

ORIGINAL RESEARCH

Artificial intelligence for detecting dental ankylosis in primary molars using panoramic radiographs—a retrospective study

Nagehan Yılmaz^{1,*}, Mustafa Hakan Bozkurt^{2,3}, Tamer Tüzüner¹, Musa Aslan⁴

¹Department of Pediatric Dentistry, Faculty of Dentistry, Karadeniz Technical University, 61080 Trabzon, Turkey

²Faculty of Engineering, Artificial Intelligence and Data Engineering, Karadeniz Technical University, 61080 Trabzon, Turkey

³Trabzon Teknokent, Mavera Information Technologies, 61080 Trabzon, Turkey

⁴OF Faculty of Technology, Software Engineering, Karadeniz Technical University, 61830 Trabzon, Turkey

***Correspondence**

nagehanyilmaz@ktu.edu.tr
(Nagehan Yılmaz)

Abstract

Background: Dental ankylosis is an eruptive abnormality that requires early diagnosis to prevent complications. This study investigated the usability and performance of various deep learning models (including transfer learning, which enhances model performance by utilizing pre-trained networks) for ankylosis detection in dental X-rays. **Methods:** Classical convolutional neural network (CNN) method, Visual Geometry Group 16-layer (VGG16), Inception V3, and MobileNet V2 deep learning models were used for classification. In total, 268 panoramic radiograph images were diagnosed: 98 as ankylosis cases and 170 as controls, with ages ranging from 4 to 15 years. Various data augmentation techniques were employed. Accuracy, sensitivity, specificity, Area Under Curve (AUC) and F1-Score metrics were assessed to evaluate the performance of the models. **Results:** The CNN network without pretraining proved insufficient, leading to the adoption of transfer learning. The accuracy, AUC, sensitivity, specificity and F1-Score values of all three models can be used, but the VGG16 and Inception V3 models generally outperformed the MobileNetV2. Based on accuracy and specificity, Inception V3 demonstrated better classification performance, while VGG16 demonstrated a more balanced performance. **Conclusions:** This study highlights the effectiveness of deep learning models, particularly VGG16, in identifying ankylosis from panoramic radiographs, emphasizing the importance of model selection for improved diagnostic outcomes.

Keywords

Ankylosis; Artificial intelligence; Deep learning; Transfer learning

1. Introduction

Dental ankylosis is an eruptive abnormality, histologically characterized as a fusion between dentin or cementum and bone resulting in the obliteration of the periodontal ligament space [1]. The fusion of bone and the tooth root leads to a vertical halt in tooth eruption, resulting in infraocclusion or possibly impaction of the affected tooth. This condition may occur during tooth eruption at any stage, whether before full emergence or after arrival at the occlusal plane [1, 2]. A significant incidence of dental ankylosis has been documented among children aged seven to eleven, with rates ranging from 1.3% to 38.5% [3, 4].

Two major theories explain dental ankylosis' etiology: the first emphasizes local factors, while the second focuses on genetic factors, although no genes have yet been identified [5]. Abnormal activation of the periodontium in mice, which is known to increase Wnt signaling, may influence cancer stem cell behavior. A correlation has also been proposed between high levels of cellular cementum and alveolar bone accumulation and ankylosed teeth development [6].

The early detection of dental ankylosis is based on clinical (the presence of infraocclusion of the respective tooth, percussion testing, and loss of tooth mobility) and preclinical findings. Because ankylosis of primary molars is associated with severe clinical outcomes, such as infraocclusion of tooth, tipping of neighbouring teeth into infraocclusion space, which may cause arch space loss, exfoliation and eruption impairments for permanent successors and an asymmetric dental pattern [1, 7].

The ankylosed tooth is unable to exert the post-eruptive motion, thus it could be located in infraocclusion from 1 mm to complete retention beneath the gingival tissue [1, 8]. In severe ankylosis cases, percussion testing and assessment of tooth mobility are not applicable [7]. Therefore, radiographic assessment is crucial for diagnosing ankylosis. Bone and root surfaces will be visible on X-ray images, as well as the absence of periodontal space [1, 7].

Panoramic radiography is a vital imaging tool in dental practice for diagnosing and planning treatment for dental and maxillofacial conditions. A wide range of advantages are associated with this procedure, including its simplicity, afford-

ability, minimal radiation exposure, reduced patient discomfort, and also the capacity to display a comprehensive bilateral view of oral structures [9, 10]. Further, interpreting dental radiographs in children can be more challenging and time-consuming. The mixed dentition, low-density areas in dental follicles, and possible pathological periapical inflammation may all be contributing factors [11–13].

Thus, disease identification can be transformed into a standardized image categorization task through artificial intelligence (AI), leveraging machine algorithms endowed with reasoning capabilities and cognitive functions, particularly relying on deep learning [10, 11]. The medical field is rapidly utilizing such technology. Particularly, image analysis using convolutional neural networks (CNNs) has demonstrated the potential to enhance practitioners' reliability and CNNs learn from the statistical patterns present in imagery by iteratively processing pairs of images and corresponding image labels, typically supplied by medical professionals. Eventually, CNNs become capable of evaluating unseen data [14].

Artificial intelligence has the potential to reduce dentists' clinical workload and enhance diagnostic accuracy, enabling earlier detection and more effective treatment alternatives, while mitigating errors stemming from fatigue and facilitating timely intervention, thereby conserving time [9, 11]. Deep learning-driven artificial intelligence algorithms have largely replaced specific conventional machine learning tasks in computer vision, such as classification, segmentation, and detection [10, 11, 15]. AI has been used in dentistry for a variety of purposes and across a wide range of modalities [10, 16–19]. This includes detecting caries using bitewing or periapical radiographs [11, 17, 20], determining ectopic eruption [11] and maxillary sinusitis with panoramic radiographs [21], diagnosing osteoarthritis on cone-beam computed tomography (CBCT) [22], identifying lateral mandibular deviation through cephalograms [18], and detecting impacted supernumerary teeth [16]. Additionally, AI has effectively identified white spot lesions [23] and hypoplastic lesions on photographic images [24].

Dental ankylosis is challenging to diagnose. Although panoramic radiography is essential for its detection, image interpretation can be complex [7, 11]. Furthermore, traditional radiographic interpretation is inherently subjective and susceptible to human error, influenced by fatigue and experience levels. A missed or delayed diagnosis of dental ankylosis can lead to more severe clinical outcomes and potentially more complex and costly interventions later on [25, 26]. The use of AI in medical imaging, particularly deep learning models, has shown promise in improving diagnostic accuracy and efficiency [11, 15]. We identified a significant gap in the literature and in clinical practice regarding the use of AI for detecting dental ankylosis, particularly in primary molars using panoramic radiographs. This gap highlights the timeliness and relevance of this study.

This study aimed to address the limitations of traditional diagnostic methods such as time constraints, high costs, and potential human errors. The objective was to develop an accurate deep learning model for detecting ankylosed primary molars from panoramic radiographs.

In our study, the null hypothesis (H_0) posited that there is

no significant difference between the performance of our AI model and chance performance. Thus, if H_0 is true, the model's predictive ability is no better than random guessing. To test this hypothesis, we employed several evaluation metrics, including accuracy, AUC (Area Under the ROC (Receiver Operating Characteristic) Curve), sensitivity, and F1-Score, as previously used in deep learning or machine learning studies [10, 11, 16, 27, 28]. Each metric is statistically connected to H_0 by setting up confidence intervals and performing hypothesis tests.

