Journal of Dentistry 147 (2024) 105105

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent

Validation of artificial intelligence application for dental caries diagnosis on intraoral bitewing and periapical radiographs

Viktor Szabó [a,1], Bence Tamás Szabó [a,*,1], Kaan Orhan [a,b,c], Dániel Sándor Veres [d], David Manulis [e], Matvey Ezhov [e], Alex Sanders [e]

[a] Department of Oral Diagnostics, Faculty of Dentistry, Semmelweis University, Budapest, Hungary
[b] Department of Dentomaxillofacial Radiology, Faculty of Dentistry, Ankara University, Ankara, Turkey
[c] Medical Design Application, and Research Center (MEDITAM), Ankara University, Ankara, Turkey
[d] Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary
[e] Diagnocat Inc, San Francisco, CA, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Deep learning
Machine learning
Dental digital radiography
Dental caries
Diagnostic imaging

ABSTRACT

Objectives: This study aimed to assess the reliability of AI-based system that assists the healthcare processes in the diagnosis of caries on intraoral radiographs.

Methods: The proximal surfaces of the 323 selected teeth on the intraoral radiographs were evaluated by two independent observers using an AI-based (Diagnocat) system. The presence or absence of carious lesions was recorded during Phase 1. After 4 months, the AI-aided human observers evaluated the same radiographs (Phase 2), and the advanced convolutional neural network (CNN) reassessed the radiographic data (Phase 3). Subsequently, data reflecting human disagreements were excluded (Phase 4). For each phase, the Cohen and Fleiss kappa values, as well as the sensitivity, specificity, positive and negative predictive values, and diagnostic accuracy of Diagnocat, were calculated.

Results: During the four phases, the range of Cohen kappa values between the human observers and Diagnocat were κ=0.66–1, κ=0.58–0.7, and κ=0.49–0.7. The Fleiss kappa values were κ=0.57–0.8. The sensitivity, specificity and diagnostic accuracy values ranged between 0.51–0.76, 0.88–0.97 and 0.76–0.86, respectively.

Conclusions: The Diagnocat CNN supports the evaluation of intraoral radiographs for caries diagnosis, as determined by consensus between human and AI system observers.

Clinical significance: Our study may aid in the understanding of deep learning-based systems developed for dental imaging modalities for dentists and contribute to expanding the body of results in the field of AI-supported dental radiology..

1. Introduction

Dental caries remains one of the most common chronic diseases worldwide [1,2]. In 2019 Wen et al. [3] estimated that there were 2.0 billion untreated cases of caries in permanent teeth. In recent decades, epidemiological studies have shown a reduction in the prevalence of caries within the populations of developed countries; however, this value is still high, especially in underdeveloped areas [4]. Therefore, the early detection of the lesions and cost-effective treatment decisions are crucial [5].

The development of dental caries begins on the tooth surface in contact with the oral cavity and progresses from the outer layer to the deeper tissues. Carious lesions in the proximal areas often develop and grow unnoticed [6]. In several cases, visually examining teeth in the posterior region can be challenging, making radiographs essential for additional diagnostic support [6,7]. Conventional radiography is the most commonly used imaging modality in dental radiology [8,9]. Intraoral radiographs [10] such as bitewings, are widely used to diagnose carious lesions [11]. They offer insights into the dental conditions spanning from the distal surface of the canine crown to that of the last erupted crown [12]. However, it is worth noting that periapical radiographs can also effectively detect caries [10,13]. Conventional

* Corresponding author.
E-mail address: szabo.bence.tamas@semmelweis.hu (B.T. Szabó).
[1] Viktor Szabó & Bence Tamás Szabó are the first co-authors.

https://doi.org/10.1016/j.jdent.2024.105105
Received 10 October 2023; Received in revised form 21 May 2024; Accepted 28 May 2024
Available online 29 May 2024

intraoral radiography is usually the most suitable radiographic modality for the recognition of proximal caries [11,13].

Anatomical variations among teeth can hinder the detection of carious lesions on radiographs, and this challenge may persist even during chairside examinations [2]. Another limitation of carious lesion assessment on intraoral radiographs is the extension of the caries; the bitewing modality has a lower sensitivity value in the case of detecting enamel carious lesions at the initial stage [2,14]. Hence, establishing a diagnosis of caries should be based on clinical decisions after the results of visual and radiological examinations have been collected [15].

Neural networks, including convolutional neural networks (CNNs), are widely used in radiological image analysis. This may be because of the large amount of digital data that are already available that serve as the basis for artificial intelligence (AI) [16]. Additionally, owing to the diversity and complexity of structures depicted on radiographs, CNNs have proven to be a fast-developing and promising choice in this field. CNNs use a combination of different layers to perform specific subtasks, and the final output is the sum of the results [17]. These networks can be used for classification, detection, and segmentation in image analysis. Classification is a wide-ranging task in which the system determines whether the searched structure, such as pathology in a given image area, is present [18]. Detection is a similar algorithm that locates and identifies specific areas of an image in which the desired lesion is located [19]. During segmentation, the neural network identifies and labels the particular part of an image in which a potential pathological lesion is located [20].

In 2015 Ronneberger et al. introduced the first U-Net as a deep learning method for biomedical image segmentation [21]. This innovation significantly reduced the training of neural networks, potentially leading to the widespread adoption of the U-Net architecture as a state-of-the art model for image detection and segmentation [22]. Notably, similar neural networks can be applied and have been used in numerous diagnostic and therapeutic tasks in dental radiology, including caries detection [15,23], staging of periodontal diseases [24], detection of periapical lesions [25], dental implant planning [26], and temporomandibular joint disorders [27].

