scientific reports

Check for updates

OPEN

End-to-end CNN-based detection of permanent first molars and prediction of root development stages from panoramic radiographs

Sukriye Turkoglu Kayaci[1]✉, Hamza Osman Ilhan[2]✉, Gorkem Serbes[3] & Hakan Arslan[4]

The aim of this study was to develop a convolutional neural network (CNN)-based end-to-end learning architecture to predict the root development stages of permanent first molar teeth using panoramic radiographs. A dataset of 1629 first molar images was labeled according to the Cvek classification and organized into five subsets (DB-1 to DB-5) based on root development stages and apical foramen status. Teeth patches were cropped using the YOLO approach, and stage prediction was performed with VGG-19, InceptionV3, and EfficientNet-B3 models optimized with the Adamax optimizer at a learning rate of 10^{-3}. The proposed method achieved high precision (98.4%) and recall (97.6%) in detecting first molar teeth. Classification performance reached average accuracies of 64.21% for DB-1, 62.66% for DB-2, and 69.64% for DB-3. For apical foramina classification, an accuracy of 84.57% was obtained in DB-4, which further improved to 94.89% in DB-5. These findings highlight the potential of CNN-based approaches in dental diagnostics, providing clinicians with an effective tool for assessing root development and supporting treatment planning.

Keywords Deep learning in dentistry, YOLO algorithm, Transfer learning, Permanent first molar, Regenerative endodontics, Root development stages

Dental caries, also known as tooth decay, is one of the most prevalent chronic diseases affecting both children and adults[1]. Specifically, The Global Burden of Disease study highlights that untreated dental caries in permanent dentition is recognized as most prevalent health issues[2]. This can lead to conditions such as pain, infection, and tooth loss, resulting in a decline in quality of life[3]. Dental diseases not only give rise to significant oral complications but also exert detrimental effects on other organ systems such as increased risk of coronary heart disease, myocardial infarction, cerebrovascular disease, and a higher risk of ischemic, hemorrhagic stroke, and cerebral ischemia[4].

Permanent first molars are highly susceptible to tooth decay and early destruction, as they are the first molars to erupt during permanent dentition and often haven't completed their maturation at that time[5,6]. Furthermore, due to their vulnerability to hypomineralization, these teeth are frequently affected by caries and pathologies that reach the pulp. It is of utmost importance to treat these teeth with proper planning[7,8]. There have been significant advancements in pulp treatments, particularly in regenerative endodontic therapies, which show promising potential. The crucial aspect lies in the application of appropriate treatment for each specific tooth. One of the most important factors determining the indication and course of treatment is whether the teeth have an open or closed apex[9,10]. In this context, the formulation of an effective treatment plan is crucial for achieving clinical success[11]. Different treatments can be applied to teeth based on the open apex/closed apex and root development levels. Essentially, whether the apex of the tooth is open or closed determines the treatment protocol. Additionally, the level of apical openness and concurrent root development significantly influence treatment planning and prognosis[12,13]. The stage of root development plays a significant role in treatment planning decisions, such as selecting an apical barrier technique using tricalcium silicate material or opting for a revascularization procedure. This aspect holds particular importance due to its potential impact on treatment

[1]Department of Pediatric Dentistry, Hamidiye Faculty of Dental Medicine, University of Health Sciences, Istanbul, Turkey. [2]Department of Computer Engineering, Faculty Of Electrical And Electronics Engineering, Yildiz Technical University, Istanbul, Turkey. [3]Department of Biomedical Engineering, Faculty Of Electrical And Electronics Engineering, Yildiz Technical University, Istanbul, Turkey. [4]Department of Endodontics, Faculty of Dentistry, Istanbul Medeniyet University, Istanbul, Turkey. ✉email: sukriyeturkoglu1@gmail.com; hoilhan@yildiz.edu.tr

outcomes and prognosis[14]. Regarding the apical foramen, which is the limitation for optical pathways of stem cell migration and rush of blood, it has been suggested that the success rate of pulpal healing was significantly reduced, becoming unpredictable when the diameter of the apical foramen was less than 1 mm. However, successful cases of revascularization have been reported even with a diameter as small as 1.0 mm[15].

While panoramic radiographs (PRs) offer significant advantages in dental treatments, their subjective interpretation can introduce challenges. Interpretations in the field of biomedical sciences can be influenced by perceptual biases or factors during the observation process, leading to potential inaccurities[16]. Irrespective of their specific discriminatory capacity, all radiographic examinations are susceptible to limited inter- and intra-examiner reliability. The reliability in question is contingent upon the examiner's level of experience[17,18]. Therefore, although the radiographic images have great advantageous in administering dental treatments, their subjective interpretation drawback can pose significant challenges such as inaccuracies in the diagnosis and treatment.

Advancements in artificial intelligence (AI), specifically deep learning (DL) methods, have the potential to revolutionize this process. AI refers to the development of computer systems that can perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation. DL, a subset of AI, involves the use of artificial neural networks with multiple layers to automatically learn and extract hierarchical representations of data. In the context of dental diagnostics, AI and DL can be employed to analyze PRs with a level of precision and efficiency that surpasses traditional human interpretation.

Utilizing DL techniques for the diagnostic analysis of dental radiographic images may offer a potential solution to address the challenges associated with low reliability, while achieving comparable or higher levels of accuracy compared to specialists[19]. For example, teeth identification and numbering on panoramic images have critical importance for dentists to accurately diagnose and plan appropriate treatments. In the study conducted by Bilgir et al., an AI based approach was proposed for the automatic detection and enumeration of teeth[20]. In this research, a comprehensive dataset consisting of 2482 anonymized panoramic radiographic images was employed, upon which the Faster R-CNN, a region-based object detection deep learning algorithm, was trained. Within the architecture of Faster R-CNN, the Inception v2 was utilized as the convolutional neural network (CNN) module. For the training of the model, 1984 images were deployed, while an additional 498 images were employed for validation and performance evaluation purposes. The performance of these models in terms of tooth detection and numbering was reported with a notable F1 score of 0.96.

In another study[21] aiming to assist diagnoses in PRs, a noteworthy study, using 1084 anonymized dental images taken from individuals aged 18 and over, was conducted by Başaran et al. Two dento-maxillofacial radiologists labeled 1084 radiographs with 10 different classes, and these labeled images were used to train a Faster R-CNN architecture. The model was trained using a comprehensive dataset of 8676 labeled instances and it achieved a remarkable intersection over union (IoU) based overlap rate of 97.5%. Apart from detecting conditions such as cavities and dental calculus, the proposed AI model has proven to be highly effective and applicable in making a wider range of diagnoses.

DL approaches have also been applied for radiographic detection of apical lesions[22], categorization of endodontic lesion based on radiographic periapical index scoring system[23], C shaped root canal detection in mandibular second molars[24], relationship between the mandibular third molar and the mandibular canal in panoramic radiographs.

In a study similar to the one proposed here, but not focused on children and specifically aimed at diagnosing adult molars, Hiraiwa et al. put forth a deep learning-based approach to determine the number of distal roots in mandibular first molars. This approach uses Cone Beam Computed Tomography (CBCT) classification as the gold standard[25]. Data for the study was retrospectively collected from an image database, including patients who underwent both dental CBCT and PR and had not received root canal treatment. These patients had undergone CBCT examination for various reasons, such as impacted teeth, mandibular deformity, and cysts in other teeth. The study made use of images from 760 mandibular first molars, obtained from 200 males and 200 females. Upon evaluation of the distal roots of these 760 mandibular first molars via CBCT, it was found that 597 teeth (78.6%) had a single root, while 163 teeth (21.4%) had an extra root. The proposed deep learning model's efficacy was assessed based on its ability to discern whether a distal root was single or accompanied by an extra root. The k-fold Cross-validation method was used on the dataset to ensure objectivity in measuring the performance of the deep learning models. The deep learning architectures employed in the study were AlexNet and GoogleNet. The models showed promising results, with accuracy rates of 87.4% and 85.3% in classifying patch images of molar teeth. Although there are significant advancements in the analysis of PRs images by using DL methods, there is a lack of research on the assessment of open/closed apices and root development levels using DL methodologies.

