



Development and validation of an artificial intelligence software for periodontal bone loss in panoramic imaging

Hakan Amasya^{1,2,3}  | Prashant Prakash Jaju⁴ | Matvey Ezhov⁵ | Maxim Gusarev⁵ | Cemal Atakan⁶ | Alex Sanders⁵ | David Manulius⁵ | Maria Golitskya⁵ | Kriti Shrivastava⁴ | Ajita Singh⁴ | Anuja Gupta⁴ | Merve Önder⁷ | Kaan Orhan^{7,8,9} 

¹Department of Oral and Dentomaxillofacial Radiology, Faculty of Dentistry, Istanbul University-Cerrahpaşa, Istanbul, Turkey

²CAST (Cerrahpaşa Research, Simulation and Design Laboratory), Istanbul University-Cerrahpaşa, Istanbul, Turkey

³Health Biotechnology Joint Research and Application Center of Excellence, Esenler, Istanbul, Turkey

⁴Department of Oral Medicine and Radiology, Rishiraj College of Dental Sciences and Research Centre, Bhopal, India

⁵Diagnocat, Inc, San Francisco, California, USA

⁶Department of Statistics, Faculty of Science, Ankara University, Ankara, Turkey

⁷Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Ankara University, Ankara, Turkey

⁸Ankara University Medical Design Application and Research Center (MEDITAM), Ankara, Turkey

⁹Department of Dental and Maxillofacial Radiodiagnostics, Medical University of Lublin, Lublin, Poland

Correspondence

Kaan Orhan, Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Ankara University, Ankara, Turkey.

Email: knorhan@dentistry.ankara.edu.tr; call53@yahoo.com

Abstract

This retrospective study is aimed at developing a web-based artificial intelligence (AI) software (DiagnoCat) for periodontal bone loss detection on panoramic radiographs and evaluating the model's performance by comparing it with clinicians' results. Separate models are trained for tooth and periodontal bone loss detection. The first model's objective was to detect teeth, segmenting their masks, and to define their numbering and developed with Mask R-CNN using pretrained ResNet-101 as a backbone. The second model was based on Cascade R-CNN architecture and used for bone loss prediction. Around 100 radiographs are evaluated by three clinicians regarding tooth identification and periodontal bone loss, separately. Ground truth is determined by the consensus and model's performance is evaluated with kappa, precision, recall, and *F*-score statistics. For tooth conditions, the overall *F*-score, accuracy, and Cohen's kappa coefficients were found to be 0.948, 0.977, and 0.933 for the binary, and 0.992, 0.988, and 0.961 for the multiclass results. For bone loss detection, the overall *F*-score, accuracy, and Cohen's kappa coefficients were found to be 0.985, 0.980, and 0.956 for the binary, and 0.996, 0.993, and 0.974 for the multiclass results. The results of this study suggest that the use of a web-based AI software (DiagnoCat) can be beneficial in detecting periodontal bone loss on panoramic radiographs.

KEYWORDS

alveolar bone, dentistry, periodontal diagnostics, periodontal disease, radiography

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *International Journal of Imaging Systems and Technology* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Periodontitis is a multifactorial chronic inflammatory disease characterized by the destruction of soft (gingiva and periodontal ligament) and hard (alveolar bone and cementum) periodontal tissues.^{1,2} The dental plaque is a predisposing factor for the disease to occur, and it initially develops as gingivitis.¹ Advanced cases are defined as the periodontitis and at this stage, bone loss occurs due to the increase in the destruction of the bone tissues supporting the tooth. Patients may be adversely affected by functional, esthetic, and psychological aspects due to various symptoms such as bleeding gums, increased tooth mobility, and tooth loss.³ The global prevalence increased by 50.2% from 1990 to 2007 and by 26.6% from 2007 to 2017.⁴ Despite all preventive measures and advancements in the diagnosis and treatment options, the disease is highly prevalent worldwide. Periodontitis is a multifactorial disease and the treatment becomes more challenging in advanced cases. Therefore, early diagnosis plays a critical role in treatment response.²

Diagnosis of periodontitis can be conducted through clinical and radiographic examinations. Although the diagnosis of the disease is mainly based on the findings of a complete clinical examination, this process is detailed and time-demanding.^{5,6} Radiographic monitoring of the bone level is useful in the diagnosis and determination of the severity of the disease. Several dental imaging techniques can be used for this purpose. In digital panoramic imaging, a two-dimensional image of all teeth and supporting tissues are acquired at once and are routinely used in dental radiographic examination. It is possible to monitor the bone level on the digital images obtained.^{6,7} Although panoramic images contribute to the diagnosis of periodontitis, they are insufficient for accurate diagnosis alone and it is recommended to be evaluated together with clinical parameters such as gingival index and mobility index.⁷

In the 1950s, the term “artificial intelligence” (AI) was introduced as the idea of developing machines capable of performing tasks normally performed by humans.⁸ AI is the part of computer science concerned with designing an intelligent computer system that exhibits characteristics we associate with intelligence in human behavior such as follows: understanding language, learning, reasoning, problem-solving, and many more.⁹ Machine learning (ML) is a sub-field of AI in which algorithms are trained through statistical analysis of the dataset, rather than being strictly coded. ML systems adapt to the data set provided in the training and an appropriate mathematical formula is developed between the input and output values as the training is completed.^{8–10} Basically, handcrafted features that define the sample can be used in the training of classification or regression algorithms. Classifier models are trained using

categorical labels, while regression models output numerical results.^{8,14} Clustering models are trained for cluster extraction of sample features to estimate pattern similarity.¹² Various ML techniques can be adopted to develop a system, and choosing the correct technique depends on proper analysis of the problem and the dataset.

