Title page

title: Validation of Artificial Intelligence Application for Dental Caries Diagnosis on Intraoral Bitewing and Periapical Radiographs

short title: AI-Based Dental Caries Detection

Bence Tamas Szabo[1], Kaan Orhan[2,3], Dániel Sándor Veres[4], Viktor Szabó[1], Maria Golitsyna[5], David Manilus[5], Matvey Ezhov[5], Alex Sanders[5]

1 Department of Oral Diagnostics, Faculty of Dentistry, Semmelweis University, Budapest, Hungary

2 Ankara University, Faculty of Dentistry, Department of Dentomaxillofacial Radiology, Ankara, Turkey

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

4 Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary

5 Diagnocat Inc, San Francisco, CA, USA

corresponding author:

Bence Tamas Szabo

47 Szentkiralyi str., Budapest, 1088, Hungary

e-mail: szabo.bence_tamas@dent.semmelweis-univ.hu

phone and fax number: +3614591500 / 59161 (ext.)

Keywords: Artificial intelligence, Deep Learning, Machine Learning, Dental Digital Radiography, Dental Caries, Diagnostic Imaging

1. Introduction

Radiology is one of the health care fields where artificial intelligence (AI) was one of the first to appear. This may be due to a large amount of digital data available here, which is one of the bases of AI [1]. Neural networks, including the convolutional neural networks (CNN), are involved in radiological image analysis. Due to the diversity of images and structures to be distinguished, different software use not only one CNN to solve a selected task, but a combination of several CNNs, each performing a specific subtask and the final output will be the sum of these [2]. This process is called computer-aided diagnosis (CAD), where the output generated by the program draws the clinician's attention to areas of the image that appear abnormality and helps to make the diagnosis [3]. This can greatly support and speed up the work of radiologists. The most commonly used imaging modality in dental radiology is conventional radiography [4, 5]. Hundreds of radiographs can be made daily at larger clinics, which could indicate the need for the construction of an AI-based system that can facilitate the work of clinicians and speed up patient care. At the same time, this large amount of images is beneficial to AI-based software, as it provides the software with a constant and large dataset, i.e. big data with which the software continues to learn and thereby evolve. In the field of dental imaging CNNs, as subgroups of deep learning, are widely used [6-17]. These networks can be used in image analysis for classification, detection, or even segmentation. Classification is a wide-ranging task in which the system decides whether the searched structure in the given area of the image is present [18]. Detection is a very similar algorithm that locates and identifies specific areas of the image where the desired lesion is found [19]. During segmentation, the software identifies and labels the particular part of the image where the potential pathological lesion is located [20]. Figure 1 illustrates this process through the recognition of dental caries.