These AI-based systems may contribute to assessments by radiologists as a decision-support tool, which may facilitate patient care. Orhan et al. [28] reported that the use of an AI tool resulted in an average reduction of 1.19 min (6.78 %) in the assessment time for a cone beam

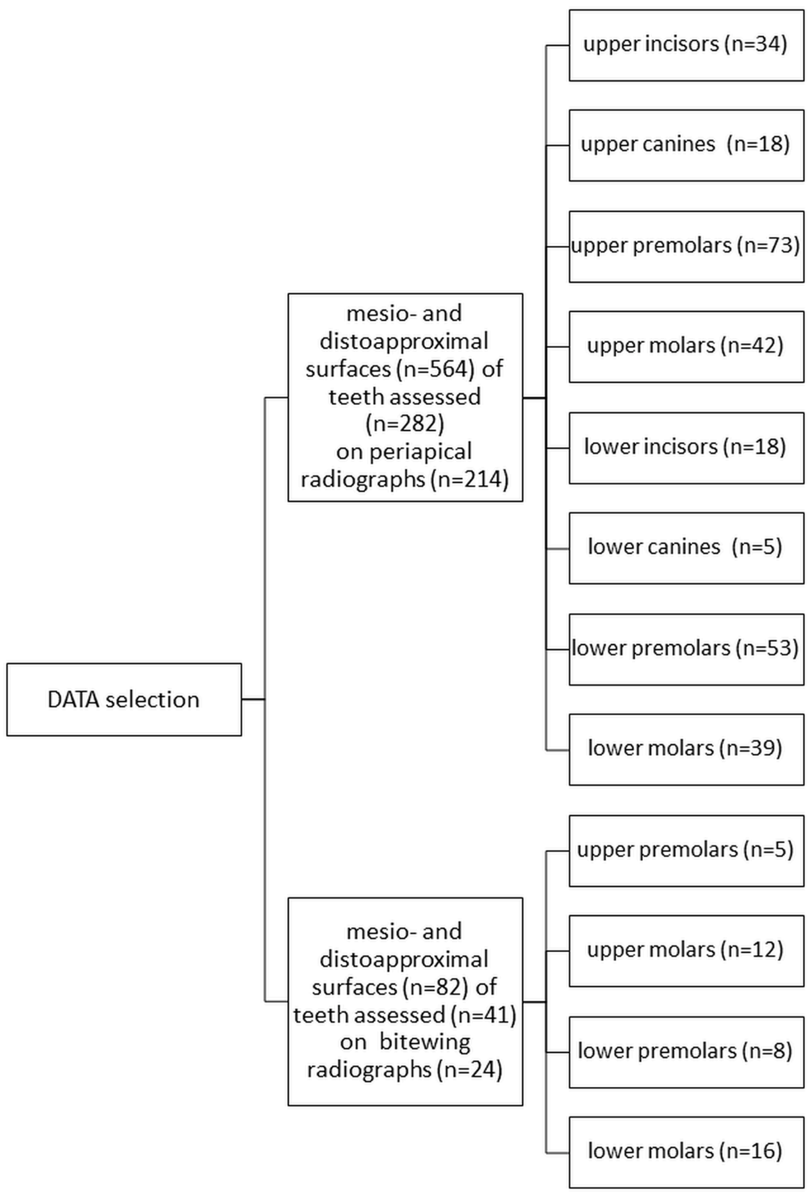

Fig. 1. Data selection flowchart. The number of selected radiographs, proximal surfaces, and the visible types of teeth are depicted.

computed tomography (CBCT) examination. Additionally, Lee et al. [2] concluded that the long-term cost of oral healthcare could potentially decrease with the accurate detection of carious lesions, a goal supported by deep learning algorithms.

Currently, there is increasing interest in the scientific literature on the use of AI-based algorithms for the diagnosis of caries [2,5,10,23,28–36]. Notably, publications based on dental radiographic data include bitewing [5,29,31,32] and periapical radiographs [2,10,34]. The aim of the present study was to investigate the accuracy — sensitivity, specificity, positive and negative predictive values, and the correctly classified proportion values — of an AI-based tool (Diagnocat, DC) assisting in the healthcare process during the radiographic assessment of caries on intraoral radiographs compared to the ground truth as determined by dentomaxillofacial radiologists. The null hypothesis for this study was that there was no agreement between human and AI observers regarding the diagnosis of carious lesions on intraoral radiographs.

2. Materials and methods

In this retrospective radiological study, 238 intraoral digital radiographs from 201 patients were selected. Selection and collection were performed using IMPAX software (v.6.5.2.657, Agfa HealthCare, Mortsel, Belgium). The selection criteria were that the coronal part of the tooth had to be *in toto* visible on the periapical or bitewing radiograph and that no structure was projected on the examined tooth: that is, only teeth with overlap-free projection were selected. A total of 214 periapical radiographs containing 282 teeth and 24 bitewing radiographs containing 41 teeth that met all examination criteria were selected (Fig. 1).

Radiographs were acquired at the Department of Oral Diagnostics, Faculty of Dentistry, Semmelweis University using a Gendex 765DC X-ray appliance (65 kV, 7 mA; Gendex Dental Systems, Hatfield, PA, USA) with Gendex GXS-700 intraoral sensors (size 1 or 2; Gendex Dental Systems, Hatfield, PA, USA). The metadata of the radiographs were recorded, including the date of the study, age of the patient, and eligible tooth or teeth shown on the radiograph, which were stored independently from the downloaded image data. The observers and the Diagnocat (DC) system were provided with encrypted anonymised image files. The study protocol was performed in accordance with the Declaration of Helsinki and approved by the Semmelweis University Regional and Institutional Committee of Science and Research Ethics (SE RKEB 138/2020). Only observers had access to anonymised image files, and only human observers had access to metadata. The sample selection process did not prioritise any specific sex.

The 626 proximal surfaces of the 323 selected teeth were evaluated by two independent observers using the IMPAX software: a fifth-year dental student (DS) and a dentomaxillofacial radiologist (DR) with more than ten years of experience. During radiographic evaluation, human observers could change certain parameters, such as brightness or contrast, and use magnification. All radiographs were evaluated using the same monitor: Samsung S24F350FHU (full HD, resolution: 1920 × 1200 pixels; Samsung, Seoul, South Korea). If the mesio- and disto-approximal surfaces were considered intact, the human observer assigned a value of '0' to the tooth. However, if a carious lesion was found on the proximal surface, a value of '1' was assigned to the tooth. The ground truth was determined during online sessions supervised by senior dentomaxillofacial radiologists, with > 20 years of experience. In cases of conflicting detection of caries on selected intraoral radiographs, consensus was reached in all cases. The DS, as observer, was not involved in determining the ground truth, nor did he have any knowledge of the results.

Regarding the model pipeline for tooth detection of the assessed AI tool, which was independently developed, a two-stage object detector, a Mask R-CNN, was used with pre-trained ResNet-101 [37]. The output of the model, for which 4500 radiographs were used for training, contained rectangular boxes and segmentation masks with the predicted numbers of teeth. The Cascade R-CNN [38] architecture was selected to generate the AI model. It is distinct from the Mask R-CNN in that it utilises an iterative process to improve box prediction and averages the results for each cascade layer to determine the class predictions [22]. None of the networks were initialised with pre-trained weights and they were trained from scratch. A combination of Jaccard loss and cross-entropy losses was utilised for the segmentation tasks. The anatomical elements were labelled separately, whereas all signs of caries were labelled as a single class and assigned to specific teeth.