Artificial neural networks are inspired by the biological neurons in the human brain. The basic network architecture consists of neurons (nodes) in discrete layers, called input, output, and intermediate layers. Each node is connected to nodes in the neighboring layer and this connection has a weight that can be calculated mathematically. Neural network models are trained by optimizing these weights in sequential trials to obtain the desired results.^{11,12} The term “deep learning” refers to deep (multi-layered) neural network architectures, with more than a few intermediate layers; the models are based on multi-layer mathematical operations for learning and inferring complex data like imagery. It is possible to develop systems with minimal human intervention using the deep learning approach, but the performance of such systems depends on the quality of the dataset provided for training; insufficient data will seriously affect the system outcomes.^{8,13} Convolutional neural networks (CNNs) are a deep learning technique that draws attention to its performance in image processing tasks, especially. The basic architecture of the CNNs is an extended neural network model with the convolution and the pooling layers, which provide automatic feature extraction.^{13,14}

AI techniques can be used in developing systems for the diagnosis of periodontitis. Several studies on the diagnosis of periodontal disease in panoramic images have been reported, and the CNN approach is commonly adopted in the developed models.^{7,15–19} This study aims to evaluate the performance of a web-based AI software (DiagnoCat) developed for tooth numbering and diagnosis of periodontal defects on digital panoramic radiographs, by comparing the results of the AI software with the hybrid results of human observers. The null hypothesis of this research is that the results of the AI system are comparable to the consensus determined by the voting of human observers for detecting periodontal bone loss in panoramic radiographs. It is suggested that the kappa, accuracy, and *F*-score values of 0.8 and above may demonstrate the similarity of the human consensus and model results.

2 | MATERIALS AND METHODS

2.1 | Dataset and preprocessing

This study was granted ethical permission by RCDS Ethical Committee (Letter.No.RCDSEC/21/28). For training AI models, 6000 anonymized panoramic radiographs

were used with 45 161 annotated periodontal bone loss instances. Periodontal diseases in annotation could be presented by segmented mask and by a bounding box (Figure 1).

Previously, a separate model was trained only on segmented annotations. If in the training data, only bounding boxes are given, these images are put through the model with prior knowledge of the bounding box coordinates to obtain pseudo segmentation for further training. The model used to predict segmentation is based on a two-stage detector with Cascade R-CNN architecture.²⁰

The inclusion criteria of the image samples were the absence of excessive image artifact, the presence of 10 or more teeth, and the absence of severe developmental anomalies. Samples that did not meet the criteria were not included in the study. There was no restriction in terms of device model and image parameters in the selection of image samples.

2.2 | ROI detector

The model used to detect regions of interest (ROI) was trained separately for tooth detection. The training dataset consisted of ~4500 images with annotated teeth,

including missing teeth. The models' objective was to detect teeth, segmenting their masks, and to define their numbering. For this task, 2-stage detector Mask R-CNN²¹ was used with pretrained ResNet-101²² as a backbone.

Predictions of the trained model for object detection were used to define the mouth area as the region of interest. The output of the model was bounding box and segmentation masks with predicted numbers for each tooth. Coordinates of the mouth area were taken as the minimum and maximum values of x and y coordinates for all detected teeth extended by a chosen number of pixels (Figure 2).

2.3 | Periodontal bone loss detector design

First, all panoramic images were enriched with pseudo segmentations from the previously trained model, then each image was cropped according to predictions from the ROI detector, and then these images were fed to the model.

The architecture chosen for the model was Cascade R-CNN,²⁰ where unlike Mask R-CNN model box object detection and segmentation is iteratively refined and

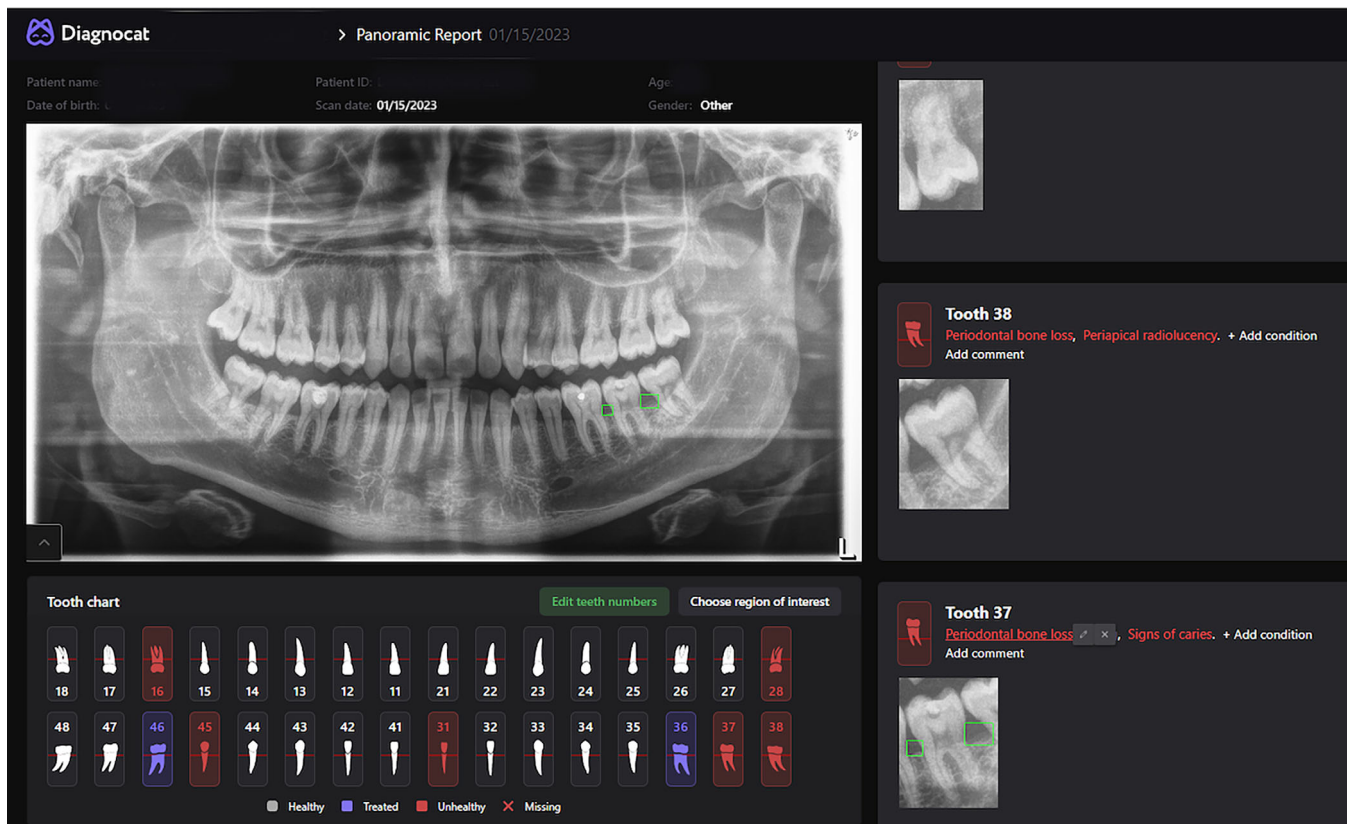


FIGURE 1 Examples of annotated data.