2. Materials and methods

This retrospective study was approved by the Scientific Research Ethics Committee of Karadeniz Technical University Faculty of Dentistry (Protocol No. 2023/18, dated 19 June 2023). Given the retrospective nature of the analysis, informed consent from individual patients was not obtained. However, to ensure confidentiality and compliance with ethical standards, all patient data was anonymized before analysis. To ensure privacy, personal identifiers were removed, and data were aggregated. A retrospective analysis was conducted on pediatric patients (aged 4 to 15) who had panoramic radiographs taken between January 2015 and August 2023 at the Karadeniz Technical University Faculty of Dentistry (Trabzon, Turkey). Ankylosis was determined by analyzing panoramic X-ray images during their classification to determine the presence or absence of ankylosis (Fig. 1).

All panoramic radiograph images were taken with the Sirona Orthophos XG3 (Dentsply Sirona, Bensheim, HE, Germany) panoramic device (62 kV, 8 mA, 14.1 seconds) and saved in Joint Photographic Experts Group (JPEG) format. Sirona Orthophos XG3 (Dentsply Sirona, Bensheim, Germany) panoramic radiography device used in this study provides high-resolution imaging and is widely used in dentistry practices. The device uses a 0.5 mm X-ray tube and digital sensor technology to provide clear and detailed panoramic images. In addition, there is a special pediatric mode that reduces radiation dose for pediatric patients and allows detailed examination of teeth and jaw structures with various imaging programs [29]. Exclusion criteria in the study included radiographs with excessive blurriness, inadequate exposure, motion artifacts, or obscure or unclear anatomical structures. Alternatively, radiographs of children aged 4 to 15 years containing ankylosed primary molars met the inclusion criteria. Children who did not exhibit dental ankylosis were designated as controls (Table 1). The VGG Image Annotator (VIA) program was used to label teeth identified with dental ankylosis.

This table illustrates the gender and age distribution of patients in the study who had radiographs with and without ankylosed teeth.

Sample size was calculated based on a previous study [11] with alpha error = 0.05 and beta error = 0.20. Thus, it was decided to use at least 200 images. To avoid data loss (such as poor-quality images), at least 240 panoramic radiographs were used. Ultimately, due to the limited number of ankylosed primary teeth cases between January 2015 and August 2023, 268 panoramic radiographs were used: 98 cases and 170 controls.

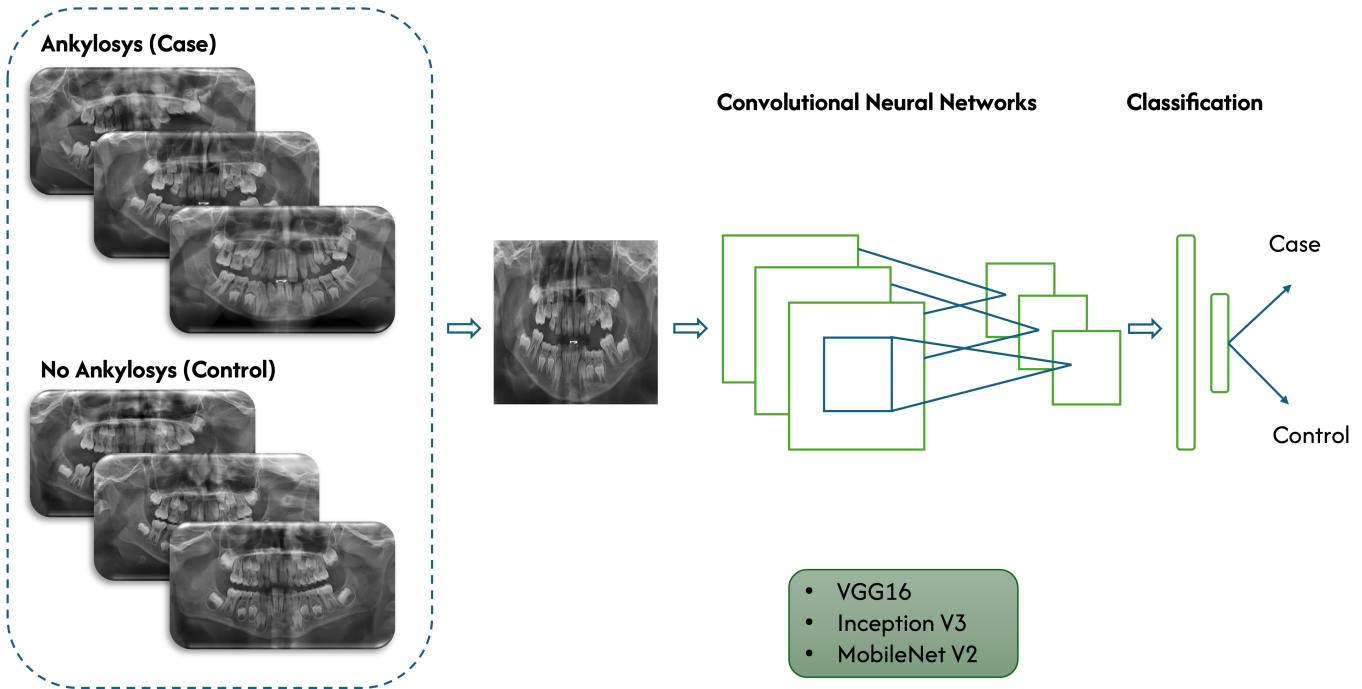


FIGURE 1. Decision flow of the system (the relevant figure illustrates the overall workflow of the study methodology).
VGG16: Visual Geometry Group 16-layer.

TABLE 1. Gender and age distribution of cases

Cases	Gender		Age (mean $\pm$ sd)
	Girl N (%)	Boy N (%)	
Ankylosed	52 (53.1)	46 (46.9)	8.13 $\pm$ 1.751
Non-ankylosed	81 (47.6)	89 (52.4)	7.18 $\pm$ 1.605

sd: Standard Deviation.

Candidate digital panoramic radiographic images were screened based on medical records. Two experienced pediatric dentists, one with 6 years of experience (NY) and the other with 16 years of experience (TT) independently classified the images as cases or controls using the same monitor and environment. When the two dentists could not reach a consensus on a diagnosis, despite discussions, the image was excluded.

Due to the relatively limited number of X-rays in the target dataset, various data augmentation techniques were applied. Initially, a network structure was created and evaluated using the classical CNN (AT&T Bell Labs, Murray Hill, NJ, USA) method. Although data augmentation was applied, CNN performed relatively poorly. Therefore, transfer learning was used to improve performance. For the purpose of transfer learning, the performance of well-known and current neural network models, such as Inception V3 (Google, Mountain View, CA, USA), MobileNet V2 (Google, Mountain View, CA, USA) and Visual Geometry Group 16-layer (Visual Geometry Group-VGG, Oxford, UK) (VGG16), which had previously demonstrated high performance, was examined (Fig. 2) [30–32]. Images identified as ankylosis and those known to be non-ankylosis were processed sequentially for data augmentation and preprocessing. Then, classification was performed using

deep learning models. Finally, the classification performance was evaluated.

Experiments were conducted using Python (Python Software Foundation, USA) programming language on an i5 processor computer with Windows (Microsoft Corporation, Redmond, WA, USA) operating system.

