RESEARCH

A hierarchical deep learning approach for diagnosing impacted canine-induced root resorption via cone-beam computed tomography

Zeynab Pirayesh[1,2], Hossein Mohammad-Rahimi[1,3], Saeed Reza Motamedian[1,4,5], Sepehr Amini Afshar[3], Reza Abbasi[3], Mohammad Hossein Rohban[3], Mina Mahdian[6], Mitra Ghazizadeh Ahsaie[7] and Mina Iranparvar Alamdari[8*]

Abstract

Objectives Canine-induced root resorption (CIRR) is caused by impacted canines and CBCT images have shown to be more accurate in diagnosing CIRR than panoramic and periapical radiographs with the reported AUCs being 0.95, 0.49, and 0.57, respectively. The aim of this study was to use deep learning to automatically evaluate the diagnosis of CIRR in maxillary incisors using CBCT images.

Methods A total of 50 cone beam computed tomography (CBCT) images and 176 incisors were selected for the present study. The maxillary incisors were manually segmented and labeled from the CBCT images by two independent radiologists as either healthy or affected by root resorption induced by the impacted canines. We used five different strategies for training the model: (A) classification using 3D ResNet50 (Baseline), (B) classification of the segmented masks using the outcome of a 3D U-Net pretrained on the 3D MNIST, (C) training a 3D U-Net for the segmentation task and use its outputs for classification, (D) pretraining a 3D U-Net for the segmentation and transfer of the model, and (E) pretraining a 3D U-Net for the segmentation and fine-tuning the model with only the model encoder. The segmentation models were evaluated using the mean intersection over union (mIoU) and Dice coefficient (DSC). The classification models were evaluated in terms of classification accuracy, precision, recall, and F1 score.

Results The segmentation model achieved a mean intersection over union (mIoU) of 0.641 and a DSC of 0.901, indicating good performance in segmenting the tooth structures from the CBCT images. For the main classification task of detecting CIRR, Model C (classification of the segmented masks using 3D ResNet) and Model E (pretraining on segmentation followed by fine-tuning for classification) performed the best, both achieving 82% classification accuracy and 0.62 F1-scores on the test set. These results demonstrate the effectiveness of the proposed hierarchical, data-efficient deep learning approaches in improving the accuracy of automated CIRR diagnosis from limited CBCT data compared to the 3D ResNet baseline model.

*Correspondence:
Mina Iranparvar Alamdari
mina.iranparvar@gmail.com
Full list of author information is available at the end of the article

Conclusion The proposed approaches are effective at improving the accuracy of classification tasks and are helpful when the diagnosis is based on the volume and boundaries of an object. While the study demonstrated promising results, future studies with larger sample size are required to validate the effectiveness of the proposed method in enhancing the medical image classification tasks.

Keywords Deep learning, Cone beam computed tomography, Impacted tooth, Root resorption

Introduction

Aside from aesthetic and occlusion concerns, one of the most significant effects of an impacted canine is pressure on the adjacent tooth root, which disrupts the blood supply, leading to root resorption (RR) [1]. Canine-induced root resorption (CIRR) is a common condition that can affect various teeth in the maxillary arch. RR of maxillary lateral incisors is common (50%). Additionally, RR of mild severity is more common (62%), with resorption more frequently located in the middle (52%) and apical (42%) thirds of the root [2]. These findings highlight the variability and prevalence of CIRR in different parts of the maxillary arch and underscore the importance of accurate diagnostic methods. CIRR is an asymptomatic condition that, on rare occasions, may result in incisor or adjacent premolar loss [3]. Two-dimensional (2D) radiographs offer limited evaluation of CIRR due to the superimposition of structures and geometric distortion [4]. With the advancement of high-resolution cone beam computed tomography (CBCT), more accurate and 3D evaluations are possible, leading to increased detection of CIRR. Using CBCT to evaluate the extent and severity of RR can provide more accurate management from no treatment to root canal therapy and extraction [5]. Recent studies have highlighted the effectiveness of CBCT imaging over 2D radiography in detecting root resorption lesions [6]. However, clinicians' ability to appropriately detect the severity of these lesions can still vary greatly [7]. Despite the effectiveness of 3D imaging in diagnosis of CIRR, this highlights the persisting challenges in accurately diagnosing and evaluating CIRR, which calls for the creation of automated and objective diagnostic techniques utilizing artificial intelligence techniques (Figure 1).

Artificial intelligence (AI) has become increasingly prominent in dentistry, particularly for the segmentation of anatomical landmarks from cone-beam computed tomography (CBCT) images, which is essential for treatment planning and monitoring in orthodontics. AI automates the identification of landmarks such as the sella, nasion, and menton, thereby improving the precision of orthodontic assessments [8] and enabling better diagnosis [9] and treatment planning [10] of conditions such as CIRR. However, the application of deep learning in medical diagnosis faces challenges, including the lack of labeled data and the high dimensionality and complexity of CBCT images [11]. Data-efficient deep learning approaches can handle small datasets and extract meaningful features from images [12] without compromising their performance [13]. This can enable faster and more accurate diagnosis of CIRR. Compared to many other 3D medical image classification studies that have used datasets with hundreds of CT/CBCT images [14], we achieved high classification performance using only 50 CBCT images, demonstrating the potential of our hierarchical deep learning system to learn from limited labeled 3D data, which is particularly beneficial in fields like dentistry and medical imaging where data annotation is costly and time-consuming.

Previous studies have demonstrated the applicability of AI in segmenting impacted teeth, such as supernumerary maxillary teeth on panoramic images [15], mandibular third molars in both panoramic [16] and CBCT images [17], and impacted canines in both panoramic [16] and CBCT images [18]. Although several studies aim to segment teeth on CBCT images, CIRR is given less attention. Given the importance of early diagnosis of canine impaction and the high prevalence of CIRR, as well as the growing number of CBCTs indicated for this purpose, we conducted a study to propose an automated approach for diagnosing CIRR in maxillary incisors. The primary aim of this study is to develop and evaluate deep learning models for the automatic diagnosis of CIRR from CBCT images. Specifically, we compare various deep learning architectures to determine which model offers the best performance in terms of accuracy, precision, recall, specificity, and F1-score. The proposed methods have the potential to improve the accuracy and efficiency of root resorption diagnosis in clinical practice, enabling early detection and treatment planning. The outcome of the present study could constitute the first step in the development of a more complex system for orthodontic treatment planning in patients with impacted maxillary canines. Moreover, it can be extended to other dental and medical applications that require the analysis of complex 3D images with limited data.