Anonymised radiographs containing no patient data were imported into the DC system. After a short analysis, the completed evaluation was displayed (Fig. 2).

The DC can identify teeth seen on radiographs, which can then be modified by the user. The detected teeth and lesions diagnosed are indicated in a separate image, as shown in Fig. 3.

The applied CNN often associates a probability value with the obtained diagnoses and indicates its location in a green-bordered area (Fig. 4).

The DC evaluated the teeth according to criteria similar to those used by human observers. If caries were detected, the term 'sign of caries' was displayed in the tooth's evaluation box. The anatomical localisation of the detected carious lesion was indicated using the DC with a green-bordered bounding box as shown in Fig. 4. If a carious lesion was highlighted by the DC on the mesio- or disto-approximal surface, it was manually recorded, and a value of '1' was assigned to the tooth (Phase 1; Fig. 5).

Four months later, the results provided by the CNN were presented to human observers who opened the AI tool and viewed the radiographic findings of all selected teeth reported by the CNN. The DS and DR, aided by the previous AI results, evaluated the same data and the findings were recorded (Phase 2). Subsequently, based on the DS, DR, and DC results, a new CNN was developed. Regardless of the radiographs selected in the present study, intraoral radiographs were used to retrain and test the new model at rates of 25 % and 75 %, respectively. Following the release of the updated and tested CNN, identical image data containing the originally selected 323 teeth were imported into the CNN, and the evaluation was executed as described above (Phase 3). Subsequently, the teeth for which human diagreement was recorded were excluded (Phase 4).

Descriptive statistics were used to quantify the proximal carious lesions. R software was used for statistical analysis [39]. To calculate Cohen's kappa coefficients and diagnostic test parameters (sensitivity, specificity, predictive values, and correctly classified proportion value) with 95 % confidence intervals, the 'epiR' R-package was employed [40], whereas for Fleiss' kappa (with 95 % confidence level) the 'irr' R-package [41] was used. The Fleiss' method was used for kappa calculations [42]. For the diagnostic test parameters, the exact binomial 95 % confidence interval was calculated to determine the performance of the DC in diagnosing carious lesions for the selected image data. Statistical significance was set at $p < 0.05$.

3. Results

A total of 626 proximal surfaces were evaluated for the 323 selected teeth, of which 41 (12.7 %) were found on the bitewing and 282 (87.3 %) were found on periapical radiographs.

During Phase 1, the DS identified 126 (39 %) cases of carious lesions; no carious lesions were identified in 197 (61 %) teeth (Table 1). The DR recorded carious lesions in 140 (43.3 %) teeth and 183 (56,7 %) teeth without carious lesions.

The CNN evaluated 322 available teeth; however, one tooth could not be identified. Among the recognised cases, caries were identified in 80 (24.8 %) cases, and in caries were not indicated in 242 (75.2 %) cases (Table 2).

Tables 1 and 2 show the results of Phase 2, in which the observers evaluated the same image data with the assistance of the CNN

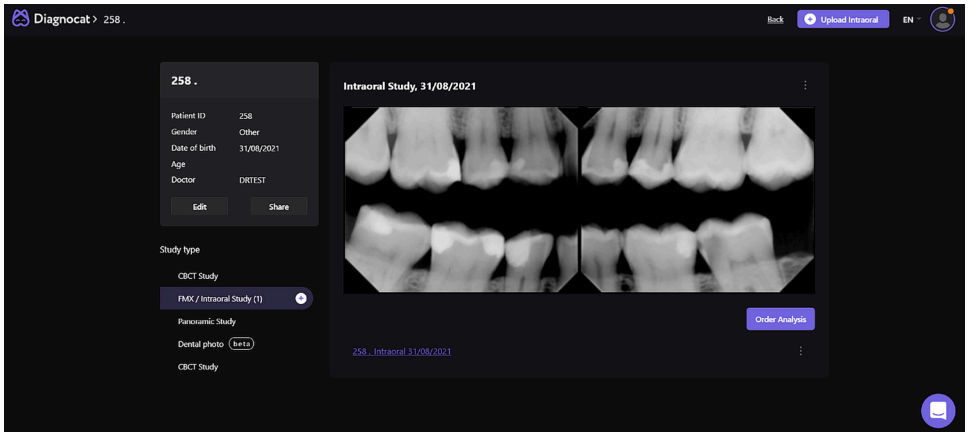

Fig. 2. User interface of the Diagnocat. It shows the bitewing radiographs of a case uploaded prior to the assessment.

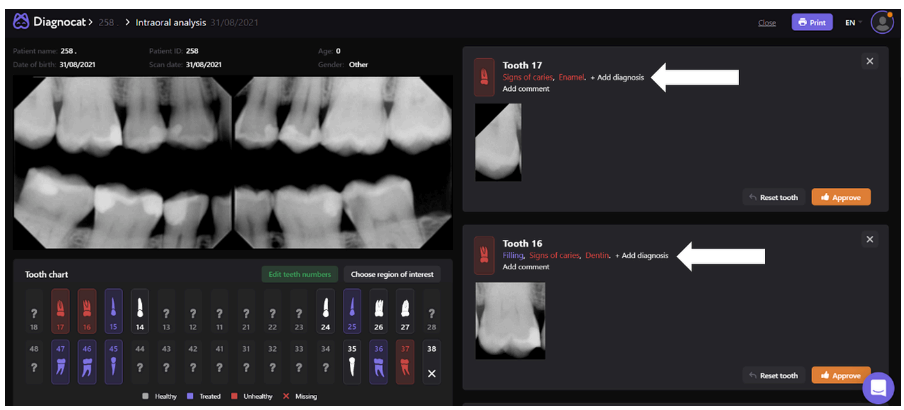

Fig. 3. Complete evaluation of a case. The detected and evaluated teeth are shown in the lower left part of the image. Arrows indicate the detected radiographic diagnoses, including caries: 'signs of caries'.

assessment results from four months earlier. The evaluation results of the imporved CNN are presented in Table 2 (Phase 3). The DC evaluated 322 teeth, as one of the teeth was not recognised, which was the same as that observed in Phases 1 and 2. The DC identified carious lesions in 115 (35.7 %) teeth but did not indicate caries in 207 (64.3 %) teeth (Table 2). The Phase 4 results are presented in Tables 1 and 2, where 21 samples with overlapping crown surfaces and human disagreement were eliminated from the data.

Cohen's and Fleiss' kappa coefficients were calculated to assess the agreement between human observers and DC. In our study, all Fleiss' kappa coefficient values differed significantly from zero ($P < 0.01$) (Table 3).