Drawing on the insights gained from the literature and the demonstrated contributions of deep learning in the field, It is believed that clinicians, particularly in dentistry, can substantially benefit from the application of DL. The usage of DL has the potential to transform the assessment of open/closed apex or developmental root stages, utilizing commonly used PRs for initial diagnosis. Such advancements can significantly enhance practicality amidst heavy clinical workload and enable more accurate treatment planning, especially in cases necessitating pulpal treatments. This potential benefit is particularly pronounced for undergraduate students and inexperienced general dentists, who rely heavily on visual assessments and often need to rapidly refer complex cases to a specialist. Moreover, the utilization of DL can greatly assist pediatric dentists in their treatment planning for pulpal treatments, both in open and closed teeth. As such, the light shed by the aforementioned DL applications in the literature reinforces our conviction in the immense value and efficacy of these emerging technologies in improving diagnostic accuracy and treatment planning in dentistry.

Although there have been significant advancements in analyzing panoramic images by using deep learning methods, there is a research gap on the detection of open/closed apices and the evaluation of root development

stages using deep learning methodologies in the existing literature, particularly on pediatric panoramic radiographs. To address this gap, following contributions are made.

- This study presents, to our knowledge, the first end-to-end system that detects permanent first molars with YOLOv7 and classifies Cvek root-development stages with a VGG-19–based classifier on pediatric panoramic radiographs.
- Targeted analyses are provided for mandibular versus maxillary teeth, which reveal anatomical differences affecting model performance.
- Binary open/closed apical foramen predictions are evaluated, including an alternative analysis that excludes Stage 4 to reflect clinical decision points.
- Different CNN architectures and hyper-parameters using five-fold cross-validation are systematically compared and precision, recall, F1 scores and confidence intervals across multiple datasets are reported.

Therefore, the aims of this study were to determine the most accurate deep learning method among the tested architectures for predicting root development stages and apical status on pediatric panoramic radiographs, to analyse mandibular and maxillary subsets as well as binary open/closed tasks, and to quantify performance using standard metrics. This study does not include a formal human-versus-AI comparison; instead, predictions are evaluated against consensus labels provided by two pediatric dentists.

Material and methods

Study design and ethical considerations

This retrospective study was conducted at the Department of Pediatric Dentistry, University of Health Sciences, Istanbul, Turkey according to the checklist for artificial intelligence in medical imaging (CLAIM) guidelines[26]. The study protocol was approved by the University of Health Sciences, Hamidiye Scientific Research Ethics Committee (Approval No: 25/244). Since this was a retrospective study, an informed consent waiver was granted by the University of Health Sciences, Hamidiye Scientific Research Ethics Committee. The study was carried out in accordance with the ethical standards of the institutional research committee and with the 1964 Declaration of Helsinki and its later amendments.

Sample size calculation

Since this study was designed as a retrospective analysis, the sample size was determined based on the availability of panoramic radiographs that met the inclusion criteria during the study period. In total, 409 pediatric patients provided 1629 first molar teeth images (817 maxillary and 812 mandibular). This number exceeded the minimum requirement estimated by a priori power analysis (G*Power 3.1). The calculation assumed a medium effect size (Cohen's d = 0.5), a significance level of $\alpha = 0.05$, and a statistical power of 0.80. Here, $\alpha = 0.05$ represents a 5% risk of Type I error, which is the conventional threshold widely adopted in biomedical research. A power of 0.80 indicates an 80% probability of detecting a true effect if one exists, thereby minimizing the likelihood of a Type II error. Under these assumptions, the minimum required sample size was estimated as 385 images. Therefore, the inclusion of 1629 images was considered sufficient to ensure statistical reliability and robust model training.

Image data set

Panoramic radiography was selected as the imaging modality because it provides a wide field of view and exposes patients to lower radiation doses than cone-beam computed tomography (CBCT), which is especially important in pediatric dentistry. The dataset comprises PRs (Vatech Digital X-ray Imaging Sytem, Seoul, Republic of Korea) from 409 children according to the manufacturer's instructions (67 kVP, 8 mA), consisting of panoramic images and corresponding label files that pinpoint the coordinates of the teeth. The images were selected from an image database of patients who admitted to pediatric dentistry clinic for routine check-ups and dental treatments between September 2022- February 2023. The patient cohort consisted of 204 girls and 205 boys with an age range of 5 to 14 years. Images of 1629 maxillary (817) and mandibular (812) first molars from these patients were used for this study. The patients having at least one first permanent molar tooth were included in the study. Patients having images with artifacts or foreign body were excluded from the study. The use of a single center and single vendor means that generalization to other populations, imaging protocols or vendors has not been established; future multi-center validation studies are required.

Image labelling

In the tooth labelling stage, two pediatric dentists initially labelled the PRs according to the coordinates of first molar teeth to extract these regions in the first stage of the proposed methodology. This process yielded a total of 1629 images: 406 (24 %) were mandibular right molar, 406 (24%) were mandibular left molar, 409 (25%) were maxillar right molar, 408 (25%) was maxillar left molar) were labelled by LabelStudio[27]. These regions were provided to the model (YOLO) specifically for the detection of the first molar teeth. Both the crowns and roots of the first molar teeth were labeled as pinpoint the coordinates. Subsequently, two pediatric dentists independently numbered the root markings based on the root development levels (according to the Cvek classification) of the children. The stage of root development was estimated from the width of apical foramen and the length of the root and categorizing teeth were divided in 5 stages; I, II, III, IV, V. The teeth with wide, divergent apical opening and a root length estimated to less than 1/2, 1/2, and 2/3 of the final root length, were I, II, and III respectively. The teeth with a wide open apical foramen and nearly completed root length were group IV and the teeth with closed apical foramen and completed root development were classified as group V according to the Cvek classification. Initially, the data were independently annotated by two pediatric dentists. After one month, the same experts re-evaluated and re-labeled the data, which showed consistency with the initial annotations. Images for which

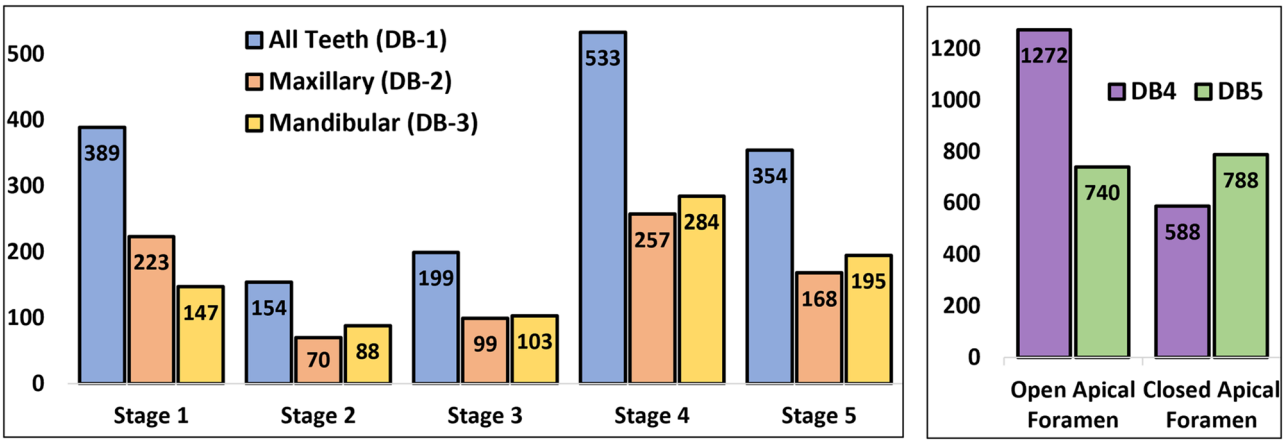

Fig. 1. Datasets used in the classification of molar roots.