Methods and materials
Study design
This study was designed as a diagnostic accuracy study aimed at evaluating the effectiveness of deep learning models in diagnosing CIRR using cone-beam computed tomography (CBCT) images. The study design follows the Standards for Reporting Diagnostic Accuracy Studies
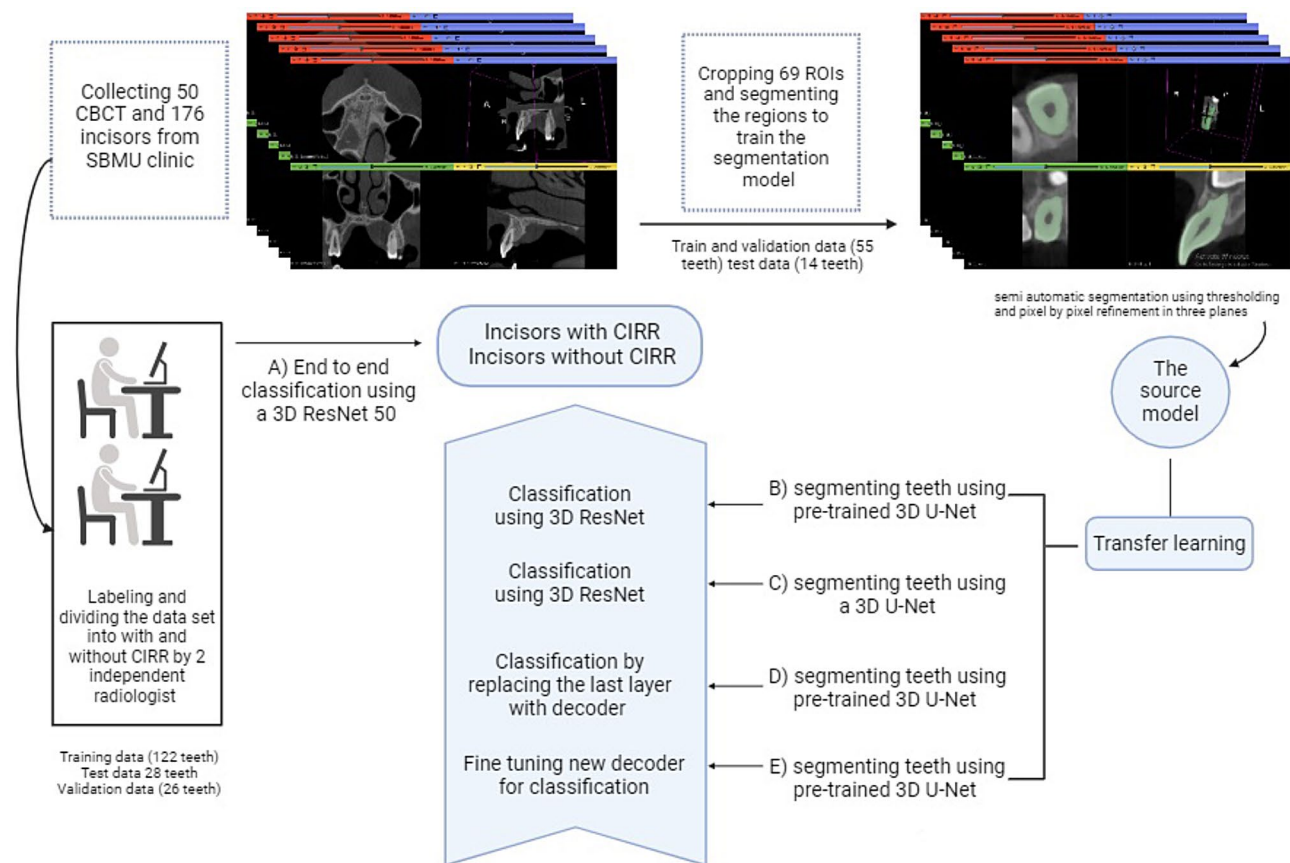
Figure. 1 Graphical abstract illustrating the study design and different deep learning approaches explored for diagnosing canine-induced root resorption from cone-beam computed tomography images (made in Biorender.com) []

(STARD) and Checklist for Artificial Intelligence in Medical Imaging.

(CLAIM) guidelines to ensure comprehensive and transparent reporting of diagnostic accuracy [19, 20].

In the present study, a novel framework based on the U-Net architecture using a two-stage training approach is proposed. We first pretrained our model by manually segmenting 3D volumes of the tooth structure. Subsequently, we fine-tuned the pretrained model for the final task, which was the detection of CIRR. We hypothesized that the model's prior knowledge of a simpler task (here, tooth segmentation) enhances its performance on the downstream task.

Dataset and data preparation

Images for this study were selected from the CBCT images of 50 patients visiting the dental clinic of Shahid Beheshti University of Medical Sciences between 2019 and 2021. Included data were from patients that referred for various dental and maxillofacial conditions particularly those with suspected impacted canines. Efforts were made to minimize bias by using a diverse dataset, which

included patients of varying ages and genders. The inclusion criteria were as follows:

1. Patients who had undergone CBCT imaging for the evaluation of impacted canines (with unilateral or bilateral impacted maxillary canines).
2. Patients who were diagnosed with either healthy incisors or CIRR.
3. Patients aged 15 years or older [4].

Exclusion criteria included:

1. Patients with previous orthodontic treatment that could affect the diagnosis of CIRR.
2. Patients with severe dental anomalies in the anterior maxillary sextant.
3. Scans with significant artifacts such as motion artifacts, beam hardening artifacts, or metal streak artifacts.
4. Scans with poor contrast or inadequate sharpness that could hinder the accurate identification of anatomical structures.

All patients who met the inclusion criteria and none of the exclusion criteria during the specified period were included in the study.

CBCT images were obtained using a NewTom VGi CBCT scanner (Verona, Italy). The following parameters were used for image acquisition: 110 kVp, 40 mAs, and 5.4 s exposure time, with an 8×8 cm field of view. We included both high-resolution images (with a voxel size of 150 μm) and standard-resolution images (with a voxel size of 300 μm).