2.1 Data preprocessing

Panoramic dental X-ray images contain extraneous spaces or bones visible outside the jaw. Therefore, a cropping process was applied to the left-right and top-bottom regions to obtain an image where only the jaw structure is visible. This eliminated unwanted areas and provided a clearer image. Visual examples of the cropping parameters used, along with before-and-after images, would enhance clarity and aid reproducibility (Fig. 3). Image pixels were normalized between [0–1] (Pnorm).

$$P_{norm} = \frac{P - P_{min}}{P_{max} - P_{min}}$$

80% of the cropped images have been designated for training (10% for validation and 10% for testing). Separate datasets for training, validation, and testing were created. Data augmentation was then applied separately to each dataset. This ensured

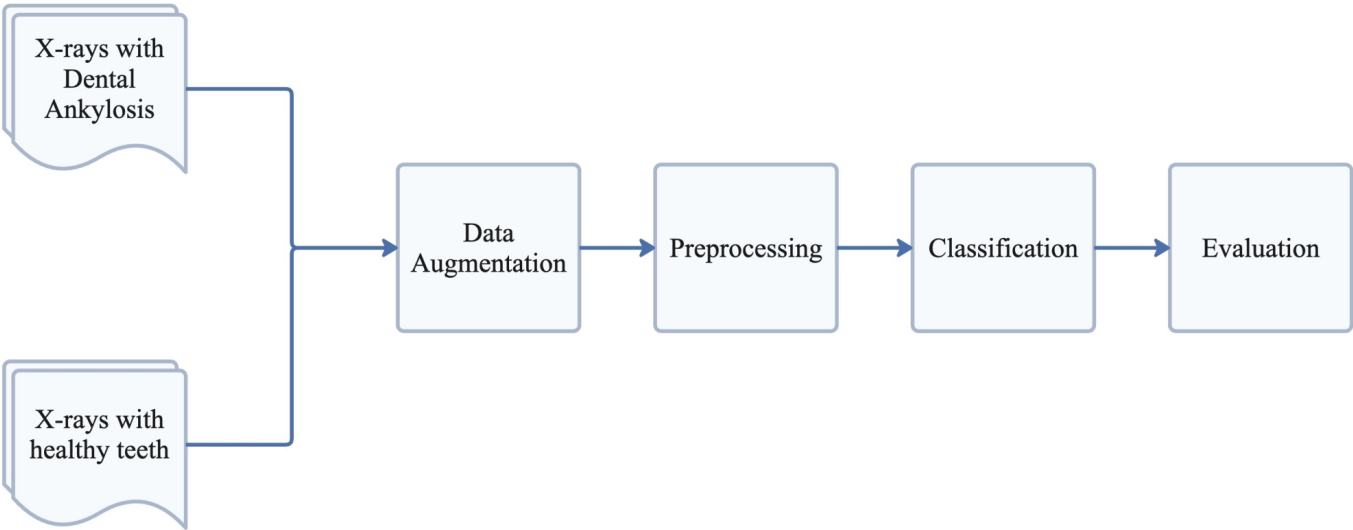


FIGURE 2. Steps of the ankylosis detection methodology.

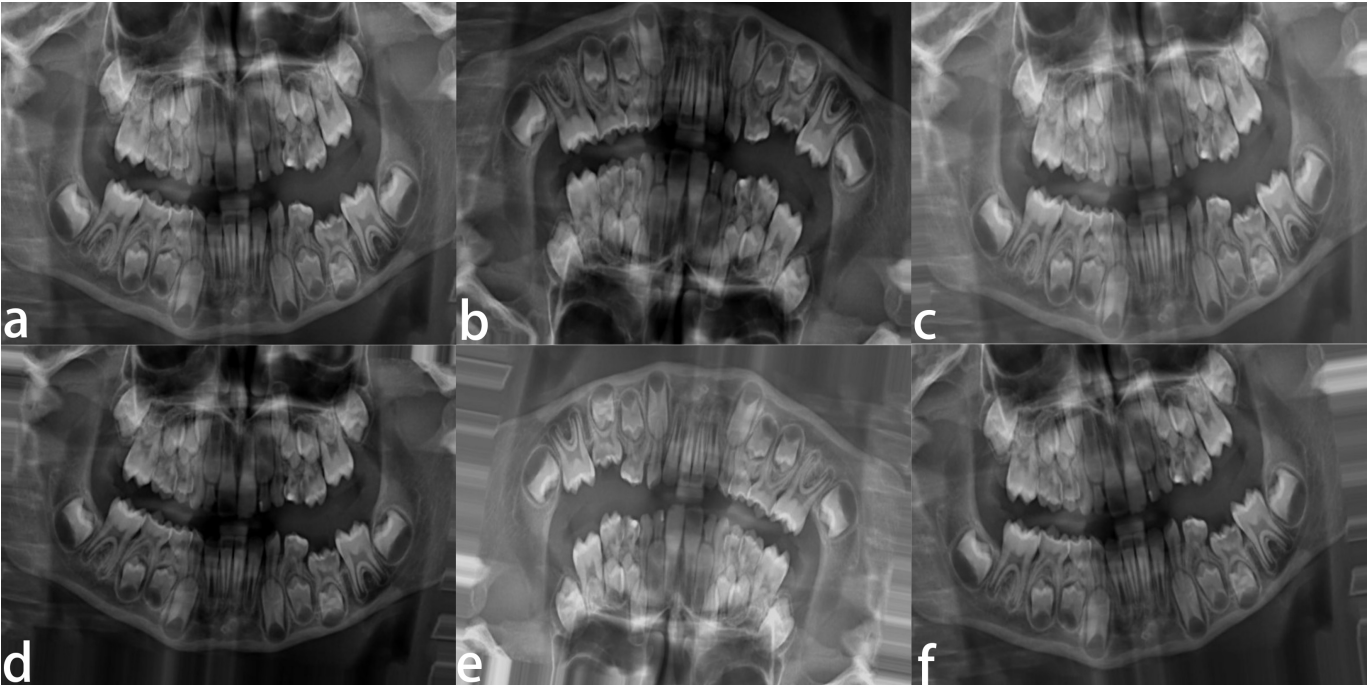


FIGURE 3. Examples of images generated randomly for data purposes. (a) Randomly generated image examples for data augmentation copy not augmented. (b) Randomly generated image examples for data augmentation copy vertical flipped and rotated. (c) Randomly generated image examples for data augmentation copy channel shifted and rotated. (d) Randomly generated image examples for data augmentation copy vertical shifted and rotated. (e) Randomly generated image examples for data augmentation vertically flipped and channel shifted, and rotated. (f) Randomly generated image examples for data augmentation vertically shifted and rotated.

that augmented images from the training dataset did not appear in the validation and test sets. The data augmentation process (horizontal shift, random vertical shift, shear transformation, random zoom, horizontal flip, vertical flip, channel shift) was performed to generate 20 new images from each original image.

The augmented images were resized to 156×156 pixels. The choice of pixel size was based on both model performance and processing efficiency. Resizing standardizes image size, making them easier to process. Additionally,

separate lists were created for training, validation, and testing: `train_images`, `train_labels`, `valid_images`, `valid_labels`, `test_images` and `test_labels` [9–11]. The data was shuffled to prevent memorization. This improved the model’s ability to generalize, leading to better results.