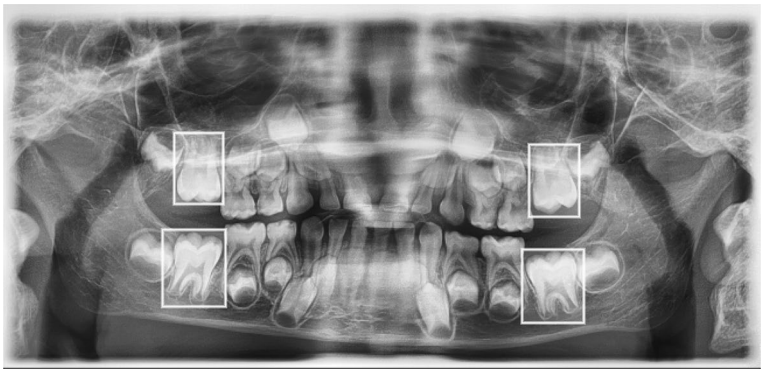

Fig. 2. Panoramic dental X-ray image with the first molars labeled.

agreement could not be achieved were excluded. Subsequently, the two pediatric dentists performed the labeling process together and resolved discrepancies through consensus. In cases where disagreement persisted, a third independent pediatric dentist with ten years of clinical experience was consulted, and the final decision was established by consensus

In the root analysis, five distinct scenarios were implemented. In the first scenario, models were applied to all extracted molar teeth patches which are named as DB-1. For the second scenario, the molar images were analyzed only for maxillary molars by creating as DB-2. In the third scenario, the mandibular molar images were analyzed as DB-3. The training of the DL models was conducted on these distinct sets. In the fourth scenario, the initial four phases stages were combined and evaluated as 'open apical foramen,' and the fifth phase was classified as 'close apical foramen'; training was then performed on this dataset (DB-4). In order to balance the sample distribution, 234 more molar region patches have been added to stage 5. In the fifth scenario, stage 4 images in the DB-3 was excluded and a new sub-dataset was created as DB-5, which utilized in the training phase of the models. A comprehensive overview of the content of the datasets and the quantity of images contained within each is illustrated in Figure 1.

Applied deep learning methods

In the proposed study, YOLO was used for cropping the permanent first molar teeth images from PRs (Fig. 2). Visual geometry group nets (VGG-19), inception V3, and EfficientNet-B3 DL models were used to predict the development stages or the status of apex foramen as open or closed. The details of the employed detection and classification algorithms are given below in a detailed manner.

You only look once

Object detection is a technological field intertwined with computer vision and image processing that focuses on identifying instances of semantic objects belonging to a particular class (such as humans, buildings, or cars) in digital images and videos. One of the prominent models in object detection is the You Only Look Once (YOLO) series. The YOLO series has evolved in parallel with the advancements in deep learning, beginning with YOLOv1 in 2015, followed by the development of YOLOv2 in 2016, YOLOv3 in 2018, YOLOv4 and YOLOv5 in 2020, and the most recent additions of YOLOv6 and YOLOv7. YOLOv7, the latest iteration in the series, expands

the model's capabilities beyond mere target identification. It has been deployed in more nuanced applications such as case segmentation and the evaluation of human postures. This demonstrates how deep learning models like YOLO continue to evolve, providing increasing value across a range of fields and use cases.[28].

Visual geometry group nets

VGG-19 is a CNN that has been trained on over a million images from the ImageNet database[29]. This network, composed of 19 layers, is capable of classifying images into 1000 distinct object categories. The model's extensive training has led to it acquiring rich feature representations applicable to a wide array of images. VGG-19, along with VGG-16, which consists of 16 convolutional layers, are renowned for their depth. The high number of convolutional layers allows these models to learn and recognize complex patterns in images, thereby boosting their performance on image classification tasks. Both models are part of the Visual Geometry Group (VGG) family of models, known for their effectiveness in handling visual data. In this study, VGG-19 was implemented in the experiments.

InceptionV3

InceptionV3, which utilizes the Inception CNN architecture, is a deep learning model designed by Google. This architecture represents a significant breakthrough as it facilitates the development of deeper and more intricate networks. Trained on the ImageNet dataset, InceptionV3 has demonstrated high performance in broad object recognition tasks. Remarkably, it manages to achieve superior accuracy with fewer parameters compared to its predecessors, which attests to its efficient learning capability[30]. The fundamental principle behind InceptionV3 is the use of convolution filters of varying sizes. This strategy allows the network to capture features at different scales and depths. Moreover, the model employs a "bottleneck"structure to increase its learning capacity. This is accomplished by parallelizing convolution layers and implementing 1×1 convolution filters, which reduces computational costs while maintaining robust learning capabilities. Given its versatility, InceptionV3 is suitable for an array of tasks, such as object recognition, object detection, and image classification.

EfficientNet-B3

EfficientNet-B3 is a deep learning model belonging to the EfficientNet family. This model was conceived utilizing Compound Scaling, a strategy designed to optimize model scaling. Compound Scaling aims to enhance model performance by taking into account pivotal dimensions of the network, including depth, width, and resolution[31]. Compared to other variants in the EfficientNet series, EfficientNet-B3 is of medium size. It possesses more parameters and exhibits a deeper structure relative to its lighter counterparts. The attributes and performance of EfficientNet-B3 generally offer superior accuracy and improved generalization capabilities over lighter models. However, its larger size may necessitate more computational resources and memory. This trade-off is an important consideration when choosing a model for specific tasks and computational constraints.

Molar tooth detection module

The initial phase of the study involved the identification of the first molars. For this purpose, YOLOv7, a prominent object detection algorithm, was employed. This algorithm operates by processing an accompanying txt file with the image. The txt file maintains the coordinates and label of the target object in YOLO format. An example dental x-ray from the dataset is illustrated in Fig. 2.

Root classification module

In the classification phase, the first molars were extracted from the dental panoramic images. These extracted images were then segmented into various phases and supplied to object classification models VGG-19, InceptionV3, and EfficientNet-B3. A few examples of these extracted teeth can be observed in Fig. 3.

Utilizing these cropped images, a variety of datasets were created. Four different scenarios were organized using these datasets, which were also detailed in the 'Image labelling' section. The specified datasets were employed to construct the following models: Model-1 represents an analysis model for all teeth. Model-2

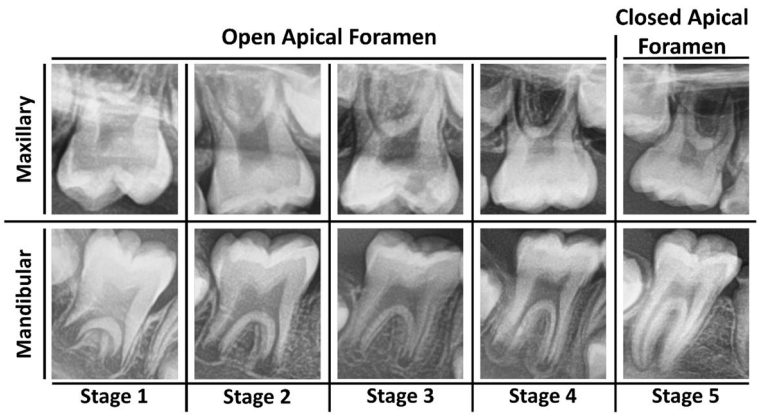

Fig. 3. Examples of tooth images divided into stages.

provides a specialized analysis model for maxillar - mandibular teeth. Model-3 examines teeth open–closed apical foramen analysis (stages I,II,III,IV,V), while Model-4 studies teeth open-closed apical foramen analysis but excludes stage 4 (thus considering stages I,II,III-V according to the Cvek classification).

Proposed approach

The system's workflow is illustrated in Fig. 4. Initially, a panoramic image is fed into the system. This image is subsequently processed by our trained YOLOv7 model, which determines the coordinates of the permanent first molars. Based on these coordinates, teeth images are extracted and fed into the chosen VGG-19 based models. The system ultimately produces output based on the model's predictions. In the molar detection phase, the dataset is partitioned into three parts. The 'valid' and 'train' folders are employed for training the YOLOv7 model, while the 'test' folder is utilized during the evaluation phase. The model, which was trained to detect permanent first molar teeth, ran for 100 epochs, in approximately 2 hours. The trained model was evaluated with 0.001 confidence value and 0.50 IOU value. The batch size was chosen as 16.

