Garcia-Canas, A., Bonfanti-Gris, M., Paraiso-Medina, S., Martinez-Rus, F., & Pradies, G. (2022). Diagnosis of interproximal caries lesions in bitewing radiographs using a deep convolutional neural network-based software. *Caries research*, *56*(5-6), 503-511.
**DOI:** 10.1159/000527491

# Diagnosis of Interproximal Caries Lesions in Bitewing Radiographs Using a Deep Convolutional Neural Network-Based Software

Angel Garcia-Canas[a], Monica Bonfanti-Gris[a], Sergio Paraiso-Medina[b], Francisco Martinez-Rus[a], Guillermo Pradies[a]

a)Department of Conservative and Prosthetic Dentistry, Faculty of Dentistry, Complutense University of Madrid,Madrid, Spain;

b) Department of Informatic Systems and Languages. Faculty of Software Engineering, Polytechnic University of Madrid, Madrid, Spain

## Abstract

The aim of this study was to evaluate the diagnostic reliability of a web-based artificial intelligence program for the detection of interproximal caries in bitewing radiographs. Three hundred bitewing radiographs of patients were subjected to the evaluation of a convolutional neural network. **First, the images were visually evaluated by a previously trained and calibrated operator with radiodiagnosis experience. After, ground truth was established and was clinically validated. For enamel caries, clinical assessment included a combination of clinical-visual and radiography evaluations. For dentin caries, clinical validation was performed by instrumentally accessing to the cavity. Secondly, the images were uploaded and analyzed by the web-based software**. Four different models were established to analyze its evaluations according to the confidence threshold (0-100%) offered by the program: model 1 (values > 0% were considered positive and values of 0% were considered negative), model 2 (values ≥ 25% were considered positive and values < 25% were considered negative), model 3 (values ≥ 50% were considered positive and values < 50% were considered negative), and model 4 (values ≥ 75% were considered positive and values < 75% were considered negative). The accuracy rate (A), sensitivity (S), specificity (E), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and areas under receiver operating characteristic curves (AUC) were calculated for the four models of agreement with the software. Models showed the following results respectively: A= 70.8%, 82%, 85.6%, 86.1%; S= 87%, 69.8%, 57%, 41.6%; E= 66.3%, 85.4%, 93.7%, 98.5%; PPV=42%, 57.2%, 71.6%, 88.6%; NPV=94.8%, 91%, 88.6%, 85.8%; PLR= 2.58, 4.78, 9.05, 27.73; NLR= 0.2, 0.35, 0.46, 0.59; AUC=0.767, 0.777, 0.753, 0.701. **Findings in the present study suggest that the artificial intelligence web-based software provides a**

**good diagnostic reliability on the detection of dental caries**. Our study highlighted model 2 for showing the best results to differentiate between healthy teeth and decayed teeth.

## Introduction

Dental caries is defined as a non-communicable multifactorial dynamic disease, biofilm-mediated and diet-modulated, which results in net mineral loss of dental hard tissues. It is determined by biological, behavioral, psychosocial and environmental factors and is considered the most prevalent condition worldwide, since it affects practically the entire population at some point in their lifes [Selwitz et al., 2007; Lee et al., 2018].

The visual-tactile method is the most frequently used technique for the diagnosis of dental caries. However, it must be complemented with the performance of radiographic tests. The radiological test of choice for the diagnosis of this condition is the bite-wing radiograph, which has higher sensitivity than panoramic radiographs, especially for the detection of interproximal lesions in posterior teeth (molars and premolars) - where clinical detection either visually or through a dental explorer is specially challenging [Selwitz et al., 2007; Schwendicke et al., 2015; **Srivastava et al. 2017**, Prados-Privado et al., 2020]. However, although it is considered to be a highly reliable method for the detection of dental caries, final diagnosis can be influenced by the operators' subjectivity, as well as by various factors including both monitor and viewing conditions [Lee et al., 2018].

During the last decades, great technological advances have been developed in all areas of our lives. This technological revolution has also affected the healthcare field, where computer-aided diagnosis has significantly evolved due to the progress of big data and artificial intelligence [Tuzoff et al, 2019]. Artificial intelligence (AI) is considered a field of engineering that implements concepts and solutions to solve complex situations [Hamet and Trembay, 2017]. Machine learning and deep learning are two subfields of artificial intelligence. **These participate** in the study and construction of algorithms that are capable of learning from information and making predictions [Mupparapu et al, 2018]. Machine learning allows algorithms to identify complex patterns among a large amount of data establishing their own rules for detecting similar patterns in new data sets. Deep learning, instead, is a specialized type of machine learning that is based on a layered learning model that processes information. The basis of deep learning are the neural networks, a system that mimics the functioning of the human brain and its neurons [Mupparapu et al, 2018; Prados-Privado et al., 2020]. Big data refers to the storage and processing of massive amounts of structured, semi-structured and unstructured data with great potential to be extracted and organized in a way that provides valuable information by using artificial intelligence.