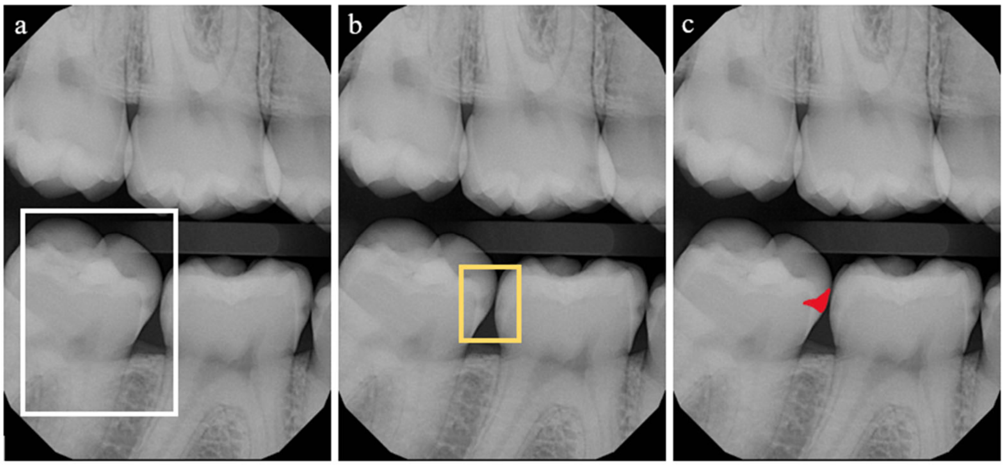

Figure 1. a. classification; b. detection; c. segmentation

This image analysis software can already be used to perform many tasks. Establishing a proper diagnosis is one of the cornerstones of any medical intervention. An important additional diagnostic method for this is radiology examination, which can be used to diagnose, for example, dental caries [11]. AI may help in the accuracy and speed of this process [8]. Dental caries is a chronic process that begins on the surface of a tooth in contact with the oral cavity and progresses from the surface to the deep, tooth tissue. Dental caries is still one of the most common chronic diseases worldwide [11, 21]. Although regular epidemiological studies have shown a reduction in the prevalence of caries within the population in developed countries in recent decades, this figure is still too high worldwide, especially in underdeveloped areas [22]. X-ray imaging is still of paramount importance in diagnosing caries. Radiologically, caries lesions can be grouped based on the surface area. Caries lesions in proximal areas often develop and grow unnoticed [23]. In several cases, visual examination of teeth in the posterior region is difficult to detect, and therefore radiographs play a key role in their diagnosis [23]. Intraoral radiographs, such as bitewing radiographs can be used to diagnose caries lesions [24], which provide information on the condition of the teeth from the distal surface of the crown of the canine to the distal surface of the last erupted crown [25]. These are the most suitable and accurate radiographic modalities in the recognition of proximal caries, and caries lesions in these regions can only be detected in time with this type of modality in most cases [24, 26]. Among these publications, which are based on radiographic data, several authors used bitewing [7, 8, 13, 15], and some of them periapical radiographs [11, 14, 17]. Currently, there is an increasing interest in the scientific literature using AI to aid in the diagnosis of caries [7, 8, 10-15, 17, 27, 28]. This study aimed to investigate the accuracy of an AI-based Diagnocat software (*Diagnocat Inc., San Francisco, CA, USA*) assisting the health care process in the radiographic diagnosis of caries on intraoral radiographs.

2. Materials and methods

In our retrospective radiological study, a total of 238 intraoral digital X-ray images of 201 patients were selected. Selection and collection were performed using the IMPAX software (v.6.5.2.657, *Agfa HealthCare, Mortsel, Belgium*). Periapical and bitewing radiographs were selected, which met all examination criteria. The selection criterion was that the coronal part of the tooth must have been *in toto* visible on the periapical or bitewing radiograph and that no structure was projected on the examined tooth, i.e. only teeth with overlap-free projection were selected. The radiographs were acquired with the aid of a Gendex 765DC X-ray appliance (65kV, 7mA; *Gendex Dental Systems, Hatfield, PA, USA*) with Gendex GXS-700 intraoral sensors (size 1 or 2; *Gendex Dental Systems, Hatfield, PA, USA*) at the Department of Oral Diagnostics, Faculty of Dentistry, Semmelweis University. Data of the selected radiographs were recorded, such as the date of the study, the age of the patient, and the eligible tooth or teeth are shown on the radiograph. The study protocol was performed according to 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 the collected data and images. No gender was preferred for sample selection.

Proximal surfaces of the 302 selected teeth were evaluated by using the IMPAX software by two independent observers: a fifth-year dental student and a dentomaxillofacial radiologist with more than ten years of experience. During the evaluation of the radiograph, human observers had the opportunity to change certain parameters such as brightness or contrast and were able to use magnification. All radiographs were evaluated on the same monitor: Samsung S24F350FHU (full HD, resolution: 1920x1200 pixels; *Samsung, Seoul, South Korea*). If the surface was considered intact, the human observer assigned a value of ’0’ to the surface, if a caries lesion was found a value of ’1’ was assigned to the surface. The anonymized radiographs containing no patient data were imported into the Diagnocat software, and after a short analysis, the completed evaluation was displayed (Figure 2).