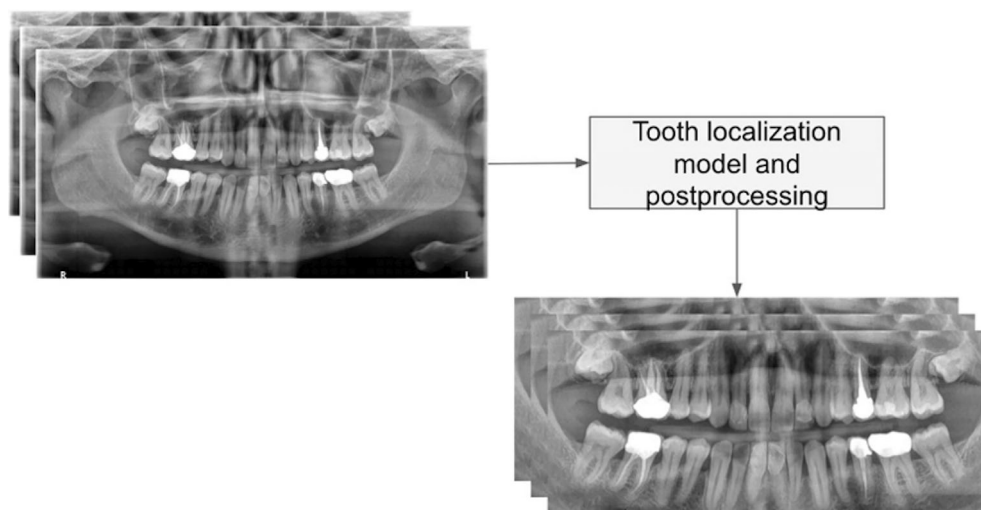


FIGURE 2 The image showing predictions of the trained model was used to define the mouth area as region of interest from panoramic radiograph.

image classification is taken as the average outcome for each cascade layer. This improves the quality of prediction and reduces overfitting.

To enhance the generalization of the model and improve performance, a variety of augmentations were applied to the input data: randomized crop, randomized rotation, randomized brightness, randomized contrast, randomized downscaling of images, randomized blur, randomized noise, optical distortion, grid distortion and contrast limited adaptive histogram equalization (CLAHE).²³

2.4 | Inference

While inference, a panoramic image first has to pass through the ROI detection like in the data preprocessing step, and then pass through the trained cascade model. Predictions of the model are then calibrated to optimize the F1-score and provided as the final output.

2.5 | Ground truth and human assessment

About 100 anonymized panoramic radiographs with permanent dentition were selected from the archive of Rishiraj College of Dental Sciences and Research Centre and analyzed by three clinicians separately. Status of each teeth region is coded as “sound”, “missing”, “restored”, “support” or “excluded”. Alveolar bone loss in approximal surfaces is scored based on the modified criteria of the 2017 World Workshop.²⁴ For staging the periodontitis severity, Stage 1 indicates <15% bone loss, Stage 2 indicates 15%–33% bone loss, and further bone loss indicates Stage 3 and 4. Originally, the distinction between the Stage 3 and 4 is proposed as the number of

tooth losses due to periodontitis; however, the data for the reason for tooth loss was not available, so the threshold between Stages 3 and 4 is determined as %80 bone loss. In addition, the bone loss pattern is saved as “vertical”, “horizontal” or “mixed” if possible, for the mesial and the distal surfaces. Observers' scores on tooth status and bone loss are recorded via an excel template. Results of each clinician are converted to binary results, as a bone loss at Stage 1 and below is “false”, and bone loss at Stage 2 and above is “true”. In determining the ground truth, three observers voted “true” or “false” for each variable. Thus, a definite consensus is determined in the binary results and finalized after being checked by an expert in dentomaxillofacial radiology.

2.6 | Statistical analysis

Agreement among three observers is demonstrated by Fleiss' Kappa coefficients. Agreement between the ground truth and the model's results is analyzed with Cohen's Kappa coefficients. The formula for the model's performance metrics (precision, recall, and *F*-score) are defined as follows²⁵:

$$\text{Precision} = \text{True Positive} / (\text{True Positive} + \text{False Positive})$$

$$\text{Recall} = \text{True Positive} / (\text{True Positive} + \text{False Negative})$$

$$F\text{-Score} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

Cohen's and Fleiss' Kappa coefficients are interpreted as “no agreement” for values <0.00; “slight agreement” for 0.01–0.2, “fair agreement” for 0.21–0.4, “moderate agreement” for 0.41–0.6, “substantial agreement” for

0.61–0.8 and “almost perfect agreement” for 0.81–1.²⁶ The statistical significance threshold is determined as $p < 0.05$.

3 | RESULTS

The Fleiss' kappa values among the three observers ranged from 0.79 to 1.00 for tooth status and 0.00 to 1.00 for periodontal assessment (Table S1). In the ground truth, for binary results, 76.5% of the teeth were found to be present (Table 1), while bone loss was determined on 64.44% of the total present approximal surfaces. The horizontal bone loss pattern was the most common bone loss pattern (Table 2).

In binary results, the overall F -score and Cohen's Kappa coefficients were 0.948 and 0.933, respectively, for the tooth conditions (Table 3), and 0.985 and 0.956 for the bone loss (Table 4), respectively. For tooth conditions, the highest F -score was achieved in the maxillary

molar region (0.964), while the lowest was found in the maxillary anterior region (0.919). For bone loss detection, the highest F -score was achieved in the mandibular premolar region (0.988), while the lowest was found in the mandibular molar region (0.979). The highest accuracy was found in the mandibular premolar and anterior (0.980) regions for tooth conditions. The highest accuracy was found in the mandibular premolar surfaces (0.985) for bone loss detection ($p < 0.05$).

In multi-class results, the overall F -score and Cohen's Kappa coefficients were found to be 0.992 and 0.961 for the tooth conditions (Table 5), and 0.996 and 0.974 for the bone loss (Table 6), respectively. For tooth conditions, the highest F -score was achieved in the maxillary molar region (0.999), while the lowest was found in the mandibular molar region (0.990). For bone loss detection, the highest F -score was achieved in maxillary molar (0.997) and anterior (0.997) regions, while the lowest was found in all mandibular regions (0.995). The highest accuracy was found in the mandibular anterior (0.994)

TABLE 1 Distribution of the data labels regarding tooth condition in ground truth.