The development of this technology has shown certain difficulties in dentistry due to the learning curve involved in the process, as well as the limited availability of large data sets of medical images. Despite this, in recent years there have been wide-raging advances in the field that have led researchers and health professionals to recognize the important role of AI. This reinforces the idea that it could be definitely used for the diagnosis of dental caries, among other oral pathologies [Mazurowski et al., 2019]. **Several studies have already been published concerning the diagnostic reliability that these new technologies can provide to dental practitioners. Although the main objective has been to radiographically segment and detect carious lesions [Bayraktar et al., 2021; Leo et al., 2020; Mertens, et al., 2021; Tripathi et al., 2019; Srivastava et al., 2017], further steps have been taken when referring to their classification. Nowadays, the clinical relevancy of these neural networks is focused on how they can differentiate different types of caries according to their position (both enamel and dentin). To do so, neural networks have been trained not only to segment anatomical structures, but also dental anatomical components such as enamel, dentin and the pulp. [Lee et al., 2021; Leo et al., 2021]**

Recently, **a web-based artificial intelligence software** that generates an automatic diagnosis of radiological images (orthopantomography and intraoral radiographs) by means of machine learning technologies (Denti.Ai; Denti.Ai Technology Inc, Toronto, ON, Canada) was released and made available on the internet. **This convolutional neural network (CNN) automatically detects and classifies dental structures, treatments and certain pathologies such as dental caries or periodontitis. To do so, it synergically works with three different AI-based modules: Model Faster R-CNN is used to identify dental structures, architecture VGG-16 is responsible for tooth classification and numbering and treatment/pathology detection is carried out by a gray-scale codification system, being indicated by bounding boxes, similar to what Srivastava et al., 2017 reported in their work. Despite this research paper focuses on the reliability evaluation of DentiAi's caries detection capacity, other neural networks have been developed and should be taken into consideration, such as Google Inception v3 CNN [Lee et al., 2018] or U-Net [Garcia Cantu et al., 2020]. Also, it is important to state that, although nowadays developed neural networks are based on a forward-propagation working process, it has also been reported that back propagation neural networks could provide with good results for this matter [Tripathi et al., 2019]. Likewise, other kinds of neural networks have been described to have reached successful outcomes, such as the one developed by Leo et al., 2021 consisting in a hybrid neural network – being this a blend of Artificial Neural Networks and Deep Neural Networks [Leo et al. 2021].**

As it was previously stated, the early diagnosis of dental caries continues to be a challenge for dentists. Artificial intelligence can perform these tasks in a simple, fast and precise way, thus helping the dentist in their daily practice, reducing work time and improving clinical decision making [Prajapati et al., 2017]. Although at present artificial intelligence in dentistry has been used mainly for the detection and identification of anatomical structures and teeth, the scientific literature about its use in dental caries diagnosis is scarce [Lee et al., 2018]. Bearing this in mind, the aim of this research was to compare the efficacy of a deep convolutional neural network-based software against the gold standard on detecting proximal caries lesions in enamel and dentin. The null hypothesis was that no significant differences would be found in the diagnostic performance of a machine learning software versus the clinical validation.

## Material and methods

*Study design and patient selection*

This study was designed as an analytical, observational, and cross-sectional study. The Standards for Reporting Diagnostic Accuracy (STARD) statement and the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) were followed in planning, conducting, and reporting this investigation [Bossuyt et al., 2015; Mongan et al., 2020; Schwendicke et al., 2021]. The sample size was calculated using a standard formula for prevalence studies [Naing et al., 2006] as follows:

$$n = \frac{Z^2 P(1-P)}{d^2}$$

where $n$ is sample size, $Z$ is the value of standard normal deviation corresponding to the level of confidence (for a 95% confidence level, the value of Z is 1.96), $P$ is the expected prevalence, and $d$ is the absolute precision. Assuming that the mean prevalence of un-treated caries in permanent teeth in Spain is 19.46% [Bravo-Pérez et al.,2020], the calculated sample size was 241 to achieve a confidence level of 95% and an absolute precision of 5%. Based on this data, a final sample size of 300 bitewing radiographs was included.