2.2 Deep learning models for dental ankylosis classification

This study employed two learning methods: training from scratch and transfer learning. Alongside the CNN method

trained from scratch, the transfer learning method was adapted to three different neural network models (VGG16, Inception V3 and MobileNet V2) for categorizing panoramic radiographic images to detect ankylosed or non-ankylosed cases [30–32].

The VGG16 architecture consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. A Red-Green-Blue (RGB) image with a size of 224×224 pixels is input to the network. The input image is passed through a series of convolutional layers. VGG16 uses 3×3 filters for each convolution and applies the Rectified Linear Unit (ReLU) activation function after each convolution. Pooling layers are applied after convolutional layers to reduce dimensions. The feature maps obtained from the convolutional and pooling layers are connected to fully connected layers. These layers perform high-level feature extraction and produce classification output [30].

Inception V3 is a deep architecture with 48 initial layers, consisting of four convolutional layers with activation functions and two max-pooling layers. The input layer of the Inception V3 model takes an image with a size of 299×299 pixels and RGB color channels. Convolutional layers are employed to extract different features from the image. The core structure of Inception V3 is the Inception modules. These modules combine different sizes and types of convolutions (1×1 , 3×3 , 5×5) and max pooling layers. This enables

the network to detect features at varying scales. In addition, Inception V3 employs global average pooling as a pooling method. In the later stages of the network, there are fully connected neural network layers and an output layer where classification results are obtained [31–33].

MobileNet V2 accepts an image of 224×224 pixels as input. The input image passes through several convolutional layers with different filter numbers. A non-linear activation function is applied after each convolutional layer. These layers extract low-level features from the image. MobileNet V2 incorporates bottleneck layers, which feature fewer filters followed by 1×1 convolutions that reduce dimensions. This reduces the number of parameters and computational costs while maintaining representational power. After the bottleneck layers, expansion layers with a higher number of filters are added. These layers use 1×1 convolutions to expand the dimensions of the features [32] (Table 2, Ref. [30–32]).

This table provides structural summaries of the artificial intelligence models used in the study.

In the study the model's performance was assessed using each of metrics (accuracy, AUC, sensitivity, and F1-Score) to comprehensively assess whether to reject H_0 . An accuracy of 50% in binary classification indicates that the model performs no better than random chance [34].

TABLE 2. Contents of model [30–32].

Model Structure Summary	
CNN	• Sequential model with 2 Conv2D layers (256 filters in the first, 128 filters in the second) • Batch Normalization after each Conv2D layer • MaxPooling after each Conv2D layer • Dropout layers with a rate of 0.7 after each MaxPooling layer • Flatten layer • Dense layers (32, 64 and 1 neuron respectively)
	• VGG16 Base Model with imagenet weights • Flatten layer • Dense layer with 1024 neuron • Dropout layer • Dense layer (output layer)
	• Inception V3 Base Model with imagenet weights • Global average pooling • Flatten layer • Dense layer with 32 neuron • Dropout layer • Dense layer with 32 neuron • Dropout layer • Dense layer (output layer)
	• MobileNet V2 Base Model with imagenet weights • Global average pooling • Flatten layer • Dense layer with 32 neuron • Dropout layer • Dense layer with 32 neuron • Dropout layer • Dense layer (output layer)
VGG16	
Inception V3	
MobileNet V2	

CNN: Classical convolutional neural network; VGG16: Visual Geometry Group 16-layer.

2.3 Evaluation metrics

In the study, accuracy, Area Under Curve (AUC), sensitivity, specificity, and F1-score were used to evaluate the model's performance. For each sample, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values were calculated based on the predicted class. Accuracy (Acc) was measured as the ratio of correctly identified samples to the total number of samples [10, 11, 16].

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}$$

Area Under Curve (AUC) represents the area under the Receiver Operating Characteristic (ROC) curve. To evaluate the classification model's performance, an AUC was determined; the closer it is to 1, the better the model is. The sensitivity formula (Sensitivity = $TP/(TP + FN)$) was used to measure the ratio of true positives (TP) to total positives (TP + FN). The specificity value was utilized as a measure of how many negative class examples were correctly predicted (Specificity = $TN/(TN + FP)$). A harmonic mean of precision and recall was calculated using the F1-score formula (F1-score = $TP/(TP + 1/2[FP + FN])$) [10, 11, 16]. In this study, x-ray images with ankylosis were labeled 1, and those without ankylosis were labeled 0.

3. Results

3.1 Loss and accuracy of the models' results

When examining the training loss and accuracy graphs for CNN, it was understood that the training loss value consistently decreased and converged to a loss value of around 0.25. Similarly, the training accuracy value steadily increased and converged to around 0.90. This indicates that the training data was well learned. A well-trained model was also shown in the loss and accuracy graphs for the VGG16 training set. How-

ever, after the 10th epoch, the loss and accuracy values did not significantly improve for the validation value. Additionally, Inception V3's loss and accuracy graphs also indicated the model was well trained. However, after the 10th epoch, the loss and accuracy values did not significantly improve. The loss and accuracy graphs for the MobileNet V2 training set showed that the model learned the training data well (Fig. 4).

A consistent decrease in loss and accuracy was not observed in the CNN validation set. Upon reaching the 25th epoch, the loss value increased, while the accuracy value decreased. In this case, overfitting began after the 25th epoch. Therefore, the best validation accuracy was 0.71. VGG16 validation accuracy and loss values did not consistently decrease after the first 10 epochs and even deteriorated. In this case, the model overfitted the training data. Therefore, the network model at the best point was saved. For Inception V3, the validation accuracy and loss values showed irregularities. While the training data consistently improved, the validation graph suggested that the Inception V3 model might have overfitted. Therefore, the network model at the best point was saved.

For MobileNet V2, instability in loss and accuracy values was observed in the validation data after the 5th epoch. The accuracy and loss values did not consistently decrease after the first 10 epochs and even showed adverse progress in some epochs. In the last 2 epochs, an increase was observed (Fig. 5). Essentially, no significant improvement in validation performance was noticed after the epoch of increasing gaps between training and validation, which indicates overfitting. This study employed the dropout method to prevent overfitting. To reduce the overfitting tendency of the model, the dropout value was initially set at 0.7 in all models. However, different dropout values (0.3, 0.5, 0.7) were tested and the most suitable value was determined.

3.2 Classification of models

Sensitivity, specificity, accuracy, AUC and F1-Score results obtained for all models used in the study are shown in Table 3.