We exported the CBCT images in the form of DICOM files. Open-source 3D Slicer software version 5.0.3 (https://www.slicer.org/) [21] was used to read and modify the DICOM files. Then, using 3D Slicer cropping features, we segmented the regions of interest (ROIs), which were the maxillary central and lateral incisors. The ROIs were exported in nrrd format for further steps. Each segmented ROI from CBCT images was analyzed independently and considered as a distinct unit. In other words our analysis is conducted at the tooth level, allowing for detailed assessment and characterization of each maxillary incisor.

Ground truth annotations

We used two independent annotation procedures for segmentation and classification tasks. During segmentation, the aim was to construct a segmentation mask of the hard tissue structure of the tooth. Two maxillofacial radiologists annotated the ROIs independently, and the final mask was the outcome of the intersection of two segmentation masks. The time taken for each segmentation was recorded using a stopwatch. Prior to segmentation, the radiologists were trained and calibrated in a joint meeting. For segmentation mask annotation, we used 3D Slicer. We first applied the thresholding function to eliminate the background and then used the painting feature to refine the voxel-by-voxel mask on the axial sections. The masks were verified on the coronal and sagittal planes for accuracy.

For the classification task, patients were divided into those with and without CIRR. Root resorption in locations not related to the impacted canine or due to other possible reasons, such as previous orthodontic treatments, was not considered a CIRR. The criterion for classifying samples as CIRR was a change in contour and outline at the apex of the tooth or on the lateral root surface compared to the normal root anatomy. Blunting and irregularities of incisors positioned in contact with the impacted canine were considered CIRR. Here, two independent maxillofacial radiologists annotated the samples. Annotation was based on raw CBCT images and previous reports in the university picture archiving and communication system. Any disagreements were resolved through consensus. In the event that a consensus was not reached, the case was excluded.

The ground truth was extracted from the same CBCT images used in the study. Specifically, the same CBCT images were processed with the deep learning (DL) algorithm for both segmentation and classification tasks. The time taken for the AI to segment each image was recorded using a stopwatch.

Data partitions

Internal validation was performed using a stratified sampling approach to ensure a balanced distribution of CIRR cases across training, validation, and test sets. In the segmentation task, we used 69 extracted ROIs (53 healthy and 16 resorbed teeth), 55 of which were used for training and validation sets. The other 14 were used for the test set.

In the classification task, 70% (n=122, 99 healthy and 23 resorbed samples) of the samples were used for the training set. Moreover, 15% (n=26, 21 healthy and 5 resorbed samples) and 15% (n=28, 22 healthy and 6 resorbed samples) of the ROIs were selected for the validation set and test set, respectively. The sampling strategy for each set was stratified sampling, where the ratio of patients with and without CIRR was approximately similar. The validation set was used for hyperparameter tuning, while the test set was used for reporting the model outcome on unseen data. Performance metrics such as mean intersection over union (mIoU) and Dice coefficient (DSC) for segmentation, and accuracy, precision, recall, and F1 score for classification, were computed for the evaluation.

Data preprocessing and augmentation

For all the training procedures, all images were initially resized to 16*16*16 voxels, converted to PyTorch tensor, and normalized with a 0 mean and 1 variance. Then, a combination of random augmentations was performed. This random augmentation was selected from a set of five options with the same probability. The five options were random flip, random elastic deformation, random anisotropy, random affine and no augmentation. All the noted augmentation implementations used default parameters. For classification, using data augmentation approaches, we increased the number of training samples from 122 to 2980. Furthermore, to address the imbalanced dataset, oversampling the class with a lower sample was achieved by augmenting.

Model architecture and training details

The fundamentals of the implemented architectures are available in the supplementary material.

The index test involved the use of deep learning models to automatically diagnose CIRR.

For this reason, we used the following training strategies and model structure:

Model A (Baseline)   We used 3D ResNet50 for end-to-end classification. Here, we used the pretrained weights of the MedicalNet framework proposed by Chen et al. [22], which were pretrained on 23 medical datasets. It has been reported that it overperformed the randomly initialized model in medical image segmentation tasks by a large margin. We froze the model's weights and trained two additional fully connected layers.

Model B   We used a 3D U-Net pretrained on the 3D MNIST dataset. Then, we passed the nonaugmented data through the model. Then, we used segmented masks for classification using 3D ResNet.

Model C   We trained a 3D U-Net for the segmentation task using our data. Then, we used the output of the model (3D masks) for classification using 3D ResNet.

Model D   We pretrained a 3D U-Net for the segmentation task using our data. Then, after replacing the last layer with fully connected layers, we fine-tune the model for our downstream classification task.

Model E   We pretrained a 3D U-Net for the segmentation task using our data. Then, after replacing the model's decoder with a new randomly initialized decoder, we fine-tune the model for our downstream classification task.

For models C, D, and E, we trained a single randomly initialized 3D U-Net and used it in different mentioned approaches. The training was performed on a Tesla K80 and Tesla T4 graphic processor unit (Nvidia Corporation, Santa Clara, CA, USA) through the Google Collaboratory platform. Hyperparameter tuning was performed based on a randomized search strategy. For training the 3D U-Net model, the learning rate, weight decay, batch size, and number of epochs were set to $2*10^{-4}$, 10^{-3}, 13 and 50, respectively. Dice loss was used with the Adam optimizer to perform the segmentation tasks. For the classification models, we used different hyperparameters for each approach, which are presented in Table 1. As

there are different models with different learning rates and batch sizes and some of them use pretrained weights, the number of epochs varies accordingly. We used binary cross-entropy loss with the Adam optimizer for all the classification tasks.