Between July 2019 and July 2020, 150 patients aged between 16 and 85 years (81 women, 69 men, mean age: 43 years) who came for a check-up of their oral health status in a private dental clinic were recruited in the study. **A non-probability convenience sampling method was performed to recruit patients undergoing this study**. The inclusion and exclusion criteria are listed in **Table 1**. All patients were evaluated and treated by the same trained clinician (A.G). The intra-examiner Kappa coefficient value prior to the start of the inspections was considered acceptable since it was found within a range of 0.81 and 0.85 (p = 0.04).

*Clinical-visual examination*

The dentist carefully conducted clinical-visual examination of each tooth using standard conditions. The lesion severity was scored according to the International Caries Detection and Assessment System (ICDAS) [Ekstrand et al., 2018]. The criteria were (0) sound tooth, (1) first visual change in enamel, (2) distinct visual change in enamel, (3) localized enamel breakdown, (4) underlying dark shadow originating from dentin, (5) distinct cavity with visible dentin, and (6) extensive distinct cavity with visible dentin. Codes of 1, 2, and 3 were used for the diagnosis of enamel caries, and codes of 4, 5 and 6 were used for the diagnosis of dentin caries.

*Radiography examination*

A total of 300 digital bitewing radiographs of premolars and permanent molars -both upper and lower- were **taken by the same clinician during the patients' appointment and were posteriorly exported in .jpeg file format. All of them were acquired and processed following identical conditions**. The X-ray unit (CS 2100; Carestream, Rochester, NY, USA) was adjusted to the manufacturer parameters of 60 kV, 7 mA, and 0.125 s, with a focus-sensor distance of 20 cm. The images were obtained using a radiovisiography sensor (RVG 6100; Carestream) positioned with a bitewing holder (XCP DS Fit; Dentsply Sirona; Charlotte, NC, USA) and the manufacturer's software (CS Imaging Version 8; Carestream). The presence of proximal enamel or dentin caries was determined using the International Caries Classification and Management System (ICCMS) **categories** [Ismail et al., 2015]: (0) no radiolucency, (1) radiolucency in the outer 1/2 of the enamel, (2) radiolucency in the inner 1/2 of the enamel ± enamel-dentin junction, (3) radiolucency limited to the outer 1/3 of dentin, (4) radiolucency reaching the middle 1/3 of dentin, (5) radiolucency reaching the inner 1/3 of dentin, clinically cavitated, and (6) radiolucency into the pulp, clinically cavitated. For enamel caries, codes of 1 or 2 were used; for dentin caries, codes of 3, 4, and 5 were used in the classification. All radiographs were anonymized and stored in an encrypted digital folder.

*Radiography assessment using a deep convolutional neural network-based software*

The pre-trained convolutional neural network-based software **(Denti.Ai) [Tuzoff et al., 2019]** was used to analyze all radiographs. Denti.Ai is an automated online software based on artificial intelligence that is capable of generating an automatic diagnosis of radiological images using deep learning technologies **in matter of seconds**. This program is compatible with orthopantomographs and intraoral radiographs, both periapical and bitewing. To carry out the task of detecting and classifying dental structures found in the radiographs, this program uses different detection modules: first, the image is processed to establish the limits of each tooth present and then marks the structures found with bounding boxes (Faster R-CNN Module). The tooth numbering module (VGG-16 Architecture) then classifies each delimited tooth region according to the *Federation Dentaire International* (FDI) dental nomenclature. Regarding the diagnosis of pathologies (carious and periapical lesions) and the detection and classification of dental treatments (composite or amalgam fillings, root canal treatments, crowns and implants, etc.), the program is executed by a coding system based on a gray scale and offers a confidence threshold for each find (from 0 to 100%) [Tuzoff et al., 2019], **bearing in mind that no** graded severity is specified in any of the pathologies encountered in the images. **Figure 1** shows an overview of caries detection process used.

In our research, once the anonymized images were uploaded to the system, study variables were selected within the categories that the program offers to search for and analyze. As dental structures were automatically detected, it was only necessary to select the "caries/decay" category.

*Clinical validation*

For enamel caries lesions, the gold standard was established by combining the clinical-visual and digital bitewing radiography assessments (A.G-C and M.B-G). With a consensus about enamel caries, the cavity was not opened for ethical reasons [Schwendicke et al., 2019]. For dentin caries, the gold standard was stablished by opening the cavity and the application of caries detector dye (Kurakay Caries Detector; Kuraray Noritake Dental, Tokyo, Japan). Detected dentin caries lesions were opened with a conical diamond bur and hand excavator until the soft carious tissue was completely removed. After the cavity was opened, lesions were restored with a filling material appropriate for the cavity.