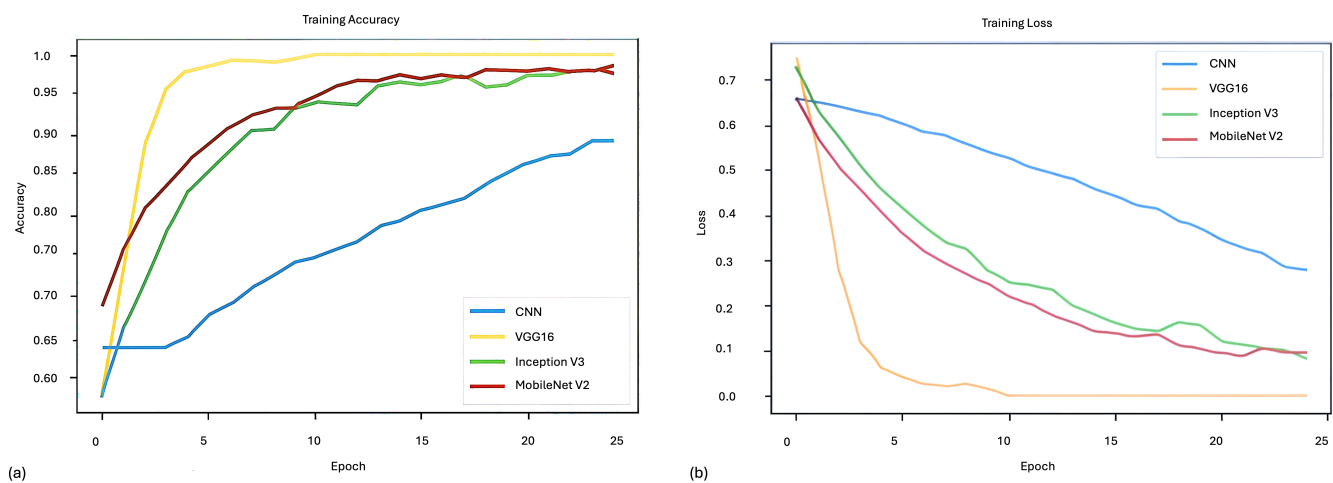


FIGURE 4. Training accuracy and loss figure. (a) Training accuracy of each model; (b) Training loss of each model. The training accuracy and loss plots show that VGG16 and Inception V3 reach faster convergence and higher accuracy, whereas CNN demonstrates the lowest performance among the tested models. Notably, VGG16 achieves the best training results, featuring the highest accuracy and the lowest loss. CNN: Classical convolutional neural network; VGG16: Visual Geometry Group 16-layer.

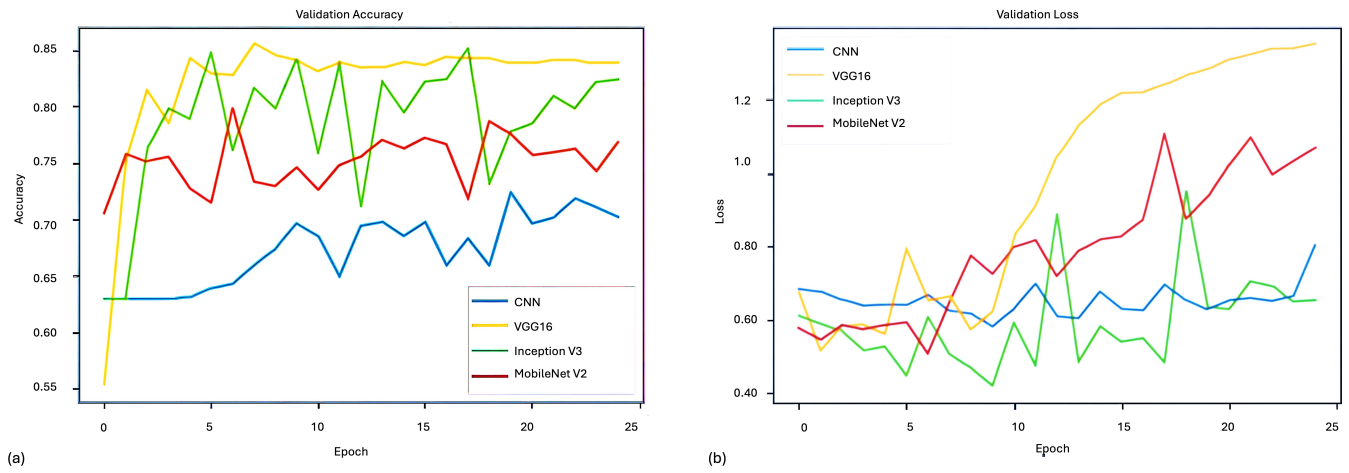


FIGURE 5. Validation accuracy and loss figure. (a) Validation accuracy of each model; (b) Validation loss of each model. The validation accuracy plot indicates VGG16 achieves the highest accuracy, but the validation loss plot reveals VGG16 also exhibits increasing loss, suggesting overfitting; in contrast, CNN shows the lowest validation accuracy among the models. CNN: Classical convolutional neural network; VGG16: Visual Geometry Group 16-layer.

TABLE 3. Comparison of the models according to accuracy, AUC score and specificity metrics.

	CNN	VGG16	MobileNet V2	Inception V3
Sensitivity	0.28	0.89*	0.77	0.79
Specificity	0.96	0.91	0.92	0.98*
Accuracy	0.71	0.90	0.86	0.91*
AUC	0.79	0.95	0.88	0.96*
F1-Score	0.42	0.87*	0.81	0.86

**It denotes the maximum values across all rows. AUC: Area Under Curve; CNN: Classical convolutional neural network; VGG16: Visual Geometry Group 16-layer.*

An analysis of the ROC curve (Fig. 6) and Table 3 values revealed that non-ankylosed images were well-detected, but the CNN model was found less effective at detecting ankylosed images. The CNN model detected ankylosis with a 0.28 sensitivity value.

Apart from CNN, sensitivity/specificity/accuracy/AUC/F1-Score values in Table 3 were differentiated as followed: VGG16 (0.89/0.91/0.90/0.95/0.87), Mobile NetV2 (0.77/0.92/0.86/0.88/0.81) and Inception V3 (0.79/0.98/0.91/0.96/0.86). According to the highest and lowest values for each metric, the sensitivity values ranged from VGG16 (0.89) to CNN (0.28), the specificity values from Inception V3 (0.98) to VGG16 (0.91), the accuracy values from Inception V3 (0.91) to CNN (0.71), the AUC values from Inception V3 (0.96) to CNN (0.79), and the F1-Score values from VGG16 (0.87) to CNN (0.42). All the highest values when comparing the metrics among the groups were summarized in Table 3.

The table reveals the highest and lowest values for each metric: sensitivity values range from the lowest in CNN to the highest in VGG16, specificity values vary from Inception V3 as the highest to VGG16 as the lowest, accuracy values go from Inception V3 at the highest to CNN at the lowest, AUC values also range from Inception V3 as the highest to CNN as the lowest, and F1-Score values vary from VGG16 at the

highest to CNN at the lowest (Higher value indicates better performance).

4. Discussion

Dental ankylosis is an eruption anomaly characterized by the fusion of the tooth root and alveolar bone, resulting in the obliteration of the periodontal ligament space. This condition is particularly prevalent among children aged 7 to 11 years, and early diagnosis is crucial to prevent serious complications such as vertical bone loss, dental asymmetry, and loss of arch space [1–5, 7].

Dental experts recommend combining clinical and radiographic examinations for ankylosed primary teeth diagnosis, with panoramic radiographs considered a routine approach [1]. Previous studies reported that CNNs can classify and segment periapical, bitewing, CBCT, and panoramic radiographs to detect caries and anatomical structures, with panoramic radiography being the most widely used technique [35, 36].