Evaluation

For the segmentation task evaluation (which was used for models C, D, and E), the mean intersection over union (mIoU) and DSC were reported. These metrics were defined as follows:

$$mIoU = \frac{Area\ of\ Intersection\ (model\ outcome\ and\ GT)}{Area\ of\ Union}$$

$$DSC = \frac{2*\ Area\ of\ Intersection}{Sum\ of\ Areas\ (model\ outcome\ and\ GT)}$$

For the classification task evaluation, the classification accuracy, precision, recall (sensitivity), and F1-score were reported. These metrics were defined as follows:

$$Classification\ accuracy = \frac{TP + TN}{\#All\ samples}$$

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

$$Sensitivity/Recall = \frac{TP}{TP + FN}$$

$$Specificty = \frac{TN}{TN + FP}$$

$$F1 - score = \frac{2\ .\ Precision\ .Recall}{Precision\ +\ Recall}$$

TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. Moreover, we reported the confusion matrix of each model's output in the test set. Since our test set was imbalanced, the criteria for selecting the best model were based on the F1 score.

Statistical analysis

The interobserver agreement for the manual segmentation of the ROIs and annotation of root resorption were analyzed to ensure consistency across the raters. The intraclass correlation coefficients (ICCs) were calculated to assess the reproducibility of the ratings provided by the two independent maxillofacial radiologists. Kohen's Kappa was calculated to measure the agreement between the two raters. All statistical analyses were performed using Python version 3.7.

Table 1 Final hyperparameters set for training various models

Hyperparameters	Model A	Model B	Model C	Model D	Model E
Learning rate	$10-4$	$10-3$	$2*10-3$	$10-5$	0.01
Total number of epochs	100	100	100	50	50
Batch size	128	128	128	128	128
Weight decay	$2*10-3$	0.05	$10-3$	$2*10-3$	0.05

Table 2 Various models' outcomes

	Accuracy	Precision	Recall	Specificity	F1-score
Model A (Baseline)	75.0% (95% CI: 68–82)	0.4 (95% CI: 0.3–0.5)	0.3 (95% CI: 0.3–0.4)	0.9 (95% CI: 0.8–0.9)	0.4 (95% CI: 0.3–0.4)
Model B	78.6% (95% CI: 72–85)	0.5 (95% CI: 0.4–0.6)	0.5 (95% CI: 0.4–0.6)	0.9 (95% CI: 0.8–0.9)	0.5 (95% CI: 0.4–0.6)
Model C	**82.1%** (95% CI: 76–88)	**0.6** (95% CI: 0.5–0.6)	**0.7** (95% CI: 0.6–0.7)	**0.9** (95% CI: 0.8–0.9)	**0.6** (95% CI: 0.5–0.7)
Model D	78.6% (95% CI: 72–85)	0.5 (95% CI: 0.4–0.6)	0.7 (95% CI: 0.6–0.7)	0.8 (95% CI: 0.8–0.9)	0.6 (95% CI: 0.5–0.6)
Model E	**82.1%** (95% CI: 76–88)	**0.6** (95% CI: 0.5–0.6)	**0.7** (95% CI: 0.6–0.7)	**0.9** (95% CI: 0.8–0.9)	**0.6** (95% CI: 0.5–0.7)

Ethics

All patient data was anonymized and processed in compliance with relevant data protection regulations to safeguard patient privacy. Adhering to recent guidelines for AI in dental research, which emphasize the importance of transparency, fairness, and accountability in AI applications [23], we ensured our models were developed in accordance. Transparency and explainability were prioritized to ensure clinicians can understand the decision-making processes of the AI models. The Shahid Beheshti University Medical Sciences Ethics Committee approved this study (IR.SBMU.DRC.REC.1401.044). The study was carried out in line with the principles of Declaration of Helsinki.

Results

Data

CBCT images of 50 patients (19 males, 31 females) were selected for this study. Their average age was 21.18 ± 12.06 years. Three patients had orthodontic devices. Thirty-five of the CBCTs had high resolution. A total of 176 ROIs were extracted, which included 34 resorbed teeth and 142 healthy teeth. The mean value for ICC was 0.9847 (± 0.0035), reflecting excellent reproducibility of tooth segmentation. Interobserver agreement between the two radiologists was high (K= 0.9014 ± 0.2130)

Model performance

The average time taken for AI-based segmentation was 7−30 milliseconds per image while the average time taken for manual segmentation, including loading 3D Slicer and obtaining the final output by experienced maxillofacial radiologists, was 11−12 min per image.

In Table 2, we present a summary of the performance of the different approaches on the test set. In all the cases, the experimental models outperformed the baseline model. Model C and Model E produced the best results among the experiments, with 82% classification accuracy (95% CI: 76 −88%) and an F1 score of 0.6 (95% CI: 0.5–0.7).