*Data analysis*

The results obtained from the machine learning software were recorded in a spreadsheet (Microsoft Excel; Microsoft Corp, Redmond, WA, USA), where the following information was included: side, tooth, location, presence or absence of caries, confidence threshold (0-100%) offered by the software, dental structure where caries was detected by software (0= No caries, 1= Enamel, 2= Dentin or 3= Both) and software's error type in incorrectly detected cases (0= No error, 1= Lack of homogeneity in grey levels due to uneven crown structure, 2= Mistaken detection of caries within dental structure, 3= Mistaken dental structure classification of the caries, 4= Lack of detection of dentin caries, 5= Dental caries detection associated with superposition of interproximal surfaces, 6= Software does not detect clinical validated caries, 7= Software detects caries but bounding box does not correctly delimit carious lesion, 8= Software detects caries in ceramic or plastic restoration and 9= Software detects dental pulp as dental caries).

Since the program offered the evaluation in a % bar representation, it was decided to establish four models according to statistical quartiles to analyze the results offered by the machine learning software. This was performed to test the software and determine whether a certain confidence threshold had to be established previous to the software analysis to obtain more reliable detection outcomes.

- Model number 1, where values greater than 0% were considered positive, and values equal to 0% were considered negative.

- Model number 2, where values greater than or equal to 25% were considered positive, and values less than 25% were considered negative.
- Model number 3, where values greater than or equal to 50% were considered positive, and values less than 50% were considered negative.
- Model number 4, where values greater than or equal to 75% were considered positive, and values less than 75% were considered negative.

The diagnostic performance of the deep convolutional neural network-based software in determining enamel and dentin caries according to the gold standard was expressed in terms of accuracy rate (A), sensitivity (S), specificity (E), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR), receiver operating characteristic (ROC) curves and areas under the curves (AUC). Accordingly, the following values were calculated: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Data was analyzed using the software SPSS 26.0 for Mac (IBM Corporation, Armonk, NY, USA).

## Results

**Table 2** shows the distribution of the analyzed teeth in this study. Out of a total of 1323 teeth, the first molar was the most studied tooth in the sample (42%), followed by the second premolar (32%), second molar (20%) and finally the first premolar (6%). The reference test showed 579 positive evaluations (caries) and 2067 negative evaluations (healthy), resulting in a sample prevalence of 21.9%. **Software's dental caries detection on enamel, dentin or enamel + dentin was found to be 33.48%, 5.03% and 10.05%, respectively, being 51.44% of cases associated with no caries detection surfaces. Values for True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) were collected for each study models: Model 1) TP=504, FP=697, TN=1370, FN=75; Model 2) TP=404, FP=302, TN=1765, FN=175; Model 3) TP=330, FP=131, TN=1936, FN=249; Model 4) TP= 241, FP=31, TN=2036, FN=338.** **Table 3** shows the values of overall accuracy, specificity, positive predictive value, positive likelihood ratio and negative likelihood ratio, which increase as the model raises the threshold confidence; and sensitivity and negative predictive value, which decrease as the model raises the threshold confidence offered by the four models established for the AI web-based software. A comparison of the ROC curves offered by the four models established for the software can be seen in **Figure 2**. The area under the curve for model 1 was 0.767 (95% CI 0.746-0.787), for model 2 was 0.777 (95% CI 0.753-0.800), for model 3 was 0.753 (95% CI 0.727-0.779), and for model 4 was 0.701 (95% CI 0.673-0.728). The Pearson's chi-squared test indicated significant differences among the four models of the machine learning software versus the gold standard ($p < 0.001$).

**During the software's testing phase, operators detected different caries detection mistakes (0-9),**

**being the results obtained for each classification category: 0=68.52%, 1=4.12%, 2=19.39%, 3=0.08%, 4=0.15%, 5=4.69%, 6=2.31%, 7=0.04%, 8=0.57% and 9=0.15%.**

## Discussion/Conclusion

This study evaluated the diagnostic performance of a deep convolutional neural network-based software for detecting proximal caries lesions in enamel and dentin. The null hypothesis was rejected because statistical differences were obtained between the results obtained from the machine learning software versus the clinical validation, being model 2 the best alternative when compared with the gold standard.