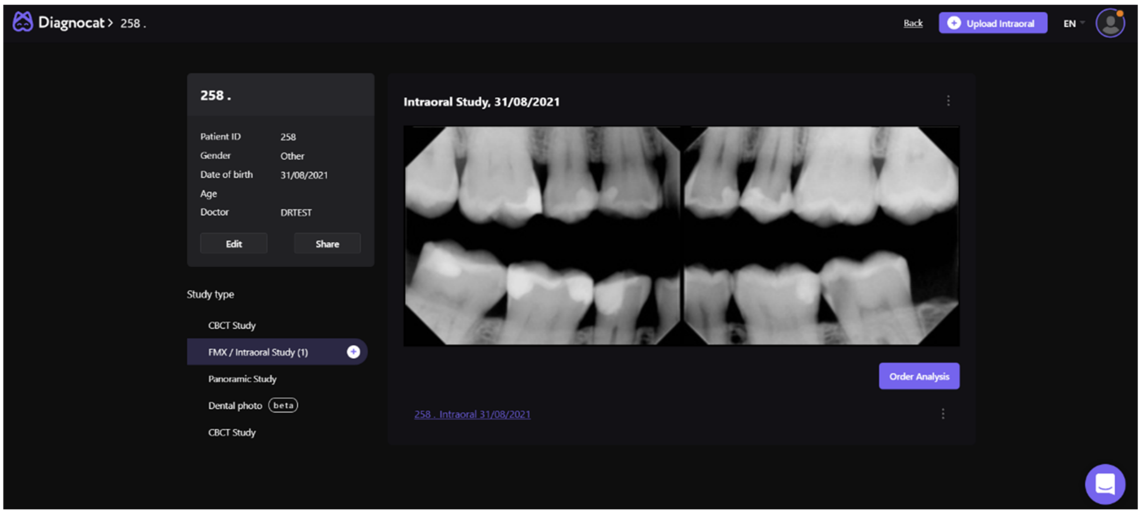

Figure 2. The user interface of Diagnocat

The software can identify the teeth seen on the radiographs, which can be modified by the user afterward. Detected teeth and lesions diagnosed on them are indicated on a separate image as shown in Figure 3.

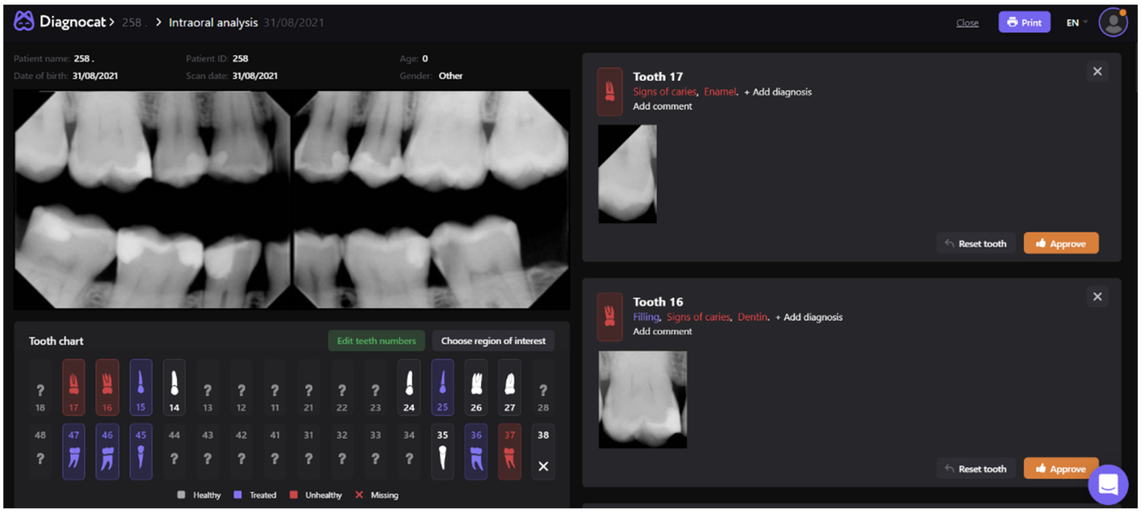

Figure 3. Completed evaluation of a case

The program often associates a probability value with the obtained diagnoses and indicates its location with a green-bordered area (Figure 4).

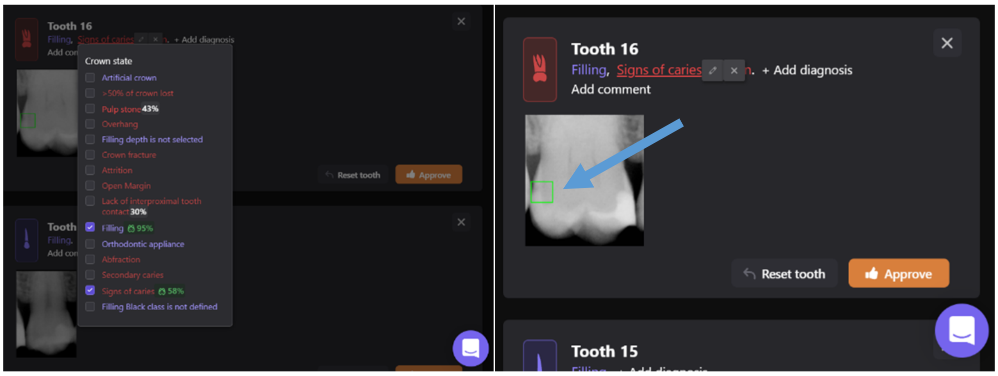

Figure 4. The probability value associated with the diagnosis and the lesion is indicated by a green rectangle

The software evaluated the given teeth according to similar criteria as human observers. If it indicated caries, a "sign of caries" appeared in the evaluation box of the tooth. The anatomical localization of the caries lesion was detected as indicated by the software with a green-bordered area as shown in Figure 4. If a caries lesion was highlighted by the software on the mesio- or distoapproximal surface, it was manually recorded and a value of '1' was assigned.