Figure 2 illustrates various models' confusion matrices when evaluated on the test set. As a pretraining model for Model C, Model D, and Model E, our segmentation model achieved an mIoU of 0.6 (95% CI: 0.6–0.7) and a DSC of 0.9 (95% CI: 0.9–0.9). Figure 3 presents examples
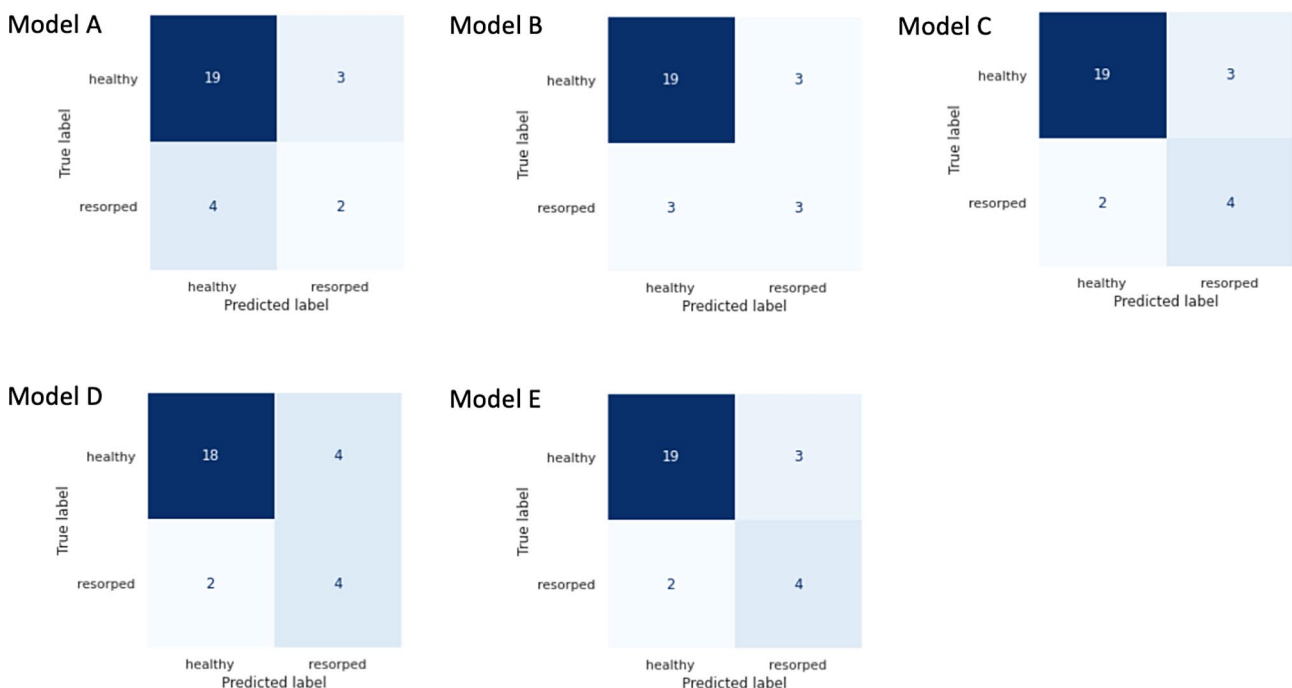
Figure. 2 Confusion matrix of various models' performance on the test set
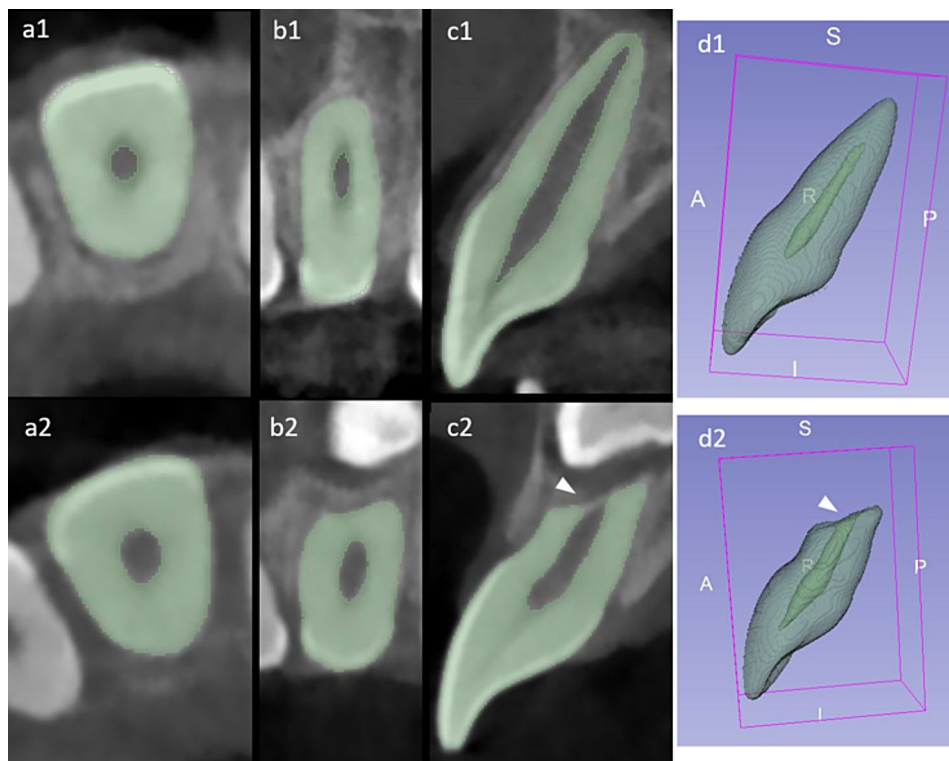
Figure. 3 Segmentation model output. *Note* the white arrowhead indicating apical CIRR in the right central incisor. a1, a2. Axial view; b1, b2 Coronal view; c1, c2 sagittal view and d1, d2 3D rendered view

of 3D U-Net results, which demonstrate the ability of our models to accurately segment the data.

Discussion

Early detection of CIRR is imperative for successful orthodontic treatment. Both clinical and radiographic examinations are essential in diagnosing impacted canines [24]. The diagnostic process typically begins with a clinical examination and palpation of the alveolar bone, followed by radiographic evaluation [25]. Traditionally, radiographic assessments of IMCs have utilized two-dimensional (2D) imaging techniques, such as intraoral periapical, occlusal X-rays, panoramic, and cephalometric radiographs [26]. However, these conventional X-rays are often limited by their low diagnostic accuracy due to factors such as image distortion, magnification, blurring, and the superimposition of different anatomical structures [27]. Consequently, the adoption of three-dimensional (3D) imaging, specifically CBCT, has gained popularity for evaluating the maxillofacial region, as it offers superior diagnostic accuracy and detailed visualization of anatomical structures [27]. While offering significant advantages, CBCT exposes patients to a higher radiation dose compared to 2D imaging techniques like panoramic and intraoral radiographs [28]. In spite of low-dose protocols that maintain sufficient image quality to reduce patient dose, effective doses of CBCT remain

higher than 2D radiographies [29]. The increased exposure must be justified by its clinical benefits to detect and evaluate CIRR. CBCT is the preferred 3D imaging modality for evaluating impacted canines and CIRR, with a 63% higher detection rate than conventional imaging. However, CBCT scan analysis requires sufficient time, skill, and specialized training [30, 31]. The present study aimed to introduce an automated approach using deep learning for diagnosing CIRR in maxillary incisors through CBCT images.

The results of our study indicate that the experimental models outperformed the baseline model in terms of classification accuracy and F1-score. Specifically, models C (classification of the segmented masks) and E (model pretraining for segmentation and then fine-tuning classification model) exhibited the highest performance, achieving a 82% classification accuracy and a F1 score of 0.62. These findings suggest that our proposed approaches are effective in improving the accuracy of the classification task. Furthermore, we evaluated the performance of our segmentation model, which was used as a pretraining model. Our segmentation model achieved good performance, suggesting that it could accurately segment tooth structures and extract meaningful features to improve the classification models. Our approach was able to automatically segment incisors in 0.07–0.3 s which was approximately 37,000 times faster

Pirayesh *et al. BMC Oral Health*   (2024) 24:982

Page 8 of 11

than the manual segmentation, highlighting its potential to enhance workflow efficiency in clinical settings.