Due to the advantages of AI technologies, dentists are increasingly using them for diagnosis, which improves diagnostic efficiency and enables earlier diagnosis and intervention [37]. Integration of AI-enabled systems into clinical applications optimizes workflow and enhances cost-effectiveness in dental practices. With AI-based imaging and diagnostic systems, the time spent on each patient can be reduced, improv-

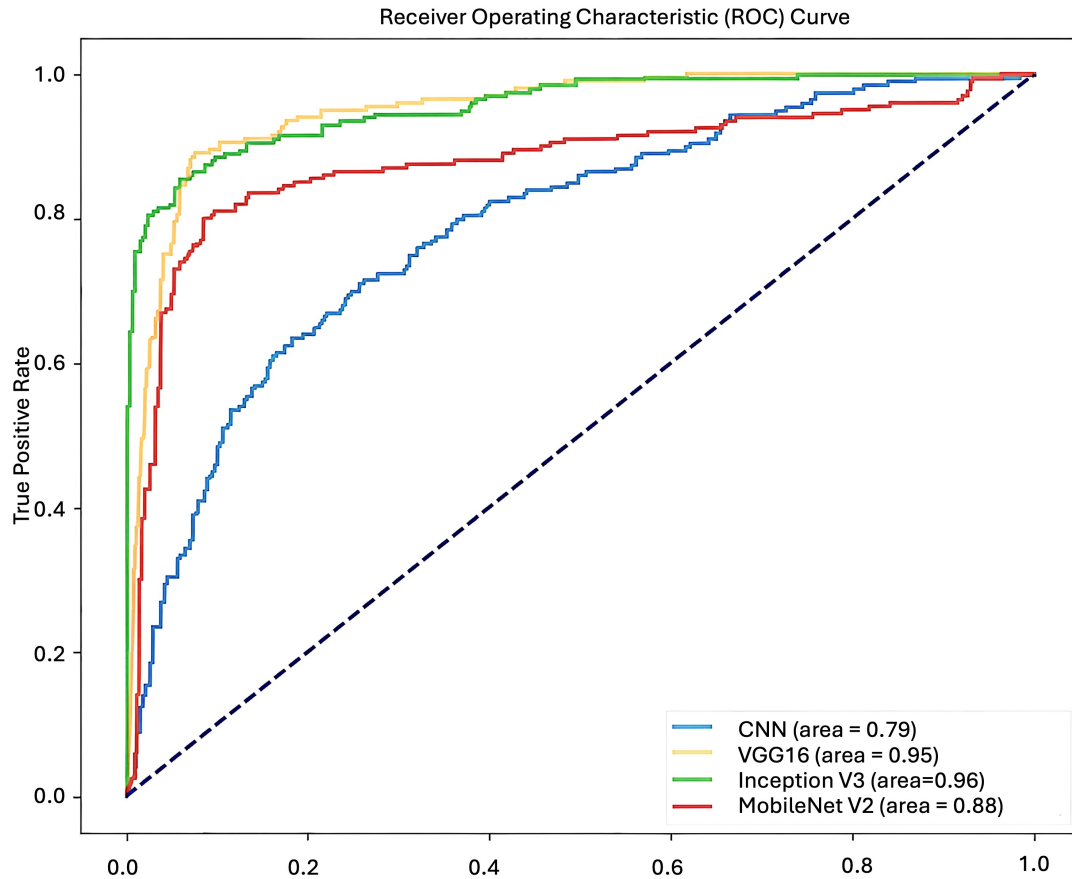


FIGURE 6. ROC curve and AUCs of the models. The validation accuracy plot shows that VGG16 attains the highest accuracy; however, the validation loss plot indicates that VGG16 also experiences increasing loss, signaling potential overfitting. In contrast, CNN exhibits the lowest validation accuracy among all the models. CNN: Classical convolutional neural network; VGG16: Visual Geometry Group 16-layer.

ing clinical efficiency. Furthermore, AI-based applications in specific areas such as caries detection reduce treatment costs by minimizing missed lesions. AI integration also streamlines administrative tasks like scheduling appointments, updating patient records, and responding to routine inquiries. By automating these processes, clinical staff can concentrate on addressing more complex patient needs, ultimately saving time and resources [38].

To determine the best AI model, various models' suitability has been tested against study methods and disease diagnosis criteria. While more advanced models exist than those used in this study, they require higher computational power. This study used models that can be easily applied to clinical settings and could potentially improve their performance with the addition of more images [10, 11, 39].

Dentists' decision-making during diagnosis has been facilitated using machine learning, with deep learning methods being employed to detect dental caries, supernumerary teeth, dental anomalies, and other dental conditions [10, 11, 40]. In this study, the effectiveness of AI, specifically the deep learning approach, in detecting ankylosis on dental panoramic x-ray images was evaluated. Aside from the sensitivity and F1-Scores of CNN, the H_0 was rejected (when considering 50% as borderline) indicating that AI-assisted deep learning models performed better than random predictions.

In earlier studies, a Region of Interest (ROI) was extracted from the original X-ray image for the deep learning model to predict diagnosis [11, 16, 20]. In our study, cases of ankylosis are localized to specific regions within the images, suggesting that training on the entire image set may not effectively learn the diagnosis of ankylosis. Therefore, during deep learning, cropping was performed on both ankylosed and non-ankylosed images to focus on ROIs. This approach aims to enable deep learning to easily identify the most relevant regions for classification, thereby enhancing the model's ability to accurately detect cases [11, 41].

It is often easier to train direct classification systems than segmentation, which requires human experts to delineate contours or record coordinates for every sample. However, human annotation can introduce errors and biases. Probability estimates in each category can be affected by these errors. Despite significant differences in diagnostic probabilities, a positive diagnosis could still be made [11, 42]. However, in comparison to segmentation-based models, classification-based models can be further improved by retraining them with manually labeled data [11].

In this study, similar to previous research, four different structure models—CNN, VGG16, Inception V3 and MobileNet V2 were used, and metrics such as accuracy, AUC score, sensitivity, specificity, and F1-Score were

evaluated [10, 11, 16, 35, 39, 43]. As observed in our study, training loss values decreased, which is consistent with the results of a study using deep learning algorithms to detect supernumerary teeth in dentistry [10]. Thus, based on the training accuracy and loss values obtained in this study, it was concluded that all models could be generally trained to acceptable levels. However, the validation loss and accuracy values differed from those reported in the literature [10]. Alongside these differences, overfitting was observed in all models at various epochs, with optimal validation values identified, but no significant improvement in validation performance after the onset of overfitting epochs. Study designs, sample types, and sizes may explain the variability in results observed in this study compared to previous findings [10].

Sensitivity is considered a crucial measure in disease screening to avoid false-negative diagnoses [10]. As observed by the sensitivity results, CNN exhibited the lowest value (0.28) compared to VGG16 (0.89), MobileNet V2 (0.77) and Inception V3 (0.79). Because of the low sensitivity of the condition, there is a risk of missing truly ankylosed cases when using the CNN model.

Findings in the dental literature report sensitivity values for VGG16 at 0.58, 0.74 and 0.85 and specificity values at 0.70, 0.79 and 0.83. For Inception V2 and/or V3, sensitivity values were reported at 0.82 and 0.90 and a specificity value at 0.78. For MobileNet V2, sensitivity values were 0.62 and 0.89, and specificity was 0.85 [10, 11, 23, 39, 44]. Similarly, our sensitivity and specificity findings were consistent with previous findings [10, 11, 23, 39, 44]. According to Mine *et al.* [10], when testing for presence of supernumerary teeth in early mixed dentition using deep learning, VGG16 had the highest sensitivity value (0.85), with all models, including AlexNet, VGG16, and Inception V3, showed high sensitivity values. In our study, VGG16 model had a higher sensitivity value (0.89) than Inception V3 (0.79). Also, MobileNet V2 showed a slightly lower value (0.77) than Inception V3. To avoid missing true positives, it is advantageous to use Inception V3, specifically VGG16, in comparison to MobileNet V2.