All model training and inference were executed on a single workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), using Python 3.8, PyTorch 2.3.0 for the YOLOv7 detection model and TensorFlow/Keras 2.12 for the classification networks.

In the classification phase of this current study, permanent first molar teeth, which were segmented from panoramic images were used. Prior to training, each cropped tooth image was resized to 224×224 pixels, normalized with the mean and standard deviation of the ImageNet dataset and randomly flipped horizontally with probability 0.5. For the VGG-19 classifier a dropout layer (p=0.5) was inserted before the final fully connected layer to mitigate overfitting. Root development stages of these teeth were classified into five categories. Then, CNN models were employed to determine the root developmental stages of teeth. The dataset, consisting of 1629 segmented tooth images, was divided into five folds. Each model was trained for 250 epochs per fold. Three optimizers (Adamax, SGD and RMSProp) and three learning rates (10^{-3}, 10^{-4} and 10^{-5}) were employed in the experiments. No class weighting or focal losses were applied; class imbalance was partially addressed by adding additional Stage 5 patches. A fifth (20%) of each fold was reserved for testing. Models were trained on each fold to measure the performance of the algorithms, and the average accuracies were computed. Different parameters and models were tested on all teeth, with the VGG-19 model yielding the most successful results. For a separate analysis of maxillar and mandibular teeth, the entire dataset (DB-1) was bifurcated as DB-2 and DB-3, and the partitioning process was applied to the folds again. The VGG-19 model was trained on these two divided datasets. A binary classification was also performed for the analysis of tooth root development. The teeth with open apical foramen correspond to the first, second, third, and fourth stages while the fifth stage is closed apical foramen. The classification model was trained on a dataset (DB-4) wherein the first four stages were categorized as 'open' and the fifth stage as 'closed'. An alternative approach involved excluding the fourth stage in this process (DB-5). Accuracy of the methods were calculated and Kappa tests were performed.

Experimental results

In this study, four models were trained, including YOLOv7, EfficientNet-B3, InceptionV3, and VGG19. The YOLOv7 model was utilized for the detection of permanent molar teeth. After various tests to achieve the best results, it is decided to train with a batch size of 16 over 100 epochs. According to the test data, the model achieved a mean Average Precision (mAP) of 99.3% at an Intersection over Union (IoU) threshold of 0.50. A more detailed evaluation results for the YOLOV7 model are displayed in Table 1. The precision value stands at 98.4% and the recall value was 97.6%. As illustrated, the detection of the first permanent molar teeth was highly successful.

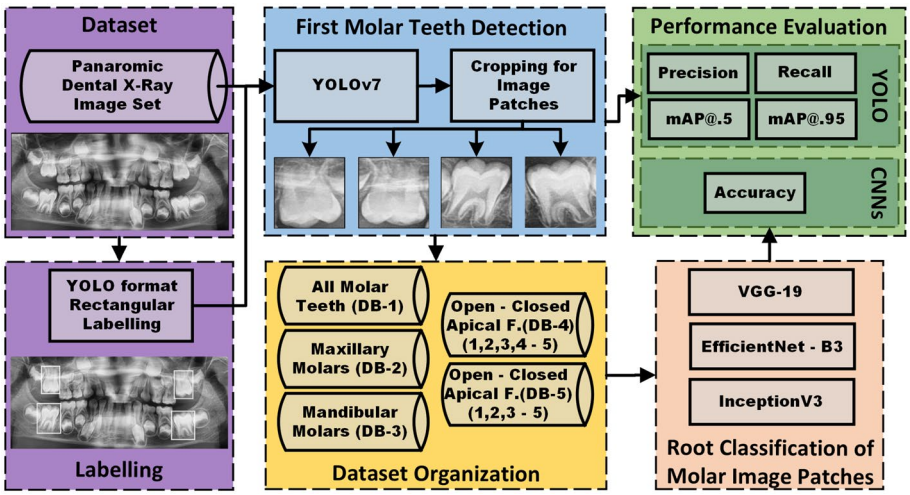

Fig. 4. The flowchart of the study.

Number of Images	Label Counts	Precision	Recall	mAP @.5	mAP @.95
63	252	0.984	0.976	0.993	0.727

Table 1. The test results of the YOLOv7 model utilized for the first molar detection in dental X-ray images.

		VGG-19		InceptionV3		EfficientNet-B3	
Opt.	**LR**	**Accuracy**	**Std.**	**Accuracy**	**Std.**	**Accuracy**	**Std.**
Adamax	10^{-3}	64%	1.9%	63%	1.7%	53%	3.2%
	10^{-4}	**64%**	**0.4%**	63%	2.1%	53%	2.1%
	10^{-5}	63%	1.6%	57%	4.7%	50%	1.6%
SGD	10^{-3}	60%	2.3%	63%	1.6%	53%	2.5%
	10^{-4}	58%	1.9%	56%	2.8%	39%	9.0%
	10^{-5}	52%	2.3%	40%	2.4%	28%	2.4%
RMSProp	10^{-3}	63%	2.1%	62%	2.7%	53%	2.3%
	10^{-4}	61%	2.0%	63%	2.3%	51%	3.6%
	10^{-5}	63%	0.9%	56%	3.7%	50%	1.9%

Table 2. Individual model performances with hyper-parameter selections on the root classification of all teeth dataset (DB-1). Significant values are in bold.

The remaining three models, EfficientNet-B3, InceptionV3, and VGG19, were used to classify the developmental stages of the roots of the detected teeth. To secure the most accurate results with these models, cross-validation method with five folds and experimented with different parameter values were utilized. All trials were trained over 250 epochs. Three different optimizers—Adamax, SGD, and RMSProp - and three different learning rate values—10^{-3}, 10^{-4}, and 10^{-5} - were tested over DB-1 which includes all first molar teeth with a five stage category. The average classification accuracies and the standard deviation values over five folds are given in Table 2.

After varying these parameters across all three models, the most successful result was achieved by the VGG19 model, which was trained with the Adamax optimizer with a learning rate of 10^{-4}. This model achieved an average accuracy of 64.21% across all folds. The results of the VGG- 19 model operated for 5 folds with different parameters are shown in Table 2.

In an effort to enhance the performance of the model, a dropout parameter to the VGG19 model was introduced, in addition to the parameters previously tested. The primary aim in utilizing the dropout parameter is to mitigate over-fitting that may occur due to the relatively small volume of data that is used, which does not present a balanced tribution across the different stages. To prevent this, the dropout parameter was added to the network architecture of VGG-19.

Another factor impacting the success of the current model is the inherent difference between the structures of maxillary and mandibular teeth. Since the apexes of the maxillary teeth are less pronounced, the classification stage becomes complicated. To address this, accuracy was improved by modeling the mandibular and maxillary teeth separately. As a result of these trainings, the accuracy rate was 69.64% for the mandibular teeth (DB-3) and 62.66% for the maxillary teeth (DB-2).

Another crucial factor when staging root developments is whether the root apexes are open or closed. In the 1st stage, the root apex is open, and they progressively close up until the 4th stage. By the 5th stage, the root apex is completely closed. To assess the model's accuracy in classifying root apices as open or closed, the teeth were categorized into two groups (DB-4): stages 1 through 4, where the root end is open, and stage 5 where the root end is closed. The model was then retrained based on these new classifications. As a result of this training, the model achieved an accuracy rate of 84.57% in the classification performance of molar teeth as open apex/closed apex over DB-4.

One of the factors influencing success in this analysis, is the significant similarity between the 4th and 5th stages. To determine how much this overlap affected accuracy, the model was retrained after excluding the fourth stage from the dataset. The 1st, 2nd, and 3rd stages were classified as the open apex group, and the 5th stage was placed in the closed apex group (DB-5). This training yielded a success rate of 94.89%. As such, this analysis proves that the high similarity between the 4th and 5th stages negatively impacts the model's accuracy.

All the additional models and scenarios throughout this study can be found in Table 3.