Since 3D image segmentation can be conducted using less data than can end-to-end 3D image classification [32], we first pretrained our model for segmenting tooth volume. Then, we fine-tuned this pretrained model for our final task, CIRR classification. This approach succeeded in outperforming our baseline, with an accuracy of 82.14%. The suggested framework is also inspired by a clinician's diagnostic process, which first looks for the volume and boundaries of the tooth and determines the presence of the CIRR. Such approaches have been used in the case of other medical imaging problems [33]. Our outcomes showed that this hierarchical approach is helpful in conditions where the diagnosis is based on the volume and boundaries of an object. Another successful approach was based on the classification of segmented masks rather than raw images. The tooth volumes were first segmented. Then, only the segmented masks were fed to the second model. As with the previous approach, this exceeded the baseline accuracy by 7.14%. This is because CIRR detection relies solely on tooth volume, not texture or color. Additionally, classifying masks enables the algorithm to aggregate global visual information for instancewise classification.

Regarding dental conditions similar to CIRR, where obtaining high-quality, annotated data faces considerable difficulties, a number of studies have investigated various methodological strategies. For example, a particularly relevant study by Mohammad-Rahimi et al. [34] investigated the efficacy of using label-efficient self-supervised learning (SSL) for detecting external cervical resorption on periapical radiographs. They trained and compared several SSL models, such as DINO, MoCo v2, and BYOL, with transfer learning baselines. The SSL models showed improved performance over the baselines, with DINO achieving 85.64% mean accuracy.

In another study by J. Huang et al. [35], the use of active learning techniques for multilabel segmentation and periapical lesion identification in CBCT volumes was investigated. These techniques rely on uncertainty quantification using a Bayesian U-Net. A higher lesion detection sensitivity of up to 84% was attained by the active learning techniques employing functions such as BALD and Max_Entropy, which outperform the nonactive learning baseline. These studies show how label-efficient methods, such as SSL and active learning, may improve model performance when working with small amounts of labeled dental imaging data. A further important aspect of our research on identifying CIRR is the ability to distinguish between conditions such as external root resorption and periapical lesions. A distinct hierarchical deep learning strategy was used in our study that leveraged segmentation pretraining transfer learning. However,

investigating SSL and active learning pretraining strategies could improve our models' performance even more. With respect to self-supervised learning (SSL) and active learning (AL), our study's customized deep learning method utilizing architectures such as U-Net and ResNet has several potential benefits, particularly given the unique goal of detecting CIRR from limited CBCT data:

The method suggested in this work offers a more explicit and interpretable way of leveraging limited annotated data by using hierarchical learning and pretraining on a relatively simple task (segmentation). Explicit knowledge transfer by learning relevant features about tooth structures during segmentation pretraining, which is directly applicable to the CIRR classification task, is more effective than implicit knowledge transfer in SSL. Compared to unsupervised pretraining in SSL, supervised pretraining on a smaller labeled segmentation dataset can provide more significant, task-specific features.

Additionally, fine-tuning the pretrained model on the final complex task of classification of the CIRR improves the model's performance on limited data. While SSL and active learning approaches have their own benefits, such as the ability to leverage unlabeled data or selectively annotate informative samples, these methods are not as beneficial as the method here. However, it should be noted that the quality and consistency of the annotations, the complexity of the underlying problem, and how the related and final tasks are connected can all impact how effective the suggested approach is. Additionally, the method is adaptable and may be used for other dental or medical imaging tasks that are based on a particular anatomical structure's volume, form, or borders.

A model can only be utilized in clinical practice if it has gained clinicians' trust through accuracy and interpretability. While SSL and AL can reduce the amount of labeled data needed, they often require iterative labeling and retraining processes, making them time-consuming and resource-intensive, and involve complex, less transparent decision-making processes. In contrast, our model can be trained in a more straightforward manner using fully labeled datasets without the long pretraining duration required by SSL. This direct application of domain-specific knowledge might at times provide a faster route to high performance. The interpretability of the models used in our work, such as U-Net and ResNet, needs careful consideration. Although these models are highly effective in classification and segmentation tasks, they often function as "black boxes," making their decision-making processes less transparent. Clearer insights into how decisions are made can be achieved with interpretability techniques such as attention mechanisms and saliency maps, which future work might investigate to enhance model transparency and clinical trust.

However, no prior work on using deep learning to diagnose CIRR in maxillary incisors with CBCT images has been reported. However, a variety of deep learning approaches have been used to diagnose dental conditions through CBCT. In a study by Setzer et al. [36] in 2020, they proposed deep learning to detect apical periodontitis in CBCT images of mandibular molars. The study reported an accuracy of 91.3% and a sensitivity of 90.3%. Although they achieved better results than did the current study in an apical periodontitis study, it is important to note that CIRR is a more challenging diagnostic task, as it might occur as faint changes in the tooth contour without obvious clinical symptoms. Reduwan et al. [37] aimed at identifying and classification of external root resorption (ERR) on 88 CBCT scans of extracted premolars, which may not fully represent the complexity of clinical cases. In contrast with our study, the ground truth was provided by only one radiologist which is prone to bias and effects the generalizability of the results. They implemented different classification models such as RF+VGG and RF+EFNET and feature selection techniques with deep learning to enhance model performance. The study reported the overall accuracy, precision, and F1-score of 81% for the best performing model which was a combination of FS+RF+VGG models. However, no values demonstrating consistency and reliability of the ground truth annotations were reported. In another study by Li et al. [38] multiple roots were segmented automatically in a single CBCT scan using U-net with Attention Gates (AGs) and Recurrent Neural Networks (RNNs). They achieved an IoU of 0.9 and DSC of 0.9 for roots in the maxilla, which is comparable to our results. It is important to note that they did not include any resorbed roots, which are more challenging to segment due to the irregular and subtle nature of resorptive changes. Separating the results of segmenting normal (mIoU=0.65) and resorbed teeth (mIoU=0.54) highlighted the increased complexity of segmenting resorbed dental structures. In a recent study by Su et al. [39] alveolar bone and teeth were segmented in 4–5 random slices of CBCTs from 389 patients. 1784 2D slices were labelled by 4 clinicians and the overlap of the identified areas was considered as PDL space. A mask RCNN with ResNet50 backbone was used and reached IoU and DSC values of 0.8 and 0.9 for tooth segmentation and a segmentation accuracy of 100% for incisors. Although they achieved higher performance, their approach involved training 2D neural networks on a few slices per CBCT, which might be due to the laborious nature of drawing polygons around anatomic structures. In contrast, we annotated each tooth in every slice in all three planes (axial, sagittal, and coronal) and used thresholding to streamline the annotation process.