Condition	Tooth (binary)	Present		Absent		
		Sound	Restorated	Missing	Support	Excluded
Positive	2448	2244	204	391	141	220
Negative	752	956	2996	2809	3059	2980
Total	3200					

TABLE 2 Distribution of the data labels regarding bone loss as the disease stage and the pattern in ground truth.

Bone loss	Surface (binary)	Stage				Bone loss pattern		
		1	2	3	4	Vertical	Horizontal	Mixed
Positive	3155	1159	406	65	57	9	1510	93
Negative	1741	3737	4490	4831	4839	4887	3386	4803
Total	4896							

TABLE 3 Performance metrics for identifying binary tooth conditions as absence or presence of teeth is demonstrated with F -score, accuracy, precision, recall and kappa coefficient values ($p < 0.05$).

Region		F -Score		Accuracy		Precision		Recall		Kappa		p value for χ^2	
Molar	Maxilla	0.962	0.964	0.973	0.975	0.995	1.000	0.930	0.930	0.941	0.945	0.000	0.000
	Mandible		0.959		0.972		0.990		0.931		0.938		0.000
Premolar	Maxilla	0.937	0.935	0.978	0.975	0.993	0.986	0.887	0.889	0.923	0.920	0.000	0.000
	Mandible		0.938		0.980		1.000		0.884		0.927		0.000
Anterior	Maxilla	0.921	0.919	0.979	0.978	1.000	1.000	0.854	0.851	0.909	0.907	0.000	0.000
	Mandible		0.923		0.980		1.000		0.857		0.912		0.000
Overall		0.948		0.977		0.996		0.904		0.933		0.000	

TABLE 4 Performance metrics for identifying binary surface conditions as absence or presence of bone loss is demonstrated with *F*-score, accuracy, precision, recall and kappa coefficient values ($p < 0.05$).

Region		<i>F</i> -Score		Accuracy		Precision		Recall		Kappa		<i>p</i> value for χ^2	
Molar	Maxilla	0.983	0.987	0.980	0.983	0.969	0.976	0.998	0.998	0.960	0.963	0.000	0.000
	Mandible		0.979		0.978		0.961		0.997		0.955		0.000
Premolar	Maxilla	0.985	0.983	0.980	0.975	0.971	0.967	1.000	1.000	0.954	0.933	0.000	0.000
	Mandible		0.988		0.985		0.975		1.000		0.968		0.000
Anterior	Maxilla	0.986	0.985	0.980	0.979	0.972	0.969	1.000	1.000	0.953	0.950	0.000	0.000
	Mandible		0.987		0.982		0.974		1.000		0.957		0.000
Overall		0.985		0.980		0.971		0.999		0.956		0.000	

TABLE 5 Performance metrics for multiclass tooth conditions as sound, restored, support, missing or excluded teeth is demonstrated with *F*-score, accuracy, precision, recall and kappa coefficient values ($p < 0.05$).

Region		<i>F</i> -Score		Accuracy		Precision		Recall		Kappa		<i>p</i> value for χ^2	
Molar	Maxilla	0.990	0.999	0.984	0.985	0.990	0.990	0.990	0.990	0.951	0.952	0.000	0.000
	Mandible		0.990		0.984		0.990		0.990		0.950		0.000
Premolar	Maxilla	0.993	0.993	0.989	0.989	0.993	0.993	0.993	0.993	0.966	0.966	0.000	0.000
	Mandible		0.993		0.989		0.993		0.993		0.966		0.000
Anterior	Maxilla	0.994	0.993	0.990	0.989	0.994	0.993	0.994	0.993	0.969	0.965	0.000	0.000
	Mandible		0.995		0.991		0.995		0.995		0.973		0.000
Overall		0.992		0.988		0.992		0.992		0.961		0.000	

TABLE 6 Performance metrics for multiclass tooth conditions as healthy, four distinct bone loss stages and three separate pattern is demonstrated with *F*-score, accuracy, precision, recall and kappa coefficient values ($p < 0.05$).

Region		<i>F</i> -Score		Accuracy		Precision		Recall		Kappa		<i>p</i> value for χ^2	
Molar	Maxilla	0.996	0.997	0.993	0.994	0.995	0.996	0.996	0.997	0.976	0.979	0.000	0.000
	Mandible		0.995		0.992		0.994		0.996		0.972		0.000
Premolar	Maxilla	0.995	0.996	0.992	0.993	0.995	0.996	0.996	0.996	0.971	0.972	0.000	0.000
	Mandible		0.995		0.992		0.995		0.995		0.970		0.000
Anterior	Maxilla	0.996	0.997	0.993	0.994	0.996	0.997	0.996	0.997	0.975	0.978	0.000	0.000
	Mandible		0.995		0.992		0.995		0.996		0.971		0.000
Overall		0.996		0.993		0.995		0.996		0.974		0.000	

region for tooth conditions, while it was found in the maxillary molar and anterior (0.994) regions for bone loss detection ($p < 0.05$).

For each tooth separately, Cohen's Kappa coefficients were changed between 0.66 and 1.00 for tooth conditions, and 0.00–1.00 for bone loss detection (Table S2). Supplementary tables can be found at: https://github.com/HakanAmasya/opg_pdl_ai.git

4 | DISCUSSION

Periodontal diseases are highly prevalent globally. In 2010, approximately 743 million people suffered from

severe periodontitis worldwide, making periodontitis the sixth most common disease.²⁷ Alongside the local destructive events in the oral cavity, researchers are interested in the relationship between periodontitis and other systemic conditions such as diabetes mellitus, atherosclerotic heart disease, metabolic syndrome, adverse pregnancy outcomes, and other chronic diseases. However, explaining the mechanisms behind these relationships remains a major challenge in this dynamic field.^{28,29}

Periodontal diseases are diagnosed through clinical and radiological examinations. Important findings such as bleeding on probing, tooth hypermobility, and gingival features are exclusive to clinical examination and are not captured in radiographs. Radiographic examinations can

reveal periodontal findings such as bone levels, bone loss, root morphology, endodontic lesions, furcation radiolucency, and intra-bony defects using various dental radiography techniques.³⁰ However, radiographic examinations usually underestimate the actual bone loss, and gingivitis signs cannot be detected since hard tissues are not involved yet [15]. Additionally, a single radiographic examination cannot determine whether bone loss is due to an existing disease or previous disease. The choice of radiographic technique depends on the advantages and disadvantages of each method, clinical findings, and patient's benefits.