A mosaic plot was prepared (Fig. 6) to illustrate the relationship between each observer's agreement during Phase 4. The x-axis denotes the findings of the human observers: a value of '0' indicates the absence of caries, and a value of '1' indicates the presence of a carious lesion.

The y-axis illustrates the results of the evaluation using the DC. TN indicates the area where both the DC and human observers assigned the value '0′, so none of the observers indicated the presence of caries on the mesio- or distoapproximal surfaces of the assessed tooth. TP shows cases where caries were present, according to both human observers and the DC. An FP result was obtained if caries were detected by the DC but not by any human observer. The size of the FN area may play a particularly important role in the development of CNNs, illustrating cases in which the CNN fails to indicate a carious lesion despite its presence, according to human observers.

The sensitivity, specificity, positive and negative predictive values, and correctly classified proportion values are listed in Table 4. It is noteworthy that the correctly classified proportion value can be considered as the diagnostic accuracy because it is expressed as the proportion of correct predictions (TP+TN) among all predictions (TP+TN+FP+FN) [43,44]. The reality of the results from the dento-maxillofacial observer in the present study was determined by calculating the listed parameters.

4. Discussion

To enhance the accuracy and performance of the radiological evaluation of caries, an image analysis system using AI has been developed [29], including the DC used in this study. The main objective of this study was to determine the reliability of the DC in the diagnosis of caries.

Several studies have examined the accuracy of caries diagnosis using AI [2,5,9,23,29,32,45,46]. Schwendicke et al. [46] used CNNs to diagnose caries based on near-infrared light transillumination images. A system, using the near-infrared light, was employed to examine 223 extracted teeth, and the resulting images were processed using Resnet18 and Resnext50 CNNs. With a caries prevalence of 41 %, the sensitivity, specificity, and accuracy values for Resnext50 were 0.59, 0.76, and 0.68, respectively. Conversely, sensitivity (0.76), specificity (0.93), and accuracy (0.86) were higher in our study. Similar to Schwendicke et al. [46], Devito et al. [9] examined 80 extracted human teeth from which bitewing radiographs were obtained. Twenty-five dental specialists and a neural network assessed the presence of proximal caries on radiographs. Significantly better results were achieved using neural networks than human observers, with a ROC curve areas of 0.884 and 0.634, respectively.

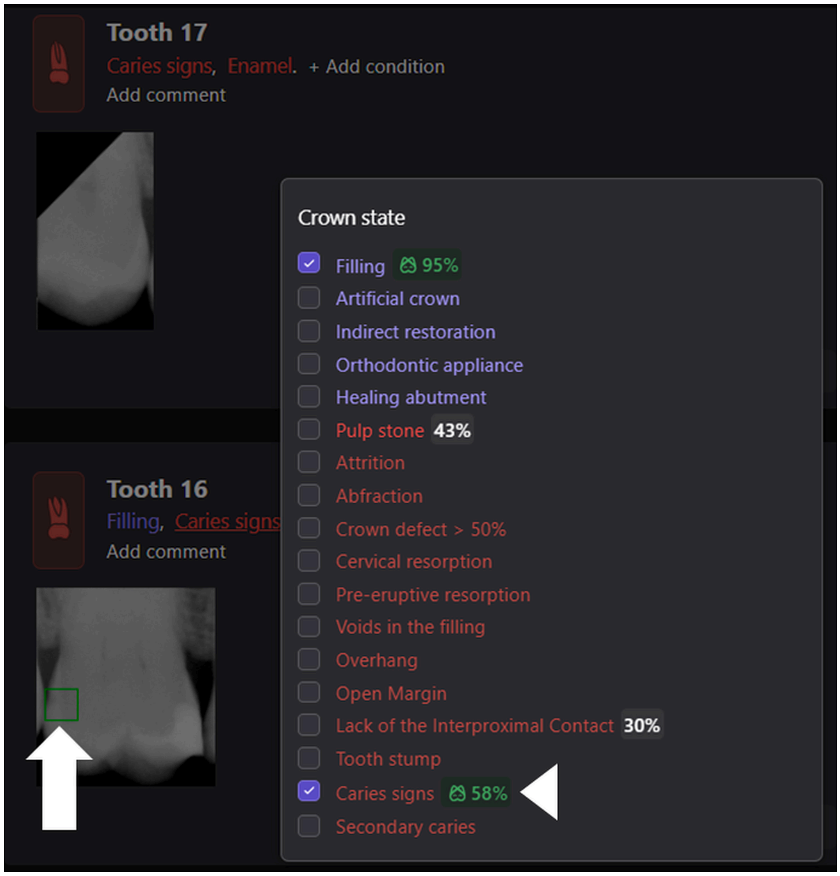

Fig. 4. The probability value associated with the diagnosis (arrowhead) and lesion is indicated by a green rectangle (arrow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

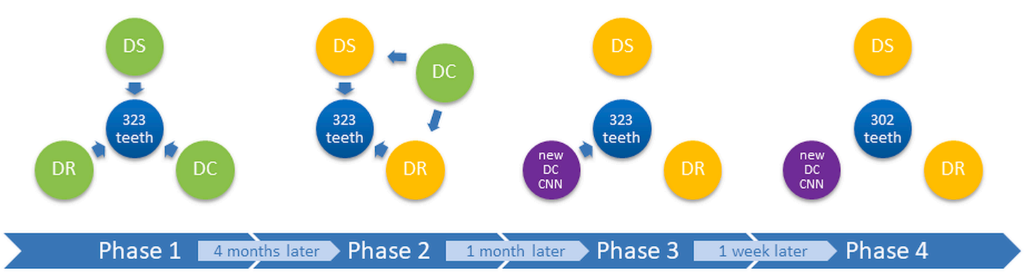

Fig. 5. The phases of evaluation. Green circles indicate the results of human and AI observers in Phase 1; orange circles indicate the results of observers carried out four months later during Phase 2; Phase 3 shows that the originally selected image data were evaluated by the updated CNN; Phase 4 demonstrates the exclusion of teeth that showed conflicting results when assessed by human observers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Agreement of human observers in Phases 1–4.