In addition to the precision and recall reported above (Table 3), F1 scores and 95% confidence intervals (CI) were calculated for the classification accuracies across the five cross-validation folds. These metrics provide a more balanced view of performance for imbalanced classes. Table 4 summarizes the mean accuracy ($\pm 95\%$ confidence interval), macro-F1 and per-class F1 scores for each dataset scenario. For the binary tasks (DB-4 and DB-5) only F1-scores for the open and closed categories are reported.

(%)	All molar teeth (DB-1)					Maxillary molar (DB-2)					Mandibular molar (DB-3)					Open–closed apical foramen			
																DB-4		DB-5	
Acc.	64.21					62.66					69.64					84.57		94.89	
Std.	0.4					3.85					2.28					3.16		0.55	
Apex	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	Open	Closed	Open	Closed
Prec.	93.1	43.3	34.8	59.7	67.2	85.9	37.7	35.6	57.6	65.1	88.1	49.4	44.1	67.2	79.2	91.1	72.7	94.2	95.7
Recall	80.2	53.2	23.1	72.8	61.5	85.2	32.8	31.3	70.8	51.2	85.7	54.5	25.2	81.7	70.2	85.8	81.9	95.5	94.3
Kappa	0.527					0.504					0.595					0.655		0.898	

Table 3. Results of all experimental scenarios.

Dataset	Mean accuracy (±95% CI)	Macro-F1	F1-stage 1	F1-stage 2	F1-stage 3	F1-stage 4	F1-stage 5	F1-open	F1-closed
DB-1 (all teeth)	64.21 ± 0.40 %	58.28 %	86.12 %	47.74 %	27.80 %	65.46 %	64.28 %	--	--
DB-2 (maxillary)	62.66 ± 4.78 %	54.51 %	85.55 %	35.09 %	33.36 %	63.60 %	56.95 %	--	--
DB-3 (mandibular)	69.64 ± 2.83 %	63.43 %	86.87 %	51.80 %	32.25 %	73.77 %	74.45 %	--	--
DB-4 (open vs closed)	84.57 ± 3.92 %	82.73 %	--	--	--	--	--	88.36 %	77.09 %
DB-5 (open vs closed w/o Stage 4)	94.89 ± 0.68 %	94.92 %	--	--	--	--	--	94.85 %	95.00 %

Table 4. Class-wised F1 scores and $\pm 95\%$ confidence intervals (CI).

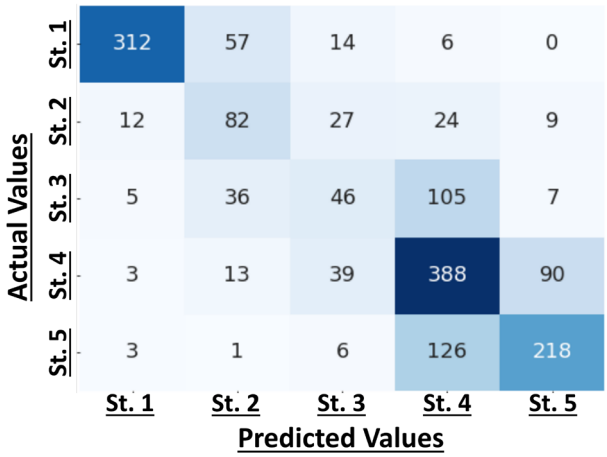

Fig. 5. The confusion matrix of VGG-19 model on DB-1 (all molars).

These F1 values were derived from the precision and recall values in Table 3 and illustrate the decrease in performance on the more ambiguous stages 2 and 3. The high F1 scores for the binary tasks demonstrate that the model is more reliable when differentiating open from closed apices than when predicting individual stages.

Regarding the class based performance analysis for all teeth, the average confusion matrix of the most successful parameter setting for the VGG-19 model, achieved with the Adamax optimizer with a learning rate of 10^{-4} is illustrated in Figure 5. The proposed classification model has shown a considerably good proficiency in identifying Stage 1 and Stage 4 teeth, by correctly classifying 312 out of 389 instances in Stage 1 and 388 out of 530 instances in Stage 4. However, the proposed model also misclassified a number of Stage 1 instances as Stage 2 and a significant proportion of Stage 4 instances as Stage 5. Regarding the Stage 2 teeth samples, the model has shown a relatively modest performance by correctly predicting 82 out of 154 instances, while a significant number of samples were incorrectly classified as Stages 3, 4, and 5. The model's performance was worst for predicting Stage 3 teeth, with the majority of these instances erroneously identified as Stage 4. In terms of Stage 5 teeth, the proposed model correctly classified 218 out of 354 instances, yet misclassified a substantial number as Stage 4. The outcomes of the proposed learning model reveal notable differences in distinguishing between certain stages. Specifically, the model frequently misclassifies Stage 1 instances as Stage 2 and a significant proportion of Stage 4 instances as Stage 5. The superiority of the model performance in distinguishing Stage 1 and Stage 4 teeth indicates that the VGG19 model has better learned extracting the features of these stages, or the extracted features of these stages are more distinct than the others. On the other hand, the poor performance seen in classifying teeth at Stage 2 and Stage 3 suggest that the features of these stages are more ambiguous, or that these stages have characteristics that overlap significantly with the other stages. The most striking observation is the

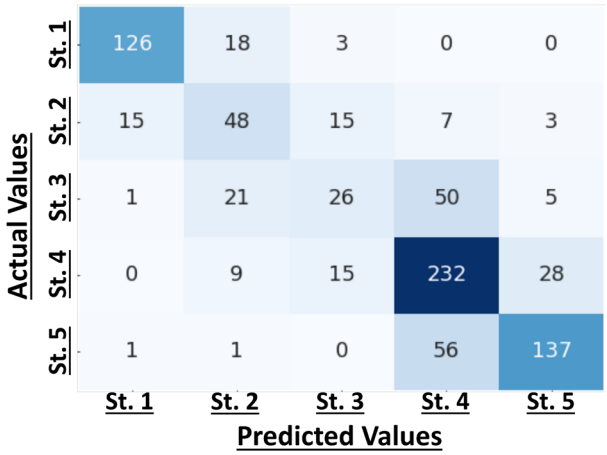

Fig. 6. The confusion matrix of VGG-19 model on DB-3 (mandibular molars).

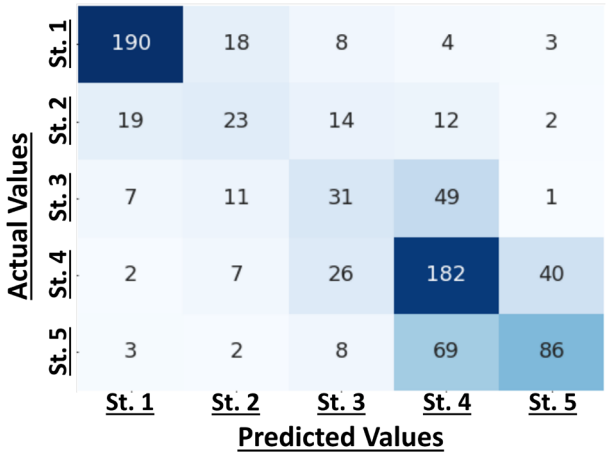

Fig. 7. The confusion matrix of VGG-19 model on DB-2 (maxillary molars).

proposed model's difficulty in distinctly separating samples belonging to Stage 4 and Stage 5. This indicates that samples from these two stages possess very similar features, leading the model to struggle in learning the differences between them.

Figure 6 displays the average confusion matrix for the most successful parameter setting of the VGG19 model, specifically applied to the mandibular teeth. From this confusion matrix, it seems that certain performance aspects reflect patterns observed in the previous all-teeth model, while others show interesting deviations. For instance, it can be noticed that a similar performance to the all-teeth model was achieved for the extreme phases (Stage 1 and Stage 5), with the correct predictions being 126 for Stage 1 and 137 for Stage 5. These phases represent the starting and ending stages of the root development. The morphological characteristics are more distinct in these phases, which allows the model to more easily distinguish the samples belonging them. Similar to the results obtained from the all-teeth model, the proposed model shows less accuracy for the Stages 2, 3 and 4. A key observation here is that the model often confuses the Stage 3 teeth for Stage 4 (50 instances), mirroring the behaviour seen in confusion matrix of the all-teeth model. In contrast, an interesting distinction can be drawn in terms of Stage 4 and Stage 5. While there was a significant confusion between these two stages in the all-teeth model, the confusion has relatively reduced for the mandibular samples. This suggests that the morphological differences between Stage 4 and Stage 5 are more recognizable in mandibular teeth compared to the all-teeth model. These findings suggest that the proposed model, while having certain common behaviours across different teeth types, may also exhibit teeth-specific performance characteristics. This reinforces the importance of considering individual anatomical structures in dental image classification tasks, as their unique characteristics may significantly influence the model performance.