Dental intraoral imaging primarily offers periapical and bitewing radiography. Periapical radiographs project teeth and periodontal structures with relatively lower radiation doses compared to extraoral devices. However, the imaging area is limited to a few teeth, and several radiographs are required for a full-mouth examination. The long-cone paralleling technique yields more accurate bone level measurements compared to the bisecting angle technique, but it requires specialized equipment. Bitewing radiographs show the condition of approximal tooth surfaces and supporting bone levels in both jaws without involving the apical region. Although horizontal bitewing radiographs are suitable for most cases, deeper bone defects require vertical alignment of the film, which can be uncomfortable.^{30,31}

Dental extraoral imaging mainly offers panoramic and cone-beam computed tomography (CBCT). Panoramic radiography demonstrates whole teeth, dentoalveolar structures, and both jaws in a single image with lower radiation compared to intraoral full-mouth series, and offers quick and straightforward image capture without the need for intraoral manipulation. Panoramic radiography is valuable for oral and maxillofacial imaging; however, drawbacks hinder its diagnostic role in periodontology, particularly due to image magnification and distortion related to projecting a large volume into a single planar image.^{30,31} Extraoral bitewing is a novel radiographic method supported by specialized panoramic imaging devices. Although the use of terms "extraoral" and "bitewing" together creates a contradiction, this technique optimizes panoramic images for approximal surfaces, essentially.³² CBCT is a specialized imaging modality that provides detailed three-dimensional images of anatomical structures, particularly in the head and neck region. It is used to assess bone quality, identify periodontal defects, and plan dental implant placement with higher precision. CBCT offers volumetric data at lower patient doses than medical CT devices, but its radiation dose is higher compared to intraoral techniques and panoramic imaging. Justification for its use should follow the "As Low as Reasonably Achievable" principles.^{30,31}

In 2014, Takeshita et al. evaluated the diagnostic accuracy of conventional and digital periapical radiography, panoramic radiography, and CBCT in assessing proximal alveolar bone loss. All imaging methods provided satisfactory accuracy in proximal bone level measurements except conventional periapical radiographs with a Han-Shin film holder. Although CBCT provided measurements closest to the control values, panoramic images are also close and recommended for initial evaluation.³³ Kumar et al. reviewed different radiographic modalities for detecting osseous defects. Intra-oral radiography was reported to be superior to panoramic radiographs, and CBCT was considered a useful complement. Justification for using CBCT over conventional techniques should consider radiation dose and clinical benefits.³⁴ Zhang et al. compared CBCT and periapical/bitewing radiography measurements with clinical attachment loss. Significant positive correlations were reported in all techniques, with CBCT showing a higher Pearson correlation coefficient with clinical measurements compared to intraoral techniques, although the difference was not significant.³⁵ Manja and Fransiari compared bitewing, periapical, and panoramic radiographs to measure alveolar bone loss. Bitewing radiographs were found to be superior in accuracy compared to periapical and panoramic images, with no significant difference among the groups.³⁶

Berghuis et al. compared panoramic and periapical radiography for detecting furcation involvement. CBCT was used to obtain reference diagnoses for all furcation sites. Panoramic and periapical radiographs were found to be relevant tools with high specificity in diagnosis furcation involvement. A combination of radiological findings and furcation probing was suggested as the ideal technique for inspecting furcation sites.³⁷ Komšić et al. compared periodontal probing, intra-surgical measurements, panoramic and CBCT-based parameters in molar furcation assessment. The correlation between intra-surgical measurements and CBCT was superior compared to clinical probing and panoramic. Different clinical and radiological modalities were found to be correlated with each other and provided satisfactory accuracy.³⁸ The American Academy of Periodontology reported the Best Evidence Consensus to address oral indications for using CBCT in 2017. The use of CBCT was suggested to be beneficial in specific scenarios in the management of periodontitis, but routine use lacked sufficient evidence.^{39,40}

Current evidence supports the use of 2D full-mouth radiographic series in addition to clinical probing parameters as the gold standards for comprehensively evaluating periodontal structures. However, the limited imaging area in a single periapical image requires multiple

images, leading to higher radiation doses than a single panoramic image. Additionally, proper registration of the entire mouth using intraoral radiography is time-consuming. CBCT imaging's volumetric examination advantage is valuable, but device costs and the relatively high radiation dose limit its justification. Interpreting CBCT images also demands more expertise than two-dimensional radiographs.⁴⁰ Panoramic imaging might not be the first choice for periodontal evaluation, but it provides information about the periodontal status of the patient and is widely used for various dental purposes.

Within its limitations, a panoramic radiograph taken for any reason provides useful information about periodontal bone levels, especially for an initial examination. For this reason, panoramic radiography, which has become almost a routine method of imaging for dental purposes, is preferred in our study. Within the scope of this study, the absence or presence of periodontal problems are classified in the mesial and the distal sides of each tooth, and measurement of bone loss quantitatively or furcation regions are excluded.

Dentomaxillofacial radiology has undergone a digital transformation with the development of digital sensor technologies. In 1997, Borg et al. compared digital periapical radiographs acquired with the charge-coupled device (CCD) and the photostimulable storage phosphor (PSP) plates in bone loss measurements. In their study, all images were acquired using a single x-ray device, and it was reported that the images obtained with the digital sensors were comparable to film-based radiographs in demonstrating bone loss. While there was no statistically significant difference between the two digital systems, the issue of underestimation of bone loss is found to be more minimal in digital systems.⁴¹ Vandenberghe et al. investigated the effect of various X-ray tube generators in the assessment of periodontal bone lesions using conventional and digital imaging receptors in 2011. The compressive research was published in two parts.^{42,43} The first part is focused on the effect of different x-ray tubes, while indirect and direct digital systems are compared with analog films secondarily. According to the results of the first part, regarding the sensors, digital systems are found to be advantageous over film-based imaging, especially in low-dose acquisitions.⁴² In the second part, several CCD and PSP systems are compared in terms of specific properties such as contrast resolution and measurement accuracy, along with the subjective quality evaluation. According to the results of the second study, there was a dose gain of around 50% by increasing the 8-bit phosphor plate system to 12-bit, while it was reported that there was no extra gain since the CCD systems in the study were at least 12 bits. The importance of bit rate for contrast resolution, which is important in

periodontal findings, is emphasized.⁴³ Also, the use of a dedicated periodontal filter is reported to increase the measurement accuracy.^{43,44} The ease of quantitative analysis of digital data paves the way for the development of clinical decision support systems using AI techniques, as well as image processing applications. Clinical decision support systems are applications that provide a second expert opinion in the decisions of clinicians.⁴⁵