For the statistical analysis R software was used (v.4.0.4., *R Foundation for Statistical Computing, Vienna, Austria*). Descriptive statistics were used to quantify the detected proximal caries lesions. For interobserver reliability tests, the Fleiss kappa coefficient was calculated to determine the level of agreement. The sensitivity, specificity, and positive and negative predictive values of the software were calculated to determine the accuracy of the software regarding the diagnosis of caries lesion for the image date we selected. The correlation was considered significant if $p < 0.05$.

3. Results

A total of 604 proximal surfaces were evaluated of the 302 selected teeth, of which 37 (12.3%) were found on bitewing and 265 (87.7%) were found on periapical radiographs.

Human observers found in 119 (39.4%) cases of caries lesions and in case of 183 (60.6%) teeth no caries lesion was determined. The software evaluated 301 available teeth since one of the teeth was not recognized, of which caries was indicated in 103 cases (34.1%) and not in 198 (65.6%) cases (Table 1). Human observers and the software both found caries in 90 (29.9%) cases, however, none of them found lesions in 169 (56.1%) cases. The software diagnosed caries in 13 (4.3%) cases when it was not stated by human observers, and in 29 (9.6%) cases, caries were not detected by the software when it was stated by human observers.

		Diagnocat	
		no caries	**caries**
humans	**no caries**	169 (56.1%)	13 (4.3%)
	caries	29 (9.6%)	90 (29.9%)

Table 1 Agreement of human observers and Diagnocat

Determining the agreement between human observers and the Diagnocat software, as an observer, the Fleiss kappa coefficient was $\kappa = 0.8$, hence the level of agreement is moderate (Table 2). In the study, Fleiss kappa coefficient values differed significantly from 0 ($p < 0.01$).

κ values	agreement
0-0.20	less than chance agreement
0.21-0.40	slight agreement

0.41-0.60	fair agreement
0.61-0.80	moderate agreement
0.81-1	substantial agreement
1	almost perfect agreement

Table 2. κ values and levels of agreement [29]

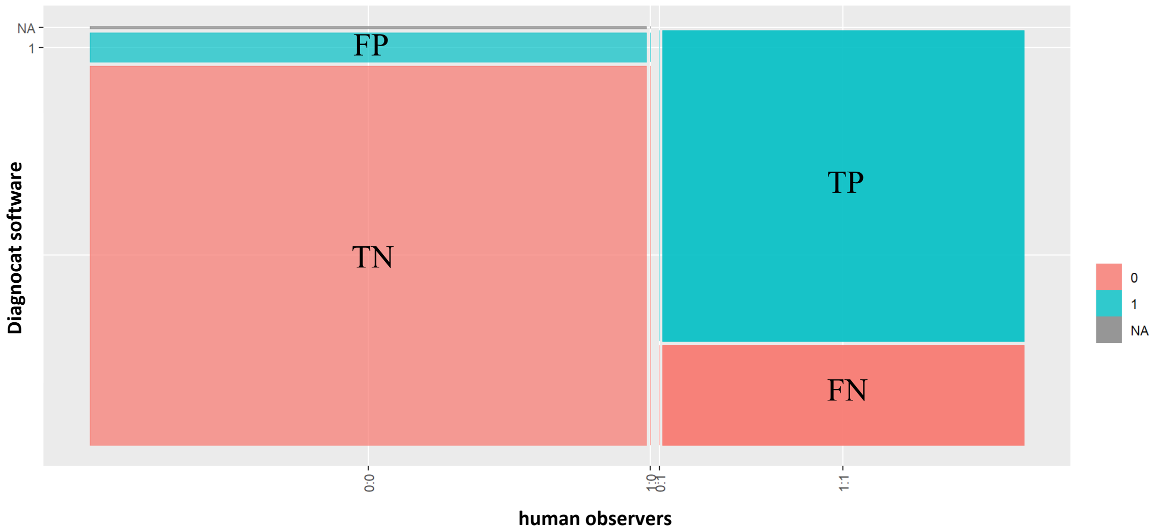

Figure 5. The X-axis contains the evaluation results of human observers: a value of '0' indicates the absence of caries and a value of '1' indicates the presence of a tooth. The Y-axis shows the evaluation of the Diagnocat software (FP - false positive, TN - true negative, TP - true positive, FN - false negative)