Limitations and future studies

As with any study, the proposed deep learning approach for diagnosing CIRR in maxillary incisors through CBCT images has limitations. One limitation is the size of the dataset used to train the deep learning models. The primary hypothesis of this work was that, even with a small sample size, the proposed hierarchical deep learning approach would accurately identify CIRR in maxillary incisors from CBCT images. To validate this, a post hoc power analysis using Python [40] and scipy library [41] was conducted to evaluate the sample size and its sufficiency to train and evaluate the performance of the proposed models. The analysis indicated that the test set was smaller than needed for optimal statistical power. This limitation was primarily due to the limited accessible data, especially in patients suspected of having impacted canines and root resorption. We employed several approaches to overcome the limited number of data samples. Our first approach was based on the assumption that we can transfer the knowledge of our model doing more straightforward tasks (here, segmentation) to final downstream tasks [42]. Although we employed several strategies to maximize the use of available data through techniques such as data augmentation and transfer learning, this constraint could affect the generalizability of the models to other datasets or populations.

Another limitation is the requirement for expert annotation of 3D segmentation masks, which can be time-consuming and labor-intensive. In addition, the quality of the segmentation masks could vary depending on the expertise of the annotators and consistency in applying the defined criteria for determining the segmentation boundaries. Furthermore, the proposed deep learning approach was tested only on CBCT images of maxillary incisors with CIRR. Future studies should evaluate the model's performance on larger, more diverse datasets, including different tooth types and locations. While our proposed method showed encouraging results in diagnosing CIRR, a more thorough assessment of the model's clinical usefulness would come from comparing it directly to radiologists using the same dataset. Further study should include a comparison involving two radiologists independently review the same CBCT images and provide a diagnosis for CIRR. Similar performance metrics ought to be calculated for the radiologists' performance to be compared with the model's.

Conclusion

While our study demonstrated promising results with an accuracy of over 80%, it is important to note that the F1 score suggests the model may still have biases or encounter difficulties with specific tasks. To the best of our knowledge, this paper is the first to develop a deep learning model for detecting CIRR using CBCT images. The

initial findings highlight the need for further research to investigate how to address the challenges posed by similar classes and to optimize the performance of our models in real-world settings. Future work need to focus on refining the model to enhance its robustness and applicability in diverse clinical scenarios.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12903-024-04718-4.

> Supplementary Material 1

Declarations

Ethics approval and consent to participate
The Shahid Beheshti University Medical Sciences Ethics Committee approved this study (IR.SBMU.DRC.REC.1401.044).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
¹Topic Group Dental Diagnostics and Digital Dentistry, ITU/WHO Focus Group AI on Health, Berlin, Germany
²Department of Orthodontics and Dentofacial Orthopedics, School of Dentistry, Zanjan University of Medical Sciences, Zanjan, Iran
³Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
⁴Dentofacial Deformities Research Center, Research Institute of Dental Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
⁵Department of Orthodontics, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran
⁶Division of Diagnostic Imaging, Department of Prosthodontics and Digital Technology, Stony Brook University School of Dental Medicine, Stony Brook, NY, USA
⁷Department of Electrical and Computer Engineering, California State University, Chico 95929, USA
⁸Department of Oral and Maxillofacial Radiology, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

References
1. Patel S, Saberi N. The ins and outs of root resorption. Br Dent J. 2018;224(9):691–9.
2. Mitsea A, Palikaraki G, Karamesinis K, Vastardis H, Gizani S, Sifakakis I. Evaluation of lateral incisor resorption caused by impacted Maxillary canines based on CBCT: a systematic review and Meta-analysis. Child (Basel). 2022;9(7).
3. Grisar K, Piccart F, Al-Rimawi AS, Basso I, Politis C, Jacobs R. Three-dimensional position of impacted maxillary canines: prevalence, associated pathology and introduction to a new classification system. Clin Experimental Dent Res. 2019;5(1):19–25.
4. Sunil G, Ranganayakulu L, Ranghu Ram R. Maxillary canine impaction-A hitch in orthodontic treatment planning. IAIM. 2018;5(6):72–6.
5. Liu M-Q, Xu Z-N, Mao W-Y, Li Y, Zhang X-H, Bai H-L, et al. Deep learning-based evaluation of the relationship between mandibular third molar and mandibular canal on CBCT. Clin Oral Invest. 2022;26(1):981–91.
6. Peralta-Mamani M, Rubira CM, López-López J, Honório HM, Rubira-Bullen IR. CBCT vs panoramic radiography in assessment of impacted upper canine and root resorption of the adjacent teeth: a systematic review and meta-analysis. J Clin Exp Dent. 2024;16(2):e198–222.
7. Alqerban A, Jacobs R, Fieuws S, Nackaerts O, Willems G, Consortium SP. Comparison of 6 cone-beam computed tomography systems for image quality and detection of simulated canine impaction-induced external root resorption in maxillary lateral incisors. Am J Orthod Dentofac Orthop. 2011;140(3):e129–39.
8. Kazimierczak N, Kazimierczak W, Serafin Z, Nowicki P, Nożewski J, Janiszewska-Olszowska J. AI in Orthodontics: Revolutionizing Diagnostics and Treatment Planning—A Comprehensive Review. J Clin Med. 2024;13(2):344.
9. Estrela C, Bueno MR, Leles CR, Azevedo B, Azevedo JR. Accuracy of cone beam computed tomography and panoramic and periapical radiography for detection of apical periodontitis. J Endod. 2008;34(3):273–9.
10. Deng Y, Sun Y, Xu T. Evaluation of root resorption after comprehensive orthodontic treatment using cone beam computed tomography (CBCT): a meta-analysis. BMC Oral Health. 2018;18:1–14.
11. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.
12. Madani A, Ong JR, Tibrewal A, Mofrad MRK. Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. Npj Digit Med. 2018;1(1):59.
13. Abdou MA. Literature review: efficient deep neural networks techniques for medical image analysis. Neural Comput Appl. 2022;34(8):5791–812.
14. Esmaeilyfard R, Bonyadifard H, Paknahad M. Dental Caries detection and classification in CBCT images using deep learning. Int Dent J. 2023.
15. Kuwada C, Ariji Y, Fukuda M, Kise Y, Fujita H, Katsumata A et al. Deep learning systems for detecting and classifying the presence of impacted supernumerary teeth in the maxillary incisor region on panoramic radiographs. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology. 2020;130(4):464-9.
16. Imak A, Çelebi A, Polat O, Türkoğlu M, Şengür A. ResMIBCU-Net: an encoder–decoder network with residual blocks, modified inverted residual block, and bi-directional ConvLSTM for impacted tooth segmentation in panoramic X-ray images. Oral Radiol. 2023;39(4):614–28.
17. Orhan K, Bilgir E, Bayrakdar IS, Ezhov M, Gusarev M, Shumilov E. Evaluation of artificial intelligence for detecting impacted third molars on cone-beam computed tomography scans. J Stomatology Oral Maxillofacial Surg. 2021;122(4):333–7.
18. Swaity A, Elgarba BM, Morgan N, Ali S, Shujaat S, Borsci E, et al. Deep learning driven segmentation of maxillary impacted canine on cone beam computed tomography images. Sci Rep. 2024;14(1):369.

19. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ open. 2016;6(11):e012799.

20. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiology Artificial Intelligence. 2020;2(2).

21. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D slicer as an image computing platform for the quantitative Imaging Network. Magn Reson Imaging. 2012;30(9):1323–41.

22. Chen S, Ma K, Zheng Y. Med3d: transfer learning for 3d medical image analysis. arXiv Preprint arXiv:190400625. 2019.

23. Rokhshad R, Ducret M, Chaurasia A, Karteva T, Radenkovic M, Roganovic J, et al. Ethical considerations on artificial intelligence in dentistry: a framework and checklist. J Dent. 2023;135:104593.

24. Dekel E, Nucci L, Weill T, Flores-Mir C, Becker A, Perillo L, et al. Impaction of maxillary canines and its effect on the position of adjacent teeth and canine development: a cone-beam computed tomography study. Am J Orthod Dentofac Orthop. 2021;159(2):e135–47.

25. Hajeer MY, Al-Homsi HK, Murad RM. Evaluation of the diagnostic accuracy of CBCT-based interpretations of maxillary impacted canines compared to those of conventional radiography: an in vitro study. Int Orthod. 2022;20(2):100639.

26. Salari B, Tofangchiha M, Padisar P, Reda R, Zanza A, Testarelli L. Diagnostic accuracy of conventional orthodontic radiographic modalities and cone-beam computed tomography for localization of impacted maxillary canine teeth. Sci Prog. 2024;107(1):00368504241228077.

27. Eslami E, Barkhordar H, Abramovitch K, Kim J, Masoud MI. Cone-beam computed tomography vs conventional radiography in visualization of maxillary impacted-canine localization: a systematic review of comparative studies. Am J Orthod Dentofac Orthop. 2017;151(2):248–58.

28. Kadesjö N, Lynds R, Nilsson M, Shi X-Q. Radiation dose from X-ray examinations of impacted canines: cone beam CT vs two-dimensional imaging. Dentomaxillofacial Radiol. 2018;47(3):20170305.

29. Andresen AK, Jonsson MV, Sulo G, Thelen DS, Shi X-Q. Radiographic features in 2D imaging as predictors for justified CBCT examinations of canine-induced root resorption. Dentomaxillofacial Radiol. 2022;51(1):20210165.

30. Becker A, Chaushu S. Etiology of maxillary canine impaction: a review. Am J Orthod Dentofac Orthop. 2015;148(4):557–67.

31. Albaker BK, Wong RW. Diagnosis and management of root resorption by erupting canines using cone-beam computed tomography and fixed palatal appliance: a case report. J Med Case Rep. 2010;4:399.

32. Sehar U, Naseem ML. How deep learning is empowering semantic segmentation. Multimedia Tools Appl. 2022;81(21):30519–44.

33. An G, Akiba M, Omodaka K, Nakazawa T, Yokota H. Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images. Sci Rep. 2021;11(1):4250.

34. Mohammad-Rahimi H, Dianat O, Abbasi R, Zahedrozegar S, Ashkan A, Motamedian SR, et al. Artificial intelligence for detection of external cervical resorption using label-efficient self-supervised learning method. J Endod. 2024;50(2):144–53. e2.

35. Huang J, Farpour N, Yang BJ, Mupparapu M, Lure F, Li J, et al. Uncertainty-based active learning by bayesian U-Net for multi-label cone-beam CT segmentation. J Endod. 2024;50(2):220–8.

36. Setzer FC, Shi KJ, Zhang Z, Yan H, Yoon H, Mupparapu M, et al. Artificial Intelligence for the computer-aided detection of Periapical Lesions in Cone-Beam Computed Tomographic images. J Endod. 2020;46(7):987–93.

37. Reduwan NH, Abdul Aziz AA, Mohd Razi R, Abdullah ERMF, Mazloom Nezhad SM, Gohain M, et al. Application of deep learning and feature selection technique on external root resorption identification on CBCT images. BMC Oral Health. 2024;24(1):252.

38. Li Q, Chen K, Han L, Zhuang Y, Li J, Lin J. Automatic tooth roots segmentation of cone beam computed tomography image sequences using U-net and RNN. J X-Ray Sci Technol. 2020;28(5):905–22.

39. Su S, Jia X, Zhan L, Gao S, Zhang Q, Huang X. Automatic tooth periodontal ligament segmentation of cone beam computed tomography based on instance segmentation network. Heliyon. 2024;10(2).

40. Foundation PS. Python (Version 1.11.4) 2023 [Software]. https://www.python.org/

41. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72.

42. Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X. Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med Image Anal. 2020;63:101693.

Publisher's Note