Higher accuracy values indicate the model's overall ability to produce correct results [16, 36] and higher AUC values demonstrate the model's ability to perform effective classification [45]. In this study, Inception V3 exhibited slightly higher accuracy (0.91) and AUC (0.96) values than VGG16 (0.90 and 0.95), respectively. MobileNet V2 (0.86 and 0.88) and CNN (0.71 and 0.79) exhibited lower values than both Inception V3 and VGG16. Compared to MobileNet V2 and CNN, Inception V3 and VGG16 may offer better detection of true positive and negative cases.

A high F1-Score value indicates the model's overall high performance and suggests balanced classification [11]. Overall, VGG16 achieved the highest performance in this metric, while Inception V3 showed very close F1-Score performance. VGG16 and Inception V3 models are both capable of classifying ankylosed and non-ankylosed images with high performance and balance. VGG16 and Inception V3 models generally excel in performance over MobileNet V2 due to their more complex architectures. These deeper models, however, may be overfitted due to limited data sizes. Even though

MobileNet V2 is slightly less performant, it is a suitable option for scenarios that require a lighter model.

According to the literature, VGG16 achieved AUC values of 0.73, accuracy values of 0.57, 0.72, 0.89, and a F1-Score of 0.56; Inception V2 and/or V3 achieved accuracy values of 0.80 and 0.88; and MobileNet V2 achieved accuracy values of 0.62 and F1-Scores of 0.57 and 0.87 [10, 11, 16, 39, 44]. Based on the combined results, the AUC and F1-Scores are aligned with the literature, indicating that VGG16 and Inception V3 could be used to detect ankylosed primary teeth from panoramic images. However, other studies using similar models have reported metric rankings that differ from ours [16, 39]. Model performance may vary based on research designs, explaining the discrepancy. Future developments in AI-assisted diagnostic models for dental ankylosis could greatly benefit from hybrid approaches and ensemble learning techniques. By combining the strengths of different deep learning architectures, hybrid models can enhance diagnostic accuracy by leveraging complementary feature extraction capabilities. Likewise, ensemble learning techniques that aggregate predictions from various models can enhance classification reliability and minimize model bias. By integrating these techniques, AI-based diagnostic systems can deliver more reliable and consistent results and ultimately aid clinical decision-making in pediatric dentistry. These approaches should be assessed in future studies in larger datasets and real-world clinical settings to optimize their applicability [46, 47].

For AI systems to be used effectively and reliably, clinical staff must undergo appropriate training and calibration processes. Dentists and auxiliary staff should be trained on AI-based tools principles, limitations and interpretation skills. Training ensures that AI systems are used effectively and safely to ensure patient safety and diagnostic accuracy. In addition, AI models need regular calibration and updates to perform optimally. This process ensures the models are constantly updated, ensuring accuracy and adaptability. AI integration into dental practices requires both technology infrastructure and the human factor to work in harmony [48].

Finally, this study explores the use of artificial intelligence to detect dental ankylosis in primary molars through the analysis of panoramic radiographs, highlighting its potential for early diagnosis, enhanced accuracy, and improved efficiency in dental care. However, this study has various limitations, including diagnosing ankylosis solely based on panoramic radiographic images, increasing case numbers statistically, and overfitting in some models. Possible biases in this study may be due to specific patient demographics, insufficient data diversity in data used in model training, variability in expert labeling, and potential model tuning biases. The limited sample size and use of panoramic radiographs alone could negatively influence dental ankylosis diagnosis. Future research should utilize three-dimensional imaging methods such as CBCT and larger sample sizes for conditions such as ectopic eruptions or hypoplasia lesions. Such adaptations could further enhance diagnostic accuracy and efficiency across a wide spectrum of dental and maxillofacial anomalies, providing valuable tools for practitioners and addressing gaps in current clinical practices. The versatility and relevance of the model extend beyond dental ankylosis in this suggestion.

5. Conclusions

Artificial intelligence (AI) methodologies offer significant potential in pediatric dentistry by reducing clinical workload and enhancing diagnostic accuracy, particularly in the early detection of dental ankylosis. This study emphasizes the effectiveness of deep learning models—specifically VGG16, Inception V3, and MobileNet V2—in identifying ankylosis from panoramic radiographic images. Compared to CNN, VGG16 showed relatively greater sensitivity and accuracy in detecting ankylosed images, highlighting the importance of model selection in dental diagnosis. Based on these findings, applying advanced AI algorithms to diagnostic workflows could enhance precision and patient outcomes by enabling earlier interventions. The limitations of certain models, such as CNN, should also be considered by dental practitioners, who may need to further optimize these models for reliable clinical use.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are available on request from the corresponding author. The data are not publicly available for ethical purposes.