A mosaic plot was prepared (Figure 5) to illustrate the relationship of each observer's agreement. The X-axis contains the findings 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 illustrates the results of the evaluation of the Diagnocat software. TN (true negative) indicates the area when both the Diagnocat software and human observers assigned value '0', so none of the observers indicated that there were caries on the mesio- or distoapproximal surfaces of the assessed tooth. The TP (true positive) area shows the cases where caries was present according to both human observers and

Diagnocat software. The false positive (FP) result is obtained if caries were detected by the software but not by any human observer. The size of the FN (false negative) area may play a particularly important role in software development, illustrating cases where the software does not indicate a caries lesion but is present according to human observers.

The values of sensitivity, specificity, and positive and negative predictive values are shown in Table 3. By calculating the listed parameters the results of the dentomaxillofacial observer were determined as reality in the present study.

	values
sensitivity	**0.76**
specificity	**0.93**
positive predictive value	**0.87**
negative predictive value	**0.85**

Table 3. Sensitivity, specificity, and positive and negative predictive values of the software based on the selected radiographs

4. Discussion

To increase the accuracy and speed of radiologic evaluation of caries, image analysis software using AI has been developed [8] as the Diagnocat software which was used in our study. The main objective of our present scientific work was to determine the reliability of the Diagnocat software in the diagnosis of caries.

Several studies in the scientific literature examine the accuracy of caries diagnostics using AI [5, 7-11, 13, 30]. Devito et al. [5] took bitewing radiographs of extracted teeth, and then these 160 images were evaluated by observers and AI-based software, which achieved better results in detecting proximal caries than human observers. Schwendicke et al. [9] used CNNs for caries diagnosis on near-infrared light transillumination (NIR-LT) images. 226 extracted teeth were examined with the DIAGNOcam® system, which works with near-infrared light, and then the obtained images were uploaded to Resnet18 and Resnext50 CNN. The software was compared with two experienced dentists, and the Fleiss kappa value for the interobserver agreement was 0.72, which is in line with our findings since in our present research, this value was 0.8. With a caries prevalence of 41%, their sensitivity value was 0.59 and their specificity value was 0.76. In contrast, in our study, the sensitivity (0.76) and the specificity (0.93) values were significantly higher obtained by us.

Kühnisch et al. [10] also used a CNN for caries diagnosis on intraoral radiographs. The examined images were previously evaluated by an observer with more than 20 years of experience. CNN was able to recognize incipient caries and caries with cavity formation with at least 90% agreement. Lee et al. [11] used 2400 periapical radiographs to train the GoogLeNet Inception v3 CNN, then processed 600 periapical radiographs with the system and examined the software efficacy in detecting caries lesions. In the case of evaluation of premolar and molar teeth, the sensitivity was 0.81 and the specificity was 0.83. The results are also comparable with our findings, and although the sensitivity value is higher compared to the value we obtained (0.76) during the caries detection, its specificity is lower compared to the value of 0.93 we determined. In their study, Cantu et al. [8] examined 141 bitewing radiographs, during which a U-Net CNN evaluated the radiographs, then after the obtained results were compared with the observed results of 7 independent experienced dentists. The software determined caries lesions with a higher sensitivity value of 75% than the human observers, who were diagnosed with only 36% sensitivity, but at the same time, the specificity values of the examiners and the software were close to each other. Compared to our study, the Diagnocat program we use had a higher specificity and a higher sensitivity when determining caries lesions, i.e. its high specificity value can support the evaluating dentist in caries diagnostics if the Diagnocat software indicates caries. Schwendicke et al. [7] examined the cost-effectiveness of AI in the diagnosis of proximal caries based on a previous study published by Cantu et al. [8] and it was found that the software using AI is more cost-effective in caries diagnostics and also more accurate than without.