Decision support systems for different purposes can be developed with various ML techniques.^{46–48} CNN models developed with deep learning have the potential for superior success in image processing.¹⁴ In 2022, Chang et al. developed a multitasking InceptionV3 model for the radiographic diagnosis of periodontitis. The data set consisted of 1836 periapical images of patients with full mouth standardized radiographs, and samples were evaluated for radiographic bone loss (RBL) and defect morphology. The last layer of the InceptionV3 model was replaced by a Global Max Pooling layer, a 1024-node fully connected layer with the rectified linear (ReLU) activation function, and the outputs were linked to two parallel fully connected layers. The RBL classification was performed by two-node, and the defect morphology classification was performed by three-node fully connected layers, both with softmax activation functions. Mean accuracies of the model were reported to be 0.88 ± 0.03 for the mild and 0.86 ± 0.03 for the severe bone loss group, with no significant difference in accuracy between the two groups ($p = 0.20$).⁴⁹ Alotaibi et al. developed a CNN model based on VGG-16 (Visual Geometry Group) network architecture with the TensorFlow and Keras libraries in Python. The data set consisted of 1724 intraoral periapical images of upper and lower anterior teeth. The model was based on 13 convolutional layers and two dense layers, and trained using 100 epochs and 16 batch sizes, and the outputs were binary (normal, abnormal) or multi-class (normal, mild, moderate, severe) categories. The total diagnostic accuracies for the alveolar bone levels were reported as 73.04% in the binary classification, and 59.42% in multi-class classification.⁵⁰ In 2023, Chen et al. developed an ensemble model utilizing the YOLOv5 and VIA labeling platform, including VGG-16 and U-Net architecture, to detect tooth position, tooth shape, periodontal bone level detection, and RBL in periapical and bitewing radiographs. The accuracy of RBL detection was reported to be 97.0%, while the overall accuracy of the model was reported to be approximately 90%.⁵¹

Ezhov et al. developed a web-based AI system for dental diagnosis using CBCT, which is the same platform as in this study (DiagnoCat), and evaluated its performance for clinical applicability with the participation of 24 dentists. The model architecture consisted of five

modules. The first two modules were ROI localization and tooth localization and numeration modules, both based on the volumetric modification of U-Net architecture. The remaining modules were described as periodontitis module, caries localization module, and periapical lesion localization modules. 1346 CBCT scans were used to train and 30 CBCT scans were examined by two groups of dentists, where one group was aided by Diagnocat and the other was unaided. The sensitivity and specificity of the developed system for periodontal bone loss were reported to be 0.9489 and 0.9661, respectively. The sensitivity values were reported to be increased in aided evaluations for periodontal bone loss conditions, with a slight reduction in specificity when compared to unaided evaluations.⁵²

Various models have been developed in research on periodontal findings in panoramic images. Shon et al. integrated U-Net and YOLOv5 together to develop a deep-learning model for classifying periodontitis stages on panoramic radiographs in 2022. In their study, YOLOv5 is used for tooth detection and numbering, while the U-Net is used to detect the boundaries to decide the stage of the periodontal disease. The proposed model had an accuracy of 0.929, with a recall and precision of 0.807 and 0.724, respectively, on average across all four stages.⁵³ In 2021, Vigil and Bharathi developed a model that classifies panoramic images as periodontally healthy or not. Histogram equalization and median filter are used in preprocessing to remove noise, and sharpen applied. Teeth and bony shapes and structures are obtained using erosion morphology operation, and located by two-dimensional Otsu (2D Otsu) threshold segmentation. Performance metrics of the model are reported as an accuracy of about 91.34%, the sensitivity of 92.8%, and *F*-score of 95.47%.¹⁵ Researchers also developed another model to diagnose periodontitis stages of mandibular area using Adaptive Center Line-Distance Based approach. The proposed model achieved an accuracy of 92.83%, precision of 97.59%, recall/sensitivity of 95.01% and *F*-score/*F*-measure of 96.28%.¹⁶ In 2020, Thanathornwong and Suebnukarn proposed a deep-learning model to identify periodontally compromised teeth on panoramic radiographs. The faster regional CNN (faster R-CNN) model is developed using a pretrained ResNet architecture and a small annotated data set. The average precision and recall rates are reported to be 0.81 and 0.80. Sensitivity, specificity, and *F*-measure are found to be 0.84, 0.88, and 0.81, respectively.¹⁷ Chang et al. proposed a hybrid of deep learning architecture for the classification of periodontitis stages on dental panoramic radiographs. The developed system analyzed the radiographic bone level in whole jaw and detected to calculate the percentage rate of the RBL. The intraclass correlation

coefficients for the classifier model and the professor, fellow, and resident's diagnoses are reported to be 0.86, 0.84, and 0.82, respectively.¹⁸ Bayrakdar et al. developed a CNN system to detect alveolar bone loss from dental panoramic radiographs using a pretrained Google Net Inception v3. The model is trained using transfer learning with the Tensor Flow with a total of 2276 images. The sensitivity, specificity, precision, accuracy, and *F1* score for the model are reported to be 0.94, 0.89, 0.89, 0.91, and 0.92, respectively.¹⁹ In our study, the performance of the AI software is evaluated with binary and multiclass scores. For tooth conditions, the overall *F*-score, accuracy, and Cohen's kappa coefficients were found to be 0.948, 0.977, and 0.933 for the binary, and 0.992, 0.988, and 0.961 for the multiclass results. For bone loss detection, the overall *F*-score, accuracy, and Cohen's kappa coefficients were 0.985, 0.980, and 0.956, respectively, for the binary results, and 0.996, 0.993, and 0.974 for the multiclass results.