		Phase 1		Phase 2		Phase 3		Phase 4	
					DR				
		no caries	caries	no caries	caries	no caries	caries	no caries	caries
DS	no caries	164 (50.8 %)	33 (10.2 %)	191 (59.1 %)	4 (1.2 %)	191 (59.1 %)	4 (1.2 %)	183 (60.6 %)	0 (0 %)
	caries	19 (5.9 %)	107 (33.1 %)	8 (2.5 %)	120 (37.2 %)	8 (2.5 %)	120 (37.2 %)	0 (0 %)	119 (39.4 %)

Cantu et al. [29] examined 141 bitewing radiographs using a U-Net CNN for radiograph evaluation. Subsequently, the obtained results were compared with the observations of seven independent experienced dentists. The CNN determined caries lesions with a higher sensitivity value of 0.75 than the human observers, who diagnosed with only 0.36 sensitivity; however, the difference between the specificity values were lower at 0.91 and 0.83, respectively. Moreover, the overall diagnostic accuracy was 0.8 for of the neural network and 0.71 for the dentists. In our study, the DC program had a higher specificity, sensitivity, and diagnostic accuracy values when determining carious lesions; that is, a

Table 2
Agreement between human observers and the DC in Phases 1–4.

		Phase 1		Phase 2		Phase 3		Phase 4	
		DC							
		no caries	caries	no caries	caries	no caries	caries	no caries	caries
DS	**no caries**	190 (58.8 %)	6 (1.9 %)	190 (59.0 %)	4 (1.2 %)	170 (52.8 %)	24 (7.5 %)	169 (56.1 %)	13 (4.3 %)
	caries	52 (16.1 %)	74 (23.0 %)	52 (16.1 %)	76 (23.6 %)	37 (11.5 %)	91 (28.3 %)	29 (9.6 %)	90 (29.9 %)
	TOTAL	242 (75.2 %)	80 (24.8 %)	242 (75.2 %)	80 (24.8 %)	207 (64.3 %)	115 (35.7 %)	198 (63.6 %)	103 (34.1 %)
DR	**no caries**	174 (53.9 %)	8 (2.5 %)	193 (59.9 %)	5 (1.6 %)	174 (54.0 %)	24 (7.5 %)	169 (56.1 %)	13 (4.3 %)
	caries	68 (21.1 %)	72 (22.4 %)	49 (15.2 %)	75 (23.3 %)	33 (10.2 %)	91 (28.3 %)	29 (9.6 %)	90 (29.9 %)
	TOTAL	242 (75.2 %)	80 (24.8 %)	242 (75.2 %)	80 (24.8 %)	207 (64.3 %)	115 (35.7 %)	198 (63.6 %)	103 (34.1 %)

Table 3
Cohen and Fleiss κ values according to the Phases 1–4.

		Phase 1	Phase 2	Phase 3	Phase 4
Cohen kappa	**DS vs. DR**	**0.66**	**0.92**	**0.92**	**1**
Cohen kappa	**DS vs. DC**	**0.58**	**0.61**	**0.60**	**0.70**
Cohen kappa	**DR vs. DC**	**0.49**	**0.62**	**0.62**	**0.70**
Fleiss kappa	**DS vs. DR vs. DC**	**0.57**	**0.72**	**0.71**	**0.80**

high specificity value can aid dentists in the evaluation caries diagnostics, particularly when the DC indicates the presence of caries. Schwendicke et al. [5] examined the cost-effectiveness of AI in the diagnosis of proximal caries, including a previous study by Cantu et al. [29], and found that AI-based assistance was more cost-effective and accurate in caries diagnosis than without AI. Lee et al. [32] investigated the role of CNNs in detecting early carious lesions and uploaded 50 bitewing radiographs to a U-Net CNN. The study showed a significant increase in sensitivity values for the three dentist observers when using the CNN (dentist 1, 85.34 %–92.12 %; dentist 2, 85.86 %–93.72 %; and dentist 3, 69.11 %–79.06 %).

In contrast, Lee et al. [2] used 2400 periapical radiographs to train the GoogLeNet Inception v3 CNN, processed 600 periapical radiographs using the system, and examined the efficacy of the neural network in detecting carious lesions. The sensitivity, specificity and diagnostic accuracy of evaluation of premolar and molar teeth were 0.81, 0.83, and 0.82, respectively. These results were comparable with our findings, and although the sensitivity value was higher than the value we obtained (0.76) for caries detection, the specificity and accuracy values were lower than the values of 0.93 and 0.86 obtained in the present study. Khan et al. [34] assessed the diagnostic efficacy of U-Net, Xnet, SegNet, and U-Net+Densenet121 deep learning architectures for the detection of caries, interradicular radiolucencies, and alveolar bone loss on periapical radiographs. U-Net and U-Net+Densenet121 achieved the best performance in caries detection, with a mean intersection over union values of 0.501 and 0.402, and with a Dice coefficient values of 0.569 and 0.453, respectively. Ari et al. [47] assessed 1169 periapical radiographs using a U-Net CNN for different radiographic findings, including carious and periapical lesions. The sensitivity for detecting caries was 0.82, which was consistent with the results of the present study (0.76). Similar to this study, Ezhov et al. [28] applied the DC system to assess CBCT images. Two groups of observers were assessed: one group performed a diagnosis supported by the DC and the other group worked without assistance from the CNN. The sensitivities for the aided and unaided groups for caries detection were 0.67 and 0.66, respectively, which are lower than our results.

The present study has several limitations. First, the major limitation was the absence of a clinical evaluation, which is essential for valid diagnosis. This could be addressed by collecting available clinical information for radiographs to be selected for future studies. On the other hand, bitewing and periapical radiographs taken using paralleling technique were consecutively selected in time if they met the selection

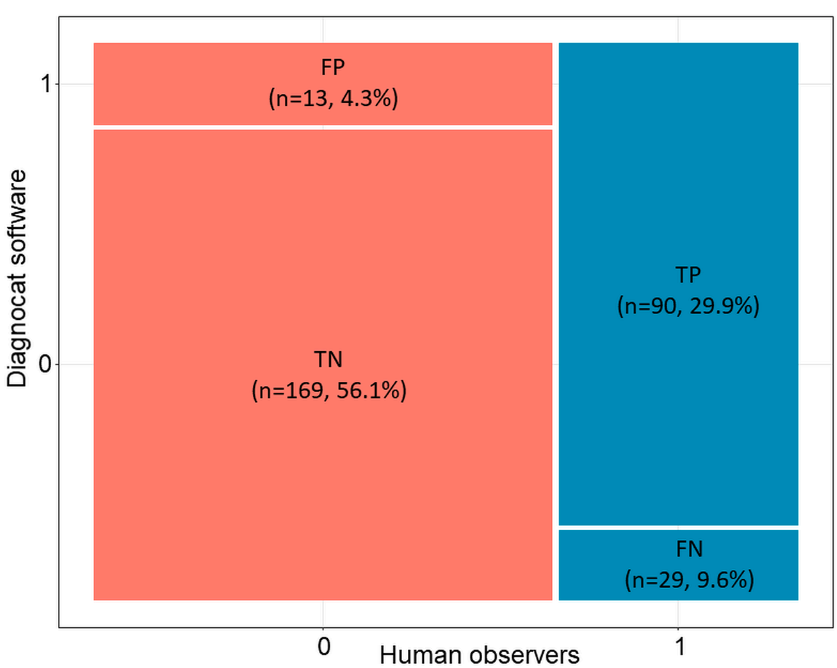

Fig. 6. The x-axis represents the evaluation results of human observers; a value of '0′ indicates the absence of caries, and a value of '1′ indicates the presence of a tooth. The y-axis shows the evaluation of the DC software (FP, false positive; TN, true negative; TP, true positive; FN, false negative).

Table 4
Sensitivity, specificity, positive and negative predictive values and correctly classified proportion values (95 % confidence interval limits) of the DC.

	Phase 1	Phase 2	Phase 3	Phase 4
sensitivity	0.51 (0.43, 0.60)	0.60 (0.51, 0.69)	0.73 (0.65, 0.81)	0.76 (0.67, 0.83)
specificity	0.96 (0.91, 0.98)	0.97 (0.94, 0.99)	0.88 (0.83, 0.92)	0.93 (0.88, 0.96)
positive predictive value	0.90 (0.81, 0.96)	0.94 (0.86, 0.98)	0.79 (0.71, 0.86)	0.87 (0.79, 0.93)
negative predictive value	0.71 (0.65, 0.77)	0.80 (0.74, 0.85)	0.84 (0.78, 0.89)	0.85 (0.80, 0.90)
correctly classified proportion value	0.76 (0.71, 0.81)	0.83 (0.79, 0.87)	0.82 (0.78, 0.86)	0.86 (0.82, 0.90)

criteria mentioned above. In the scientific literature, bitewing radiographs have proven to be a superior imaging modality, especially for the diagnosis of early caries detection [48]. Thus, guidelines and position statements on radiographic modalities for caries diagnosis recommend bitewing radiographs for proximal sites if radiological imaging is justified [49,50], although dentists request a larger number of periapical images than bitewing radiographs [51]. Thus, we considered it reasonable to assess how reliable an AI tool can detect carious lesions on bitewing or periapical radiographs. Additionally, we aimed to determine the reliability of the applied AI tool with a larger number of bitewing radiographs during future studies. Another possible limitation is that the estimated depth of the detected carious lesion was not recorded, which may have affected the reliability of the assessed AI tool. In addition, we consider it necessary to increase the sample size and group the teeth according to their position in the dental arch to obtain more detailed information from the assessed CNN.

The present study examined the potential development of caries diagnostics through the application of AI on intraoral radiographs. The sensitivity (0.76), specificity (0.93), and diagnostic accuracy (0.86) of the DC CNN for caries detection were comparable to those in the available literature. Based on the present study, the reliability of the DC has not yet reached a level at which it can be used independently to diagnose caries. However, it can significantly aid dentists in evaluating intraoral radiographs, given the moderate to substantial agreement between the observers and CNN, as indicated by the determined Fleiss' kappa values. In the future, our research aims to broaden its scope by including additional test data and performing an intraobserver reliability test. We strive to contribute to the development of AI-based applications in diagnostic radiology to ensure the delivery of the highest quality and most efficient patient care.

5. Conclusions

Based on the agreement between human and CNN observers, as well as the sensitivity, specificity, and diagnostic accuracy values that are in line with the findings of the scientific literature, it can be concluded that DC has the potential to significantly assist dentists in the diagnosis of caries and enhance the efficiency of their work.

Acknowledgments

This study did not receive any specific grants from funding agencies in the public, commercial, or non-profit sectors.

CRediT authorship contribution statement

Viktor Szabó: Writing – original draft, Investigation. **Bence Tamás Szabó:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Kaan Orhan:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Dániel Sándor Veres:** Visualization, Formal analysis. **David Manulis:** Software, Data curation. **Matvey Ezhov:** Software, Resources. **Alex Sanders:** Resources, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

David Manulis, Matyey Ezhov and Alex Sanders are employees of Diagnocat Co. Ltd.. Kaan Orhan is a scientific research advisor for the Diagnocat Co. Ltd., San Francisco CA.

References

[1] R.H. Selwitz, A.I. Ismail, N.B. Pitts, Dental caries, Lancet 369 (2007) 51–59, https://doi.org/10.1016/S0140-6736(07)60031-2.

[2] J.H. Lee, D.H. Kim, S.N. Jeong, S.H. Choi, Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm, J. Dent. 77 (2018) 106–111, https://doi.org/10.1016/j.jdent.2018.07.015.

[3] P.Y.F. Wen, M.X. Chen, Y.J. Zhong, Q.Q. Dong, H.M. Wong, Global burden and inequality of dental caries, J. Dent. Res. 101 (2022) 392–399, https://doi.org/10.1177/00220345211056247.

[4] J.E. Frencken, P. Sharma, L. Stenhouse, D. Green, D. Laverty, T. Dietrich, Global epidemiology of dental caries and severe periodontitis – a comprehensive review, J. Clin. Periodontol. 44 (18) (2017) S94–S105, https://doi.org/10.1111/jcpe.12677. Suppl.

[5] F. Schwendicke, J.G. Rossi, G. Gostemeyer, K. Elhennawy, A.G. Cantu, R. Gaudin, A. Chaurasia, S. Gehrung, J. Krois, Cost-effectiveness of artificial intelligence for proximal caries detection, J. Dent. Res. 100 (2021) 369–376, https://doi.org/10.1177/0022034520972335.

[6] J.R. Keenan, A.V. Keenan, Accuracy of dental radiographs for caries detection, Evid. Based Dent. 17 (2016) 43, https://doi.org/10.1038/sj.ebd.6401166. –43.

[7] A. Wenzel, Radiographic modalities for diagnosis of caries in a historical perspective: from film to machine-intelligence supported systems, Dentomaxillofac. Radiol. 50 (2021) 20210010, https://doi.org/10.1259/dmfr.20210010.

[8] C.M. Pine, J.J. ten Bosch, Dynamics of and diagnostic methods for detecting small carious lesions, Caries. Res. 30 (1996) 381–388, https://doi.org/10.1159/000262348.