AUTHOR CONTRIBUTIONS

NY, MHB and TT—conceived the ideas. NY, MHB and MA—collected the data. NY, MHB, MA and TT—analysed and checked the data; and led the writing.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The present retrospective study received approval from Scientific Research Ethics Committee of Karadeniz Technical University Faculty of Dentistry (Protocol No. 2023/18). The informed consent is exempted according to rules of the Scientific Research Ethics Committee of the Faculty of Dentistry at Karadeniz Technical University.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Eşian D, Bica CI, Stoica OE, Dako T, Vlasa A, Bud ES, *et al.* Prevalence and manifestations of dental ankylosis in primary molars using panoramic x-rays: a cross-sectional study. *Children*. 2022; 9: 188.
- [2] Yoon G. Eruption guidance of horizontally impacted permanent first molar with primary retention of primary second molars. *Journal of the Korean Academy of Pediatric Dentistry*. 2020; 47: 219–227.
- [3] Messer LB. Ankylosed primary molars: results and treatment recommendations from an eight-year longitudinal study. *Pediatric Dentistry*. 1980; 2: 37–47.
- [4] Noble J. Diagnosis and management of the infraerupted primary molar. *British Dental Journal*. 2007; 203: 632–634.
- [5] Hua L. To extract or not to extract? Management of infraoccluded second primary molars without successors. *British Dental Journal*. 2019; 227: 93–98.
- [6] Wu Y, Yuan X, Perez KC, Hyman S, Wang L, Pellegrini G, *et al.* Aberrantly elevated Wnt signaling is responsible for cementum overgrowth and dental ankylosis. *Bone*. 2019; 122: 176–183.
- [7] Ducommun F. Diagnosis of tooth ankylosis using panoramic views, cone beam computed tomography, and histological data: a retrospective observational case series study. *European Journal of Orthodontics*. 2018; 40: 231–238.
- [8] Kuroi J. Infraocclusion of primary molars: an epidemiologic and familial study. *Community Dentistry and Oral Epidemiology*. 1981; 9: 94–102.
- [9] Kaya E. Proposing a CNN method for primary and permanent tooth detection and enumeration on pediatric dental radiographs. *Journal of Clinical Pediatric Dentistry*. 2022; 46: 293–298.
- [10] Mine Y, Iwamoto Y, Okazaki S, Nakamura K, Takeda S, Peng TY, *et al.* Detecting the presence of supernumerary teeth during the early mixed dentition stage using deep learning algorithms: a pilot study. *International Journal Paediatric Dentistry*. 2022; 32: 678–685.
- [11] Liu J. Artificial intelligence-aided detection of ectopic eruption of maxillary first molars based on panoramic radiographs. *Journal of Dentistry*. 2022; 125: 104239.
- [12] Sams CM. Pediatric Panoramic radiography: techniques, artifacts, and interpretation. *Radiographics*. 2021; 41: 595–608.
- [13] Sabbadini GD. A review of pediatric radiology. *Journal of the California Dental Association*. 2013; 41: 575–581, 584.
- [14] Schwendicke F. Artificial intelligence for caries detection: value of data and information. *Journal of Dental Research*. 2022; 101: 1350–1356.
- [15] Leite AF, Gerven AV, Willems H, Beznik T, Lahoud P, Gaëta-Araujo H, *et al.* Artificial intelligence-driven novel tool for tooth detection and segmentation on panoramic radiographs. *Clinical Oral Investigations*. 2021; 25: 2257–2267.
- [16] Kuwada C, Arijji 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: 464–469.
- [17] Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, *et al.* Detecting caries lesions of different radiographic extension on bitewings using deep learning. *Journal of Dentistry*. 2020; 100: 103425.
- [18] Takeda S. Landmark annotation and mandibular lateral deviation analysis of posteroanterior cephalograms using a convolutional neural network. *Journal of Dental Sciences*. 2021; 16: 957–963.
- [19] Vishwanathaiah S. Artificial intelligence its uses and application in pediatric dentistry: a review. *Biomedicine*. 2023; 11: 788.
- [20] Lee JH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of Dentistry*. 2018; 77: 106–111.
- [21] Murata M, Arijji Y, Ohashi Y, Kawai T, Fukuda M, Funakoshi T, *et al.* Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. *Oral Radiology*. 2019; 35: 301–307.
- [22] Lee KS, Kwak HJ, Oh JM, Jha N, Kim YJ, Kim W, *et al.* Automated detection of tmj osteoarthritis based on artificial intelligence. *Journal of Dental Research*. 2020; 99: 1363–1367.
- [23] Askar H, Krois J, Rohrer C, Mertens S, Elhennawy K, Ottolenghi L, *et al.* Detecting white spot lesions on dental photography using deep learning: a pilot study. *Journal of Dentistry*. 2021; 107: 103615.
- [24] Alevizakos V. Artificial intelligence system for training diagnosis and differentiation with molar incisor hypomineralization (MIH) and similar pathologies. *Clinical Oral Investigations*. 2022; 26: 6917–6923.
- [25] Doufexi A. Tooth ankylosis: etiology, diagnosis and treatment review and case series. *European Journal of Dental and Oral Health*. 2025; 6: 12–15.

- [26] Hadi A. Ankylosed permanent teeth: incidence, etiology and guidelines for clinical management. *Medical Dental Research*. 2018; 1: 1–11.
- [27] Adedigba AP. Performance evaluation of deep learning models on mammogram classification using small dataset. *Bioengineering*. 2022; 9: 161.
- [28] Zhang J. Optimization and performance evaluation of deep learning algorithm in medical image processing. *Frontiers in Computing and Intelligent Systems*. 2024; 7: 67–71.
- [29] Dentsply Sirona. Orthophos XG 3 digital panoramic X-ray for practical diagnostics. 2025. Available at: <https://dspconnect.s3.amazonaws.com/Sirona/2Dand3DImaging/OrthophosXG3Brochure.pdf> (Accessed: 30 April 2025).
- [30] Hussain M, Taher T, Almourad MB, Mafarja M. Optimizing VGG16 deep learning model with enhanced hunger games search for logo classification. *Scientific Reports*. 2024; 14: 31759.
- [31] Szegedy C. Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV and 27–30 June 2016. Institute of Electrical and Electronics Engineers: Piscataway, NJ. 2016.
- [32] Sandler M. Mobilenetv2: inverted residuals and linear bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT and 16 December 2018. Institute of Electrical and Electronics Engineers: Piscataway, NJ. 2018.
- [33] Alla SSM. Brain tumor detection using transfer learning in deep learning. *Indian Journal of Science and Technology*. 2022; 15: 2093–2102.
- [34] Aggarwal CC. Classifier performance and evaluation. In Aggarwal CC (ed.) *Machine learning for text* (pp. 209–234). 1st edn. Springer: Cham. 2018.
- [35] Asci E, Kilic M, Celik O, Cantekin K, Bircan HB, Bayrakdar IS, *et al.* A deep learning approach to automatic tooth caries segmentation in panoramic radiographs of children in primary dentition, mixed dentition, and permanent dentition. *Children*. 2024; 11: 690.
- [36] Schwendicke F. Convolutional neural networks for dental image diagnostics: a scoping review. *Journal of Dentistry*. 2019; 91: 103226.
- [37] Sun XY, Yuan C, Wang XZ, Wang X, Feng XP, Tai BJ, *et al.* Report of the National investigation of resources for oral health in China. *Chinese Journal of Dental Research*. 2018; 21: 285–297.
- [38] Schwendicke F, Rossi JG, Göstemeyer G, Elhennawy K, Cantu AG, Gaudin R, *et al.* Cost-effectiveness of artificial intelligence for proximal caries detection. *Journal of Dental Research*. 2021; 100: 369–376.
- [39] 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. *Journal of Endodontics*. 2024; 50: 144–153.
- [40] Marsillac Mde W. Dental anomalies in panoramic radiographs of pediatric patients. *General Dentistry*. 2013; 61: 29–33.
- [41] Mooney GC, Morgan AG, Rodd HD, North S. Ectopic eruption of first permanent molars: presenting features and associations. *European Archives of Paediatric Dentistry*. 2007; 8: 153–157.
- [42] Noothout JMH, De Vos BD, Wolterink JM, Postma EM, Smeets PAM, Takx RAP, *et al.* Deep learning-based regression and classification for automatic landmark localization in medical images. *IEEE Transactions on Medical Imaging*. 2020; 39: 4011–4022.
- [43] Hartman H. Exploring the potential of artificial intelligence in paediatric dentistry: a systematic review on deep learning algorithms for dental anomaly detection. *International Journal of Paediatric Dentistry*. 2024; 34: 639–652.
- [44] Ahn Y. Automated mesiodens classification system using deep learning on panoramic radiographs of children. *Diagnostics*. 2021; 11: 1477.
- [45] Martens D. Performance of classification models from a user perspective. *Decision Support Systems*. 2011; 51: 782–793.
- [46] Kçük DB, Imak A, Özçelik STA, Çelebi A, Türkoğlu M, Sengur A, *et al.* Hybrid CNN-transformer model for accurate impacted tooth detection in panoramic radiographs. *Diagnostics*. 2025; 15: 244.
- [47] Kang S. Diagnostic accuracy of dental caries detection using ensemble techniques in deep learning with intraoral camera images. *PLOS ONE*. 2024; 19: e0310004.
- [48] Mahesh Batra A. A new era of dental care: harnessing artificial intelligence for better diagnosis and treatment. *Cureus*. 2023; 15: e49319.

How to cite this article: Nagehan Yılmaz, Mustafa Hakan Bozkurt, Tamer Tüzüner, Musa Aslan. Artificial intelligence for detecting dental ankylosis in primary molars using panoramic radiographs—a retrospective study. *Journal of Clinical Pediatric Dentistry*. 2025; 49(6): 120-130. doi: 10.22514/jocpd.2025.133.