**The caries diagnosis through bitewing radiographs presents some limitations due to the lack of agreement among professionals, especially in those inexperienced [Srivastava et al., 2017; Cantu et al., 2020]. Also - as mentioned in Lee et al. 2021 - these radiological images are especially vulnerable to false-positive and false-negative diagnoses due to dental structure overlapping**. The introduction of artificial intelligence programs to assist in caries detection and diagnosis could provide the dental professional with a more reliable and accurate diagnosis [Lee et al., 2018; Geetha et al., 2020; Prados-Privado et al., 2020]. **Nevertheless, although these new technologies might undoubtedly help clinicians, they can also elevate the number of cases where detection of carious lesions do not clinically exist (FP), increasing invasive therapy decisions where they are not needed [Mertens et al., 2021]. Some articles presented a reliable way to accurately detect caries regions from the input bitewing radiographs, consisting of the development of a two-step detection and refinement process where two different models can be utilized to detect 1) dental caries segmentations and 2) structure segmentations [Lee et al., 2021]. Also, a back propagation and Local Binary Pattern (LBP) CNN based on genetic algorithms was developed by Tripathi et al., ensuring successful results due to the selection of the better neural population – the genetic algorithm selects the NN which produces least amount of errors from the LBP- and the establishment of a learning rate [Tripathi et al., 2019].**

Currently, several scientific studies related to the use of convolutional neural networks in the diagnosis of caries in intraoral radiographs have been published [Yu et al., 2006; Li at al., 2007; Albahbah et al., 2016; Srivastava et al., 2017; Lee et al., 2018; Cantu et al., 2020; Geetha et al., 2020]. All of them showed positive results in favor of artificial intelligence, but the web-based program tested in this study (Denti.Ai) was not included in any of them. The results of our study show similar results compared to related works, which share a similar methodology [Yu et al., 2006; Li at al., 2007; Albahbah et al., 2016; Srivastava et al., 2017; Lee et al., 2018; Cantu et al., 2020; Geetha et al., 2020]. The highest overall accuracy is presented by model 4 with 86.1%, followed by model 3 (85.6%), model

2 (82%) and model 1 (70.8%). Similarly happened with the specificity values, model 4 had the highest capacity to detect healthy teeth with 98.5%, followed by model 3 (93.7%), model 2 (85.4%) and model 1 (66.3%). As the confidence threshold increased, the program improved its ability to detect true negatives and, therefore, increased the percentage of total correct answers. However, the opposite occurred for sensitivity values. The program worsened its ability to detect decayed teeth: model 1 showed the best results (87%), followed by model 2 (69.8%), model 3 (57%) and model 4 (41.6%). Both testing parameters - sensitivity and specificity - are inherent properties of the test regardless of the disease prevalence in the sample studied and offer us an approximation of the program's capacity to identify caries. However, both values are given by inverse probability weighting, which limits the possibility of making intuitive clinical interpolations for each patient [Gansky 2003]. **Nevertheless, according to Bayraktar et al., 2021, AI-based software's can successfully detect sound dental surfaces, demonstrating they have a higher rate of specificity than sensitivity in bitewing radiographs [Bayraktar et al., 2021].**

Regarding the probability of presenting dental caries with a positive result, as the value of the threshold confidence that determined the presence of cavities increased, the PPV increased. This may be due to the fact that the higher the percentage offered by the web-based software, the greater the probability of success, and, as a result, model 4 offered a percentage of 88.6% and model 1 of 42%. The difference between models is less for the NPV, being 94.8% for model 1 and 85.8% for model 4. One aspect to be highlighted is that these values, despite they are extremely useful when making clinical decisions and transmitting information to patients, are highly influenced by the condition prevalence in the sample studied. When the prevalence is low, NPV increases and PPV decreases, being the opposite when the prevalence is high. For this reason, PPV and NPV should not be used as rates to compare two different diagnostic methods, nor to extrapolate results obtained in other studies.

For this reason, it is necessary to determine alternative statistical tests that do not depend on the prevalence of the disease and offer a clinical utility that allows to reflect the diagnostic reliability of the test, such as the likelihood ratio. The likelihood ratio is the likelihood that a given test result would be expected in a patient with the target characteristic compared to the likelihood that same result would be expected in a patient without the target characteristic. According to the ranges of likelihood ratio values [Hayden and Brown, 1999], model 4 presented a highly relevant PLR and a bad NLR, model 3 a good PLR and a regular NLR, model 2 a regular PLR and NLR, and model 1 a regular PLR and a good NLR.

Another parameter independent of prevalence and with high clinical utility is the ROC curve, obtained from the sensitivity and specificity values of the test. The area under this curve is the best indicator of the predictive ability of the test, allowing to establish comparisons between different diagnostic tests.

In this research, model 2 offered the highest value of the area under the curve (0.777), followed by model 1 (0.767), model 3 (0.753), and finally model 4 (0.701).

**In our study, the AI web-based software did not classify dental caries according to their radiological severity. Despite this, when performing a manual-operator categorization of the structural-affected lesions, small interproximal enamel caries were the most repeatedly detected ones. Recent published articles confirmed that these new CