungyung Lee, Soyoung Park, Jonghyun Shin. Automated processing of intraoral clinical photographs using deep learning techniques. Journal of Clinical Pediatric Dentistry. 2026; 50(1): 114-125. doi: 10.22514/jocpd.2026.011.