Figure 7 shows the average confusion matrix for the optimal parameter setting of the VGG19 model, as applied to the maxillary teeth. As it can be seen in this confusion matrix, several interesting findings emerge when the proposed model is applied to the classification problem of the maxillary teeth. For example, although the proposed model shows relatively strong performance for Stage 1 by correctly predicting 190 instances, it exhibits some weaknesses in the prediction of Stage 2 samples and discerning the distinction between Stage

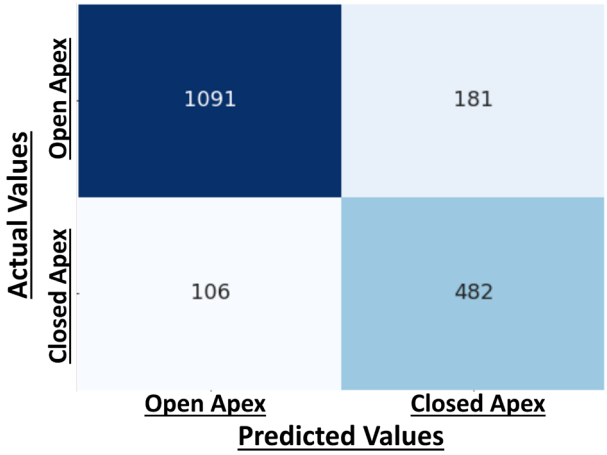

Fig. 8. The confusion matrix of VGG-19 model on DB-5 (open-closed apex analysis with all stages).

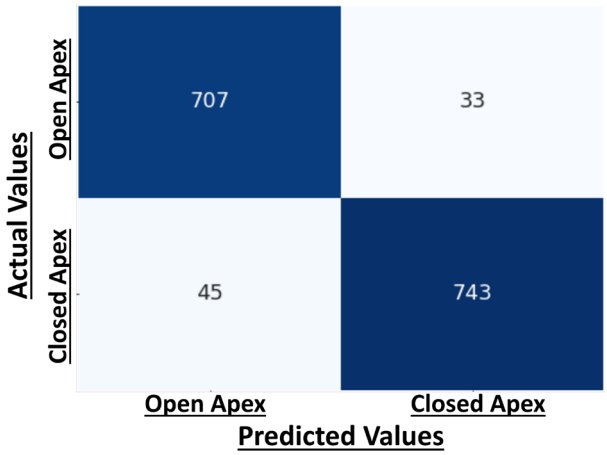

Fig. 9. The confusion matrix of VGG-19 model on DB-5 (open–closed apex analysis with all stages except 4th stage).

3 and 4, and Stage 4 and 5. A large number of Stage 2 teeth are misclassified as Stage 1 and Stage 4. A similar challenge arises in differentiating between Stage 3 and Stage 4 teeth, where a high number of Stage 3 teeth are misclassified as Stage 4. Furthermore, despite the model's reasonable success in predicting Stage 4 teeth, a large portion of Stage 5 teeth are misclassified as Stage 4. This indicates that the proposed model fails to distinguish between the final stages of tooth development in maxillary teeth. Overall, the model's performance on maxillary teeth is noticeably weaker compared to its performance on mandibular teeth.

Figure 8 displays the average confusion matrix for the best parameter configuration of the VGG19 model when employed for an open versus closed root analysis involving all stages. The consolidated matrix results from aggregating the individual confusion matrices of each fold during evaluation, applying the same optimal VGG19 model configuration referred to previously. Additionally, the Figure 9 presents the average confusion matrix for the folds using the most successful parameter setting of the VGG19 model for the open versus closed root analysis across all stages, excluding the 4th stage.

In the early stages, first molar teeth are particularly susceptible to cavities, as mentioned previously. Decay in these teeth can rapidly progress towards the pulp. The molar teeth, which often have an open apical foramen during the initial eruption phases, are subjected to various treatments. The open or closed apical foramen conditions of the tooth's root are one of the key factors in the treatment planning. The possible appropriate treatments (such as the conventional root canal therapy or regenerative endodontic treatment) applied can be decided after determining the status of apical foramen. Moreover, to administer the correct treatment, it's imperative to assess the apical foramen in conjunction with other symptoms present in the tooth. These show the necessity of having a learning model which can be employed to determine the open/closed apical foramen conditions of the tooth's root. Figure 8 displays the average confusion matrix for the best parameter configuration of the VGG19 model when employed for an open versus closed root analysis involving all stages. The consolidated matrix results from aggregating the individual confusion matrices of each fold during evaluation, applying the same optimal VGG19 model configuration referred to previously.

According to the Cvek classification, root development is divided into five stages. Stage 4, which describes a root development level where the apical foramen is nearing closure, has shown similar appearances to stage 5. Consequently, there is a potential for confusion between stages 4 and 5 in the determination. There are no distinctions between the mandibular teeth and maxillary teeth concerning these five stages. But, in the maxillary jaw, due to the superimpositions created by the anatomical structures on the teeth roots, the likelihood of confusion between stages 4 and 5 increases as root development progresses. In the proposed study, a modified train/test data organization is also applied to be able to reduce the effect of similarity seen between phases 4 and 5, which has a potential to induce confusion in models. To do so, the image samples belonging to stage 4 were extracted from the train and test data. Later, a new classification approach, in which the open apical foramen consists of images of stages 1, 2 and 3, and the closed apical foramen was just made of stage 5 images, was carried out. Figure 9 presents the average confusion matrix for the folds using the most successful parameter setting of the VGG19 model for the open versus closed root analysis across all stages, excluding the 4th phase. As it can be seen in Fig. 9, a more balanced data distribution was achieved with the exclusion of stage 4 samples, and this causes a better generalization ability for the models resulting in enhanced classification performance.

Discussion

A comprehensive view of the dental arches can be provided by the PRs that are commonly used in dental diagnostic scan. However, the usage of PRs suffer from inherent limitations in the visualization of three-dimensional anatomical structures, such as the palatal root of the maxillary molar. These regions in PRs may be obscured or distorted due to overlapping anatomical structures in two-dimensional plane. On the other hand, the Cone-beam computed tomography (CBCT) imaging may offer a more accurate and detailed representation of the dental anatomy resulting in a better assessment of complex root morphology and pathology. However, the CBCT based imaging produces higher radiation doses compared to PRs[32–34], and this makes the usage of CBCT less preferable for routine practice especially in pediatric patients who are more sensitive to radiation exposure[35,36]. Therefore, the Panoramic radiography is chosen as the main imaging technique in the proposed study that can be trusted as a valuable diagnostic tool due to its wide availability, lower radiation dose, and cost-effectiveness.

The recognition of panoramic radiographs as a valuable method for evaluating root development stages indicates that the use of CBCT may often be unnecessary, consistent with the ALARA principle. This is particularly relevant in the assessment of early developmental stages, such as Stage 1 and Stage 2, where imaging involves very young patients. In these cases, reliance on panoramic radiography plays a critical role in minimizing radiation exposure while still providing sufficient diagnostic information. Also; the dataset used in this study was obtained from a single institution and involved a relatively narrow age range. This limitation may affect the generalizability of the model to different populations and imaging systems. No external validation was performed in the present study; however, future work will involve multicenter studies using datasets from various institutions and imaging devices to assess the robustness and generalizability of the model.