[9] K.L. Devito, F. de Souza Barbosa, W.N. Felippe Filho, An artificial multilayer perceptron neural network for diagnosis of proximal dental caries, Oral. Surg. Oral. Med. Oral. Pathol. Oral. Radiol. Endod. 106 (2008) 879–884, https://doi.org/10.1016/j.tripleo.2008.03.002.

[10] X. Lin, D. Hong, D. Zhang, M. Huang, H. Yu, Detecting proximal caries on periapical radiographs using convolutional neural networks with different training strategies on small datasets, Diagnostics 12 (2022) 1047, https://doi.org/10.3390/diagnostics12051047.

[11] K. Kamburoğlu, E. Kolsuz, S. Murat, S. Yüksel, T. Özen, Proximal caries detection accuracy using intraoral bitewing radiography, extraoral bitewing radiography and panoramic radiography, Dentomaxillofac. Radiol. 41 (2012) 450–459, https://doi.org/10.1259/dmfr/30526171.

[12] A. Wenzel, Bitewing and digital bitewing radiography for detection of caries lesions, J. Dent. Res. 83 (2004) 72–75, https://doi.org/10.1177/154405910408301s14.

[13] Z.Z. Akarslan, M. Akdevelioglu, K. Gungor, H. Erten, A comparison of the diagnostic accuracy of bitewing, periapical, unfiltered and filtered digital panoramic images for approximal caries detection in posterior teeth, Dentomaxillofac. Radiol. 37 (2008) 458–463, https://doi.org/10.1259/dmfr/84698143.

[14] F. Schwendicke, G. Göstemeyer, Conventional bitewing radiography, Clin. Dent. Rev. 4 (2020) 22, https://doi.org/10.1007/s41894-020-00086-8.

[15] H. Amasya, M. Alkhader, G. Serindere, K. Futyma-Gąbka, C. Aktuna Belgin, M. Gusarev, M. Ezhov, I. Różyło-Kalinowska, M. Önder, A. Sanders, A.L.F. Costa, S. L.P. Castro Lopes, K. Orhan, Evaluation of a decision support system developed with deep learning approach for detecting dental caries with cone-beam computed tomography imaging, Diagnostics 13 (2023) 3471, https://doi.org/10.3390/diagnostics13223471.

[16] Y. Mintz, R. Brodie, Introduction to artificial intelligence in medicine, Minim. Invasive Ther. Allied Technol. 28 (2019) 73–81, https://doi.org/10.1080/13645706.2019.1575882.

[17] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, C.I. Sanchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, 10.1016/10.1016/j.media.2017.07.005.

[18] M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H.J. Bae, N. Kim, Deep learning in medical imaging, Neurospine 16 (2019) 657–668, https://doi.org/10.14245/ns.1938396.198.