Panoramic imaging is not the ideal tool for periodontal evaluation, due to the inherent limitations of the technique. In cases where the primary purpose of imaging is periodontal causes, the imaging technique preferred in our study constitutes a limitation. However, panoramic imaging is a technique that is used to examine not only for periodontal reasons but also many other conditions, and it is a technique where the entire jaw and teeth can be imaged with a relatively low radiation dose. Once the image is acquired, it contains information regarding the periodontal status of the patient. Clinical decision support systems can provide expert opinion on the primary justification criteria, as well as be useful in demonstrating other findings that are not in the main target but may be overlooked. For this reason, we suggest that a software, which provides expert opinion in evaluating periodontal findings on panoramic images can be useful as a clinical decision support system. Also, image samples in this study were not obtained using a single imaging device and fixed imaging parameters. This situation may be considered as another limitation. On the other hand, the AI system developed in our study is web-based and open to online access (<https://diagnocat.com>). This scenario requires the acceptance of radiographs produced in various configurations in different clinics. Such flexibility in model development contributes to potential users' greater benefits in utilizing the system, independent of the devices and parameters from which they generate data in the clinic. Issues such as the reliability of the radiographic projection of clinical bone loss or the effect of different devices are beyond the scope of our study.

Panoramic imaging is not the ideal tool for periodontal evaluation, due to the inherent limitations of the technique. In cases where the primary purpose of imaging is

periodontal causes, the imaging technique preferred in our study constitutes a limitation. However, panoramic imaging is a technique that is used to examine not only for periodontal reasons but also many other conditions, and it is a technique where the entire jaw and teeth can be imaged with a relatively low radiation dose. Once the image is acquired, it contains information regarding the periodontal status of the patient. Clinical decision support systems can provide expert opinion on the primary justification criteria, as well as be useful in demonstrating other findings that are not in the main target but may be overlooked. For this reason, we suggest that a software, which provides expert opinion in evaluating periodontal findings on panoramic images can be useful as a clinical decision support system. A software providing expert opinion on periodontal findings in panoramic images can serve as a clinical decision support system. This study's images were not acquired using a single imaging device and fixed parameters, which might be considered a limitation. However, the flexibility of the AI system to handle images from various configurations contributes to its usability. We suggest that data from assorted sources can be useful in reducing device-dependent factors, in a web-accessed system, to evaluate the periodontal bone loss using panoramic images.

5 | CONCLUSION

The evaluated AI software (DiagnoCat) achieved significant success compared to ground truth determined by three clinicians. Clinical decision support systems can aid in evaluating alternative findings beyond the primary imaging reason. Using such software can improve service quality.

CONFLICT OF INTEREST STATEMENT

Matvey Ezhov, Maxim Gusarev, Maria Golitsyna, David Manulius, and Alex Sanders are employees of Diagnostics Co. Ltd. Kaan Orhan is a scientific research advisor for Diagnostics Co. Ltd., San Francisco CA. Hakan Amasya, Prashant Prakash Jaju, Cemal Atakan, Kriti Shrivastava, Ajita Singh, Anuja Gupta, and Merve Önder have no potential competing interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Hakan Amasya  <https://orcid.org/0000-0001-7400-9938>
Kaan Orhan  <https://orcid.org/0000-0001-6768-0176>

REFERENCES

- Han P, Bartold PM, Ivanovski S. The emerging role of small extracellular vesicles in saliva and gingival crevicular fluid as diagnostics for periodontitis. *J Periodontol Res.* 2022;57(1):219-231.
- Baima G, Iaderosa G, Corana M, et al. Macro and trace elements signature of periodontitis in saliva: A systematic review with quality assessment of ionomics studies. *J Periodontol Res.* 2022;57(1):30-40.
- Stöhr J, Barbaresco J, Neuenschwander M, Schlesinger S. Bidirectional association between periodontal disease and diabetes mellitus: A systematic review and meta-analysis of cohort studies. *Sci Rep.* 2021;11(1):1-9.
- James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet.* 2018;392(10159):1789-1858.
- Farhadian M, Shokouhi P, Torkezaban P. A decision support system based on support vector machine for diagnosis of periodontal disease. *BMC Res Notes.* 2020;13(1):1-6.
- Machado V, Proença L, Morgado M, Mendes JJ, Botelho J. Accuracy of panoramic radiograph for diagnosing periodontitis comparing to clinical examination. *J Clin Med.* 2020;9(7):2313.
- Krois J, Ekert T, Meinhold L, et al. Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep.* 2019;9(1):1-6.
- Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res.* 2021;99(7):769-774.
- Borges AFS, Laurindo FJB, Spínola MM, Gonçalves RF, Mattos CA. The strategic use of artificial intelligence in the digital era: systematic literature review and future research directions. *IJIM.* 2021;57:102225.
- Meghil MM, Rajpurohit P, Awad ME, McKee J, Shahoumi LA, Ghaly M. Artificial intelligence in dentistry. *Artificial Intelligence Dentistry Dent Rev.* 2022;2(1):100009.
- Johari M, Esmaeili F, Andalib A, Garjani S, Saberhari H. Detection of vertical root fractures in intact and endodontically treated premolar teeth by designing a probabilistic neural network: an ex vivo study. *Dentomaxillofac Radiol.* 2017;46(2):20160107.
- Yadav AK, Chandell SS. Solar radiation prediction using artificial neural network techniques: A review. *Renew Sustain Energy Rev.* 2014;33:772-781.
- Mao K, Chen L, Wang M, Xu R, Zhao X. Classification of hand-wrist maturity level based on similarity matching. *IET Image Process.* 2021;15(12):2866-2879.
- Rawat W, Zenghui W. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* 2017;29(9):2352-2449.
- Vigil MSA, Bharathi VS. Detection of periodontal bone loss in mandibular area from dental panoramic radiograph using image processing techniques. *Concurr Comput Pract Exp.* 2021;33(17):e6323.
- Vigil MSA, Bharathi VS. Classification of periodontitis stages in mandibular area from dental panoramic radiograph using adaptive center line-distance based image processing approach. *JAIHC.* 2022;14:1-11.
- Thanathornwong B, Suebnukarn S. Automatic detection of periodontal compromised teeth in digital panoramic radiographs