Clinically, misclassification between Stage 4 (nearly closed apex) and Stage 5 (closed apex) has limited impact on treatment planning because both stages often receive similar interventions–such as apical barrier techniques or regenerative procedures. Therapeutic decisions are generally driven by the size of the apical foramen and the presence of symptoms rather than by the exact developmental stage. Our finding that the model frequently confuses Stages 4 and 5 therefore reflects not only the limitations of two-dimensional panoramic imaging but also genuine clinical ambiguity. Future work may explore merging these stages or adopting ordinal classification schemes to better capture their continuum.

The mandibular first molar is the first erupting permanent tooth and is susceptible to a high risk of decay and subsequent pulp damage, necessitating frequent endodontic intervention[24,25]. The outcome of endodontic treatment relies on several contributing factors. Among all radiographic examinations, panoramic radiographs are commonly employed as a routine diagnostic tool especially during endodontic treatments. They are frequently utilized to assess the various situations and configurations of root canal structures. Deep learning studies have been conducted on dental root using panoramic radiographs[22,23,37]. Panoramic radiographs are used to examine the open/closed apical foramen and root development stages in the current study.

Our analysis also highlighted that Stages 2 and 3 were the most difficult to distinguish. The morphological differences between these consecutive stages are subtle on panoramic radiographs, and superimposition of anatomical structures particularly in the maxillary arch–further blurs stage boundaries. This is consistent with the low F1 scores reported for Stages 2 and 3 (Table 4). Future research could explore ordinal losses that respect the ordered nature of development or merge adjacent stages (e.g., combining Stages 2 and 3) when clinically appropriate.

In a critical meta-analytical study, the stage of root develeopment is identified as one of the predictors of success of pulpotomy treatment[8]. It was observed that open-apex teeth demonstrated more successful outcomes in pulpotomy treatments, which suggested a possible association with incresased vascular and cellular proliferation in the pulp. The evaluation of open/closed apex and root development stages through deep learning plays a crucial role in determining the treatment success. This is because the main criteria for success lies in making an accurate diagnosis and implementing appropriate treatments. In the present study, a CNN model was successfully trained and applied to evaluate whether the root apexes were open or closed, based on the Cvek classification. The model demonstrated an ability to detect open/closed apex with an accuracy of 84.57% . In the analysis conducted by excluding the 4th group, a success rate of 94.89% was observed. It is believed that the similarity between the images of the 4th and 5th stages has influenced the model's performance. Clinically, due to the nearly closed definition of the root canal in the 4th group according to Cvek, it is considered that the treatments applied to 5th can be implemented. As mentioned above, the state of open/closed apical foramen influences treatment planning. The Clinical Considerations for a Regenerative Procedure by the American

Association of Endodontists, drawing on Cvek's classification of root development, outlines stage-specific treatment recommendations. Teeth classified as Stages 1, 2, and 3 are generally best managed with regenerative endodontic procedures due to their incomplete root formation. At Stage 4, however, the increased thickness and strength of the root canal walls permit either a regenerative protocol or the use of an apical MTA plug followed by conventional root canal obturation. This same treatment strategy is also considered appropriate for Stage 5 teeth, in which root development is complete and the apex is fully closed[14,38]. However, what is particularly emphasized here is the importance of classifying root development not only based on open/closed apical foramen but also according to the different developmental stages. Several radiographic classifications have been proposed to facilitate the interpretation of different stages of root development. These classifications encompass the range from the initiation of crown calcification to the completion of root maturation. Two common indexes, namely the Moorrees and Demirjian classifications, incorporate multiple subdivisions[39,40]. Primarily used in forensic sciences, these indexes aid in estimating the age of individuals and determining their biological identification in criminal, civil, or administrative contexts[41]. Conversely, it is more suitable for clinicians to employ an index that specifically supports therapeutic decision-making in their daily practice by solely focusing on root development. The Cvek classification fulfills this requirement, as it was introduced for radiographic diagnosis of traumatized teeth that have already erupted, encompassing five subdivisions[40]. Radiographic interpretation in any biomedical field is an inherently imperfect visual process that can be adversely affected by various factors. These factors include inherent biases in perception, reasoning abilities, and the cognitive decision-making process of each individual observer. In a study evaluating the intra-observer and inter-observer agreements during the radiographic assessment of different stages of root development using the dichotomous, Moorrees, Demirjian, and Cvek classifications, the Cvek classification demonstrated a high level of agreement. Furthermore, the study also assessed the impact of observer experience on the results[42]. Evaluation of PRs of two experienced pediatric dentist according to the Cvek classification constitute our data in the current study. Success in training with all teeth is 64.21%, success in training maxillar teeth is 62.66%, success in training mandibular teeth is 69.64%, success in open apical foramen (first 4 stages) - closed apical foramen (5th stage) analysis is 84.57%, success in open apical foramen (first 3 stages)––closed apical foramen (5th stage) analysis is 94.89%. According to the classification-based (Cvek root classification stages) analysis, the lower values in certain groups indicates the need for improvement in this aspect. Deep learning systems have been applied to various medical fields; however, there is still a lack of sufficient applications in the field of dentistry. Therefore, a performance evaluation of deep learning methods was conducted for the detection of open or closed apical foramen of first molars and the assessment of root development levels in pediatric panoramic radiographs, which are closely related to the fields of endodontics and especially pediatric dentistry.

The developmental five stages of the molar roots progress in a sequential manner, which means each stage is a continuation of the previous one with incremental changes. As it can be seen from Fig. 5, this continuum can potentially blur the boundaries between adjacent stages. This haziness makes it hard to accurately distinguish between adjacent stages for the proposed model. The misclassified samples between Stage 4 and Stage 5, which are the last stages of the development and presumably exhibit slight differences, is the particular evident of this phenomenon. These findings highlight the inherent challenge in using deep learning models to classify progressively changing and closely linked stages. It underscores the need for careful fine-tuning of the model to account for the delicacies in the permanent molar teeth root development stages. To overcome this challenge, various CNN models, in which fine tuning by testing different optimizers and learning rates were applied in the proposed study.

The present work has several limitations. First, all radiographs were acquired at a single institution using the same Vatech system; no external validation was conducted, so generalizability to other populations, imaging parameters or vendors is unknown. Second, annotations were generated by two pediatric dentists working independently and then reaching consensus; inter- or intra-rater reliability were not measured, and the model's performance has not been compared directly against clinicians with different levels of experience. Third, per-case prediction scores were not retained, so ROC curves and AUC metrics could not be computed. Fourth, despite the addition of additional Stage 5 samples, the dataset remained imbalanced, and class-weighted or ordinal loss functions were not used; these approaches warrant exploration. Finally, although panoramic radiographs are widely available and involve lower radiation doses than CBCT, their two-dimensional nature introduces overlapping anatomy and may limit accuracy. Future work should include multi-center external validation, reader-study comparisons between AI and dentists, computation of observer variability, and investigation of weighted or ordinal losses. CBCT imaging could enhance accuracy but must be balanced against increased radiation exposure, particularly in children[43]. Despite these limitations, the higher accuracy observed in the binary open/closed tasks demonstrates the feasibility of leveraging deep learning for automated assessment of root development on panoramic radiographs.

Conclusion

This study demonstrates that, convolutional neural network can assist clinicans in detecting the permanent first molar teeth and evaluating the open/closed apical foramen. This potential enhances the efficiency and accuracy in the clinical workflow of dentists, particularly pediatric dentists and endodontists, who administer treatment on open/closed apices. Additionally, this facilitates the determination of which tooth will undergo which treatment modality (regenerative endodontics or conventional endodontic treatments) in the clinical setting.

Data availability

Dataset can be accessible through the Yildiz-Hi-Lab Github website, available at: https://github.com/Yildiz-Hi-Lab/MolarRootStage.

Received: 28 January 2025; Accepted: 30 September 2025
Published online: 05 November 2025

References

1. Kassebaum, N. et al. Global burden of untreated caries: A systematic review and metaregression. *J. Dental Res.* **94**, 650–658 (2015).
2. Kassebaum, N. J. et al. Global, regional, and national prevalence, incidence, and disability-adjusted life years for oral conditions for 195 countries, 1990–2015: a systematic analysis for the global burden of diseases, injuries, and risk factors. *J. Dental Res.* **96**, 380–387 (2017).
3. Lourdes-Ribeiro, M. L. et al. Adolescents' and parents' perspective on impact of dental caries and malocclusion on quality of life. *J. Dent. Child.* **89**, 136–142 (2022).
4. Seitz, M. W. et al. Current knowledge on correlations between highly prevalent dental conditions and chronic diseases: An umbrella review. *Prevent. Chronic Dis.* **16**, E132 (2019).
5. Chrysanthakopoulos, N. A. Reasons for extraction of permanent teeth in Greece: A five-year follow-up study. *Int. Dental J.* **61**, 19–24 (2011).
6. Byahatti, S. M. & Ingafou, M. S. Reasons for extraction in a group of Libyan patients. *Int. Dent. J.* **61**, 199–203 (2011).
7. Lakhani, S., Noble, F., Rodd, H. & Cobourne, M. T. Management of children with poor prognosis first permanent molars: An interdisciplinary approach is the key. *Br. Dent. J.* **234**, 731–736 (2023).
8. Oreano, M. D., Santos, P. S., Borgatto, A. F., Bolan, M. & Cardoso, M. Association Between Dental Caries and Molar-Incisor Hypomineralisation in First Permanent Molars: A Hierarchical Model. (Community Dentistry and Oral Epidemiology, 2022).
9. Solomon, R. V. et al. Coronal pulpotomy technique analysis as an alternative to pulpectomy for preserving the tooth vitality, in the context of tissue regeneration: A correlated clinical study across 4 adult permanent molars. *Case Rep. Dent.* **2015**, 916060 (2015).
10. Linsuwanont, P., Wimonsutthikul, K., Pothimoke, U. & Santiwong, B. Treatment outcomes of mineral trioxide aggregate pulpotomy in vital permanent teeth with carious pulp exposure: The retrospective study. *J. Endod.* **43**, 225–230 (2017).
11. Ather, A., Patel, B., Gelfond, J. A. & Ruparel, N. B. Outcome of pulpotomy in permanent teeth with irreversible pulpitis: A systematic review and meta-analysis. *Sci. Rep.* **12**, 19664 (2022).
12. Dawood, A. & Patel, S. The dental practicality index-assessing the restorability of teeth. *Br. Dent. J.* **222**, 755–758 (2017).
13. Eliyas, S., Briggs, P., Harris, I., Newton, J. & Gallagher, J. Development of quality measurement instruments for root canal treatment. *Int. Endod. J.* **50**, 652–666 (2017).
14. Kim, S., Malek, M., Sigurdsson, A., Lin, L. & Kahler, B. Regenerative endodontics: A comprehensive review. *Int. Endod. J.* **51**, 1367–1388 (2018).
15. Lia, J., Zhengb, L., Daraqelc, B., Liud, J. & Hue, Y. Treatment outcome of regenerative endodontic procedures for necrotic immature and mature permanent teeth: A systematic review and meta-analysis based on randomised controlled trials. *Oral Health Prev. Dent.* **21**, 141–152 (2023).
16. Manning, D. J., Gale, A. & Krupinski, E. Perception research in medical imaging. *Br. J. Radiol.* **78**, 683–685 (2005).
17. Nardi, C. et al. Accuracy of orthopantomography for apical periodontitis without endodontic treatment. *J. Endod.* **43**, 1640–1646 (2017).
18. Nardi, C. et al. Is panoramic radiography an accurate imaging technique for the detection of endodontically treated asymptomatic apical periodontitis?. *J. Endod.* **44**, 1500–1508 (2018).
19. Lin, P., Huang, P. & Huang, P. Automatic methods for alveolar bone loss degree measurement in periodontitis periapical radiographs. *Comput. Methods Prog. Biomed.* **148**, 1–11 (2017).
20. Bilgir, E. et al. An artificial ıntelligence approach to automatic tooth detection and numbering in panoramic radiographs. *BMC Med. Imaging* **21**, 1–9 (2021).
21. Başaran, M. et al. Diagnostic charting of panoramic radiography using deep-learning artificial intelligence system. *Oral Radiol.* **38**, 1–7 (2022).
22. Ekert, T. et al. Deep learning for the radiographic detection of apical lesions. *J. Endod.* **45**, 917–922 (2019).
23. Moidu, N. P., Sharma, S., Chawla, A., Kumar, V. & Logani, A. Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. *Clin. Oral Invest.* **26**, 651–658 (2022).
24. Zhang, L. et al. A lightweight convolutional neural network model with receptive field block for c-shaped root canal detection in mandibular second molars. *Sci. Rep.* **12**, 17373 (2022).
25. Hiraiwa, T. et al. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. *Dentomaxillofac. Radiol.* **48**, 20180218 (2019).
26. Mongan, J., Moy, L. & Kahn Jr, C. E. Checklist for artificial intelligence in medical imaging (claim): A guide for authors and reviewers (2020).
27. Tkachenko, M., Malyuk, M., Holmanyuk, A. & Liubimov, N. Label studio: Data labeling software (2020–2022). Open source software. https://github.com/heartexlabs/label-studio.
28. Cao, L., Zheng, X. & Fang, L. The semantic segmentation of standing tree images based on the yolo V7 deep learning algorithm. *Electronics (Basel)* **12**, 929 (2023).
29. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2014).
30. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. arXiv:1512.00567 (2016). Accessed 01 June 23.
31. Tan, M. & Le, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv:1905.11946 (2019). Accessed 01 June 23.
32. Ludlow, J. A manufacturer's role in reducing the dose of cone beam computed tomography examinations: Effect of beam filtration. *Dentomaxillofac. Radiol.* **40**, 115–122 (2011).
33. Signorelli, L., Patcas, R., Peltomäki, T. & Schätzle, M. Radiation dose of cone-beam computed tomography compared to conventional radiographs in orthodontics strahlenbelastung beim digitalen volumentomogramm im vergleich zu konventionellen kieferorthopädischen radiologischen unterlagen. *J. Orofac. Orthop.* **77**, 9–15 (2016).
34. Roberts, J., Drage, N., Davies, J. & Thomas, D. W. Effective dose from cone beam CT examinations in dentistry. *Br. J. Radiol.* **82**, 35–40 (2009).
35. Patcas, R., Signorelli, L., Peltomäki, T. & Schätzle, M. Is the use of the cervical vertebrae maturation method justified to determine skeletal age? A comparison of radiation dose of two strategies for skeletal age estimation. *Eur. J. Orthod.* **35**, 604–609 (2013).
36. Pearce, M. S. et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: A retrospective cohort study. *Lancet* **380**, 499–505 (2012).
37. Zhang, Y. et al. Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection. *Sci. Data* **10**, 380 (2023).
38. Li, J., Zheng, L., Daraqel, B., Liu, J. & Hu, Y. Treatment outcome of regenerative endodontic procedures for necrotic immature and mature permanent teeth: A systematic review and meta-analysis based on randomised controlled trials. *Oral Health Prevent. Dent.* **21**, b4100877 (2023).
39. Pallás, J. M. A. & Villa, J. J. *Métodos de investigación clínica y epidemiológica* (Elsevier Health Sciences, 2019).
40. Cvek, M. Prognosis of luxated non-vital maxillary incisors treated with calcium hydroxide and filled with Gutta-Percha. A retrospective clinical study. *Dent. Traumatol.* **8**, 45–55 (1992).

41. Krishan, K., Kanchan, T. & Garg, A. K. Dental evidence in forensic identification-An overview, methodology and present status. *Open Dent. J.* **9**, 250 (2015).
42. Plascencia, H. et al. Intra-and interobserver agreement during the assessment of the different stages of root development using 4 radiographic classifications. *J. Endod.* **47**, 906–913 (2021).
43. Cohen, M. D. Pediatric CT radiation dose: How low can you go?. *Am. J. Roentgenol.* **192**, 1292–1303 (2009).

Author contributions
Data were acquired and annotated by S.K.T. The experiments were carried out by H.O.I and G.S., while S.K.T and H.A. performed the data analysis. The manuscript was collaboratively drafted and reviewed by all authors.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.T.K. or H.O.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025