Lee et al. [13] investigated the role of CNNs in the detection of early caries lesions, and 50 bitewing radiographs were uploaded to the U-Net CNN. The study showed that the sensitivity values of the three dentist observers increased significantly by using the software: from 85.34% to 92.12%, from 85.86% to 93.72%, and from 69.11% to 79.06. Similar to our study, Ezhov et al. [12]applied the Diagnocat CNN software to assess CBCT images. Two groups of observers were examined: one of the groups performed a diagnosis supported by the Diagnocat, and the other group was unaided by the CNN. The sensitivity for aided and unaided groups for caries detection were 0.67 and 0.66, respectively, which are lower than our results.

Similar to the limited number of publications available in the scientific literature, our present study examined the possible development of caries diagnostics using AI on intraoral radiographs. The sensitivity (0.76) and specificity values (0.93) of the Diagnocat software for caries detection proved to be comparable to the available literature data. Based on our present study, the reliability of the Diagnocat software does not yet reach the level where it can be used independently to diagnose caries, but at the same time, it can highly support dentists during the evaluation of intraoral radiographs, since there is a moderate, close to a substantial agreement between the observers and the CNN based on the determined Fleiss kappa value. In the future, we plan to expand our research by including additional test data and performing an intraobserver reliability test. We must strive to contribute to the development of AI-based software and ensure the highest and most efficient patient care available.

5. Conclusions

Based on the agreement of the human and CNN observers, as well as the sensitivity and specificity values which are in line with the findings of the actual scientific literature, it can be concluded that the Diagnocat CNN may greatly help in evaluating dentist's work in the diagnosis of caries.

6. Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

7. References

1. 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
2. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, Jawm 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. https://doi.org/10.1016/10.1016/j.media.2017.07.005
3. R. Takahashi, Y. Kajikawa, Computer-aided diagnosis: A survey with bibliometric analysis, Int. J. Med. Inform. 101 (2017) 58-67. https://doi.org/10.1016/j.ijmedinf.2017.02.004
4. 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
5. 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
6. J.-J. Hwang, Y.-H. Jung, B.-H. Cho, M.-S. Heo, An overview of deep learning in the field of dentistry, Imaging Sci. Dent. 49 (2019) 1-7. https://doi.org/10.5624/isd.2019.49.1.1
7. 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
8. AG. Cantu, S. Gehrung, J. Krois, A. Chaurasia, JG. Rossi, R. Gaudin, K. Elhennawy, Schwendicke F., 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
9. 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
10. J. Kühnisch, Meyer, O., Hesenius, M., Hickel, R., Gruhn, V., Caries Detection on Intraoral Images Using Artificial Intelligence, J. Dent. Res. 101 (2021) 158-165. https://doi.org/10.1177/00220345211032524
11. 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.
12. 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
13. 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
14. 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
15. 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. (2021). https://doi.org/10.1007/s11282-021-00577-9
16. S. Kurt Bayrakdar, K. Orhan, I.S. Bayrakdar, E. Bilgir, M. Ezhov, M. Gusarev, Shumilov E., 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

17. 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
18. M. Kim, Yun J., Cho Y., Shin K., Jang R., Bae H.-J., Kim N., Deep Learning in Medical Imaging, Neurospine 16 (2019) 657–668. https://doi.org/10.14245/ns.1938396.198
19. J. Ker, Wang L., Rao J., Lim T., 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. RH. Selwitz, AI. Ismail, NB. Pitts, Dental caries, Lancet 369 (2007) 51-59. https://doi.org/10.1016/S0140-6736(07)60031-2
22. J.E. Frencken, Sharma P., Stenhouse L., Green D., Laverty D., Dietrich T., Global epidemiology of dental caries and severe periodontitis – a comprehensive review, J. Clin. Periodontol. 44 Suppl 18 (2017) S94–S105. https://doi.org/10.1111/jcpe.12677
23. J.R. Keenan, Keenan A.V., Accuracy of dental radiographs for caries detection, Evid. Based Dent. 17 (2016) 43–43. https://doi.org/10.1038/sj.ebd.6401166
24. K. Kamburoğlu, Kolsuz E., Murat S., Yüksel S., Özen T., 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
25. Wenzel A., Bitewing and digital bitewing radiography for detection of caries lesions, J Dent Res 83 (2004) 72-75. https://doi.org/10.1177/154405910408301s14
26. 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) 458-63. https://doi.org/10.1259/dmfr/84698143
27. H. Mohammad-Rahimi, SR. Motamedian, MH. Rohban, J. Krois, SE. 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
28. 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
29. A. J. Viera, J. M. Garrett, Understanding interobserver agreement: the kappa statistic, Fam Med 37 (2005) 360-3
30. 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

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.