- using faster regional convolutional neural networks. *Imaging Sci Dent.* 2020;50(2):169-174.
18. Chang HJ, Lee SJ, Yong TH. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Sci Rep.* 2020;10(1):1-8.
 19. Kurt S, Çelik Ö, Bayrakdar İŞ, et al. Success of artificial intelligence system in determining alveolar bone loss from dental panoramic radiography images. *Cumhuriyet Dent J.* 2020;23(4):318-324.
 20. Cai Z, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018:1-9.
 21. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*; 2017:1-12.
 22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2016:1-9.
 23. Zuiderveld K. Contrast limited adaptive histogram equalization. In: Heckbert PS, ed. *Graphics Gems IV*. Academic Press; 1994:474-485.
 24. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: framework and proposal of a new classification and case definition. *J Periodontol.* 2018;89(1):159-172.
 25. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F- score, with implication for evaluation. In: Losada DE, Fernández-Luna JM, eds. *Advances in Information Retrieval*. Springer; 2005:345-359.
 26. Richard LJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
 27. Peres MA, Macpherson LMD, Weyant RJ, et al. Oral diseases: a global public health challenge. *Lancet.* 2019;394(10194):249-260.
 28. Genco RJ, Mariano S. Clinical and public health implications of periodontal and systemic diseases: an overview. *Periodontol 2000.* 2020;83(1):7-13.
 29. Winning L, Linden GJ. Periodontitis and systemic disease: association or causality? *Curr Oral Health Rep.* 2017;4(1):1-7.
 30. Corbet EF, Ho DKL, Lai SML. Radiographs in periodontal disease diagnosis and management. *Aust Dent J.* 2009;54(1):27-43.
 31. Mol A. Imaging methods in periodontology. *Periodontol 2000.* 2004;34(1):34-48.
 32. Johnson KB, Mol A, Tyndall DA. Extraoral bite-wing radiographs: A universally accepted paradox. *J Am Dent Assoc.* 2021;152(6):444-447.
 33. Takeshita WM, Iwaki LCV, Da Silva MC, Tonin RH. Evaluation of diagnostic accuracy of conventional and digital periapical radiography, panoramic radiography, and cone-beam computed tomography in the assessment of alveolar bone loss. *Contemp Clin Dent.* 2014;5(3):318-323.
 34. Kumar V, Arora K, Udupa H. Different radiographic modalities used for detection of common periodontal and periapical lesions encountered in routine dental practice. *Oral Hyg Health.* 2014;2:163.
 35. Zhang W, Rajani S, Wang BY. Comparison of periodontal evaluation by cone- beam computed tomography, and clinical and intraoral radiographic examinations. *Oral Radiol.* 2018;34(3):208-218.
 36. Manja CD, Fransiari ME. A comparative assessment of alveolar bone loss using bitewing, periapical, and panoramic radiography. *Bali Med J.* 2018;7(3):636-638.
 37. Berghuis G, Cosyn J, Bruyn HD, Hommez G, Dierens M, Christiaens V. A controlled study on the diagnostic accuracy of panoramic and peri-apical radiography for detecting furcation involvement. *BMC Oral Health.* 2021;21(1):1-10.
 38. Komšić S, Plančak D, Kašaj A, Puhar I. A comparison of clinical and radiological parameters in the evaluation of molar furcation involvement in periodontitis. *Acta Stomatol Croat.* 2019;53(4):326-336.
 39. Kim DM, Bassir SH. When is cone-beam computed tomography imaging appropriate for diagnostic inquiry in the management of inflammatory periodontitis? An American Academy of periodontology best evidence review. *J Periodontol.* 2017;88(10):978-998.
 40. Mandelaris G, Scheyer ET, Evans M, et al. American Academy of periodontology best evidence consensus statement on selected oral applications for cone-beam computed tomography. *J Periodontol.* 2017;88(10):939-945.
 41. Borg E, Gröndahl K, Gröndahl HG. Marginal bone level buccal to mandibular molars in digital radiographs from charge-coupled device and storage phosphor systems. an in vitro study. *J Clin Periodontol.* 1997;24(5):306-312.
 42. Vandenberghe B, Livia C, Bosmans H, Yang J, Jacobs R. A comprehensive in vitro study of image accuracy and quality for periodontal diagnosis. Part 1: the influence of X-ray generator on periodontal measurements using conventional and digital receptors. *Clin Oral Investig.* 2011;15(4):537-549.
 43. Vandenberghe B, Bosmans H, Yang J, Jacobs R. A comprehensive in vitro study of image accuracy and quality for periodontal diagnosis. Part 2: the influence of intra-oral image receptor on periodontal measurements. *Clin Oral Investig.* 2011;15(4):551-562.
 44. Baksi BG. Measurement accuracy and perceived quality of imaging systems for the evaluation of periodontal structures. *Odontology.* 2008;96(1):55-60.
 45. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med.* 2020;3(1):1-10.
 46. Amasya H, Yildirim D, Aydogan T, Kemaloglu N, Orhan K. Cervical vertebral maturation assessment on lateral cephalometric radiographs using artificial intelligence: comparison of machine learning classifier models. *Dentomaxillofac Radiol.* 2020;49(5):20190441.
 47. Sin Ç, Akkaya N, Aksoy S, Orhan K, Öz U. A deep learning algorithm proposal to automatic pharyngeal airway detection and segmentation on CBCT images. *Orthod Craniofac Res.* 2021;24(2):117-123.
 48. Bayrakdar I, Orhan K, Akarsu S, et al. Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. *Oral Radiol.* 2022;68(4):468-479.
 49. Chang J, Chang MF, Angelov N, et al. Application of deep machine learning for the radiographic diagnosis of periodontitis. *Clin Oral Investig.* 2022;26(11):6629-6637.
 50. Alotaibi G, Awawdeh M, Farook FF, Aljohani M, Aldhafiri RM, Aldhoayan M. Artificial intelligence (AI) diagnostic tools: utilizing

- a convolutional neural network (CNN) to assess periodontal bone level radiographically—a retrospective study. *BMC Oral Health*. 2022;22(1):399.
51. Chen CC, Wu YF, Aung LW, et al. Automatic recognition of teeth and periodontal bone loss measurement in digital radiographs using deep-learning artificial intelligence. *J Dent Sci*. 2023;18(3):1301-1309.
 52. Ezhov M, Gusarev M, Golitsyna M, et al. Clinically applicable artificial intelligence system for dental diagnosis with CBCT. *Sci Rep*. 2021;11(1):15006.
 53. Shon HS, Kong V, Park JS, et al. Deep learning model for classifying periodontitis stages on dental panoramic radiography. *Appl Sci*. 2022;12(17):8500.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Amasya H, Jaju PP, Ezhov M, et al. Development and validation of an artificial intelligence software for periodontal bone loss in panoramic imaging. *Int J Imaging Syst Technol*. 2023;1-12. doi:[10.1002/ima.22973](https://doi.org/10.1002/ima.22973)