[19] J. Ker, L. Wang, J. Rao, T. Lim, Deep learning applications in medical image analysis, Comput. Sci. 6 (2018) 9375–9389, https://doi.org/10.1109/ACCESS.2017.2788044.
[20] M.H. Hesamian, W. Jia, X. He, P. Kennedy, Deep learning techniques for medical image segmentation: achievements and challenges, J. Digit. Imaging. 32 (2019) 582–596, https://doi.org/10.1007/s10278-019-00227-x.
[21] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, navab, N., Hornegger, J., Wells, W., Frangi, A. (eds), in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, 9351, Springer, Cham, 2015, https://doi.org/10.1007/978-3-319-24574-4_28.
[22] K. Orhan, C. Aktuna Belgin, D. Manulis, M. Golitsyna, S. Bayrak, S. Aksoy, A. Sanders, M. Önder, M. Ezhov, M. Shamshiev, M. Gusarev, V. Shlenskii, Determining the reliability of diagnosis and treatment using artificial intelligence software with panoramic radiographs, Imaging. Sci. Dent. 53 (2023) 199–208, https://doi.org/10.5624/isd.20230109.
[23] J. Kühnisch, O. Meyer, M. Hesenius, R. Hickel, V. Gruhn, Caries detection on intraoral images using artificial intelligence, J. Dent. Res. 101 (2021) 158–165, https://doi.org/10.1177/00220345211032524.
[24] X. Li, D. Zhao, J. Xie, H. Wen, C. Liu, Y. Li, W. Li, S. Wang, Deep learning for classifying the stages of periodontitis on dental images: a systematic review and meta-analysis, BMC Oral Health 23 (2023) 1017, https://doi.org/10.1186/s12903-023-03751-z.
[25] S. Sadr, H. Mohammad-Rahimi, S.R. Motamedian, S. Zahedrozegar, P. Motie, S. Vinayahalingam, O. Dianat, A. Nosrat, Deep learning for detection of periapical radiolucent lesions: a systematic review and meta-analysis of diagnostic test accuracy, J. Endod. 49 (2023) 248–261, https://doi.org/10.1016/j.joen.2022.12.007, e3.
[26] S.K Bayrakdar, K. Orhan, I.S. Bayrakdar, E. Bilgir, M. Ezhov, M. Gusarev, E. Shumilov, A deep learning approach for dental implant planning in cone-beam computed tomography images, BMC Med. Imaging 21 (2021) 86, https://doi.org/10.1186/s12880-021-00618-z.
[27] T.H. Farook, J.J. Dudley, Automation and deep (machine) learning in temporomandibular joint disorder radiomics: a systematic review, Oral. Rehabil. 50 (2023) 501–521, https://doi.org/10.1111/joor.13440.
[28] M. Ezhov, M. Gusarev, M. Golitsyna, J.M. Yates, E. Kushnerev, D. Tamimi, S. Aksoy, E. Shumilov, A. Sanders, K. Orhan, Clinically applicable artificial intelligence system for dental diagnosis with CBCT, Sci. Rep. 11 (2021) 1–12, https://doi.org/10.1038/s41598-021-01678-5.
[29] A.G. Cantu, S. Gehrung, J. Krois, A. Chaurasia, J.G. Rossi, R. Gaudin, K. Elhennawy, F. Schwendicke, Detecting caries lesions of different radiographic extension on bitewings using deep learning, J. Dent. 100 (2020) 103425, https://doi.org/10.1016/j.jdent.2020.103425.
[30] H. Mohammad-Rahimi, S.R. Motamedian, M.H. Rohban, J. Krois, S.E. Uribe, E. Mahmoudinia, R. Rokhshad, M. Nadimi, F. Schwendicke, Deep learning for caries detection: a systematic review, J. Dent. 122 (2022) 104115, https://doi.org/10.1016/j.jdent.2022.104115.
[31] I.S. Bayrakdar, K. Orhan, S. Akarsu, Ö. Çelik, S. Atasoy, A. Pekince, Y. Yasa, E. Bilgir, H. Sağlam, A.F. Aslan, A. Odabaş, Deep-learning approach for caries detection and segmentation on dental bitewing radiographs, Oral. Radiol. 38 (2021) 468–479, https://doi.org/10.1007/s11282-021-00577-9.
[32] S. Lee, S.I. Oh, J. Jo, S. Kang, Y. Shin, J.W. Park, Deep learning for early dental caries detection in bitewing radiographs, Sci. Rep. 11 (2021) 1–14, https://doi.org/10.1038/s41598-021-96368-7.
[33] H. Devlin, T. Williams, J. Graham, M. Ashley, The ADEPT study: a comparative study of dentists' ability to detect enamel-only proximal caries in bitewing radiographs with and without the use of AssistDent artificial intelligence software, Br. Dent. J. 231 (2021) 481–485, https://doi.org/10.1038/s41415-021-3526-6.
[34] H.A. Khan, M.A. Haider, H.A. Ansari, H. Ishaq, A. Kiyani, K. Sohail, M. Muhammad, S.A. Khurram, Automated feature detection in dental periapical radiographs by using deep learning, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. 131 (2021) 711–720, https://doi.org/10.1016/j.oooo.2020.08.024.
[35] B. Üzen H Dayı, İ.B. Çiçek, Ş.B. Duman, A novel deep learning-based approach for segmentation of different type caries lesions on panoramic radiographs, Diagnostics 13 (2023) 202, https://doi.org/10.3390/diagnostics13020202.
[36] F. Oztekin, O. Katar, F. Sadak, M. Yildirim, H. Cakar, M. Aydogan, Z. Ozpolat, T. Talo Yildirim, O. Yildirim, O. Faust, U. Rajendra Acharya, An explainable deep learning model to prediction dental caries using panoramic radiograph images, Diagnostics 13 (2023) 226, https://doi.org/10.3390/diagnostics13020226.
[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, arXiv: 1512.03385 (2015). 10.48550/arXiv.1512.03385.
[38] Z. Cai, N. Vasconcelos, R.C.N.N. Cascade, Delving into high quality object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1–9.
[39] V.R.C. Team, R: A Language and Environment For Statistical Computing. R Foundation For Statistical Computing, Austria, 2022. https://www.R-project.org/. accessed 13 February.
[40] T.N.M. Stevenson, C. Heuer, J. Marshall, J. Sanchez, R. Thornton, J. Reiczigel, J. Robison-Cox, P. Sebastiani, and P. Solymos. epiR: tools for the analysis of epidemiological Data. https://CRAN.R-project.org/package=epiR/, 2021 (accessed 13 February 2022).
[41] J. Lemon, M. Gamer, I. Fellows, Puspendra Singh Irr: Various Coefficients of Interrater Reliability and Agreement, 2019. https://www.r-project.org/. accessed 13 February 2022.
[42] J.L. Fleiss, Measuring nominal scale agreement among many raters, Psychol. Bull. 76 (1971) 378–382, https://doi.org/10.1037/h0031619.
[43] A.M. Šimundić, Measures of diagnostic accuracy: basic definitions, EJIFCC 19 (2009) 203–211.
[44] A.C. Stahl, A.S. Tietz, B. Kendziora, M. Dewey, Has the STARD statement improved the quality of reporting of diagnostic accuracy studies published in European Radiology? Eur. Radiol. 33 (2023) 97–105, https://doi.org/10.1007/s00330-022-09008-7.
[45] M. Moran, M. Faria, G. Giraldi, L. Bastos, L. Oliveira, A. Conci, Classification of approximal caries in bitewing radiographs using convolutional neural networks, Sensors 21 (2021) 5192, https://doi.org/10.3390/s21155192.
[46] F. Schwendicke, K. Elhennawy, S. Paris, P. Friebertshauser, J. Krois, Deep learning for caries lesion detection in near-infrared light transillumination images: a pilot study, J. Dent. 92 (2020) 103260, https://doi.org/10.1016/j.jdent.2019.103260.
[47] T. Ari, H. Sağlam, H. Öksüzoğlu, O. Kazan, İ.Ş. Bayrakdar, S.B. Duman, Ö. Çelik, R. Jagtap, K. Futyma-Gąbka, I. Różyło-Kalinowska, K. Orhan, Automatic feature segmentation in dental periapical radiographs, Diagnostics 12 (2022) 3081, https://doi.org/10.3390/diagnostics12123081.
[48] N. Takahashi, C. Lee, J.D. Da Silva, H. Ohyama, M. Roppongi, H. Kihara, W. Hatakeyama, S. Ishikawa-Nagai, M. Izumisawa, A comparison of diagnosis of early stage interproximal caries with bitewing radiographs and periapical images using consensus reference, Dentomaxillofac. Radiol. 48 (2019) 20170450, https://doi.org/10.1259/dmfr.20170450.
[49] K.W. Neuhaus, J. Kühnisch, A. Banerjee, S. Martignon, D. Ricketts, F. Schwendicke, M.H. van der Veen, S. Doméjean, M. Fontana, A. Lussi, A. Jablonski-Momeni, F. M. Mendes, G. Douglas, G. Schmalz, G. Campus, J. Aps, K. Horner, N. Opdam, M. C. Huysmans, C.H. Splieth, ORCA-EFCD consensus report on clinical recommendations for caries diagnosis. Paper II: caries lesion activity and progression assessment, Caries Res. (2024) 1, https://doi.org/10.1159/000538619. -1.
[50] J. Kühnisch, J.K. Aps, C. Splieth, A. Lussi, A. Jablonski-Momeni, F.M. Mendes, G. Schmalz, M. Fontana, A. Banerjee, D. Ricketts, F. Schwendicke, G. Douglas, G. Campus, M. van der Veen, N. Opdam, S. Doméjean, S. Martignon, K.W. Neuhaus, K. Horner, M.D. Huysmans, ORCA-EFCD consensus report on clinical recommendation for caries diagnosis. Paper I: caries lesion detection and depth assessment, Clin. Oral. Investig. 28 (2024) 227, https://doi.org/10.1007/s00784-024-05597-3.
[51] P. Grieco, A. Jivraj, J. Da Silva, Y. Kuwajima, Y. Ishida, K. Ogawa, H. Ohyama, S. Ishikawa-Nagai, Importance of bitewing radiographs for the early detection of interproximal carious lesions and the impact on healthcare expenditure in Japan, Ann. Transl. Med. 10 (2022) 2, https://doi.org/10.21037/atm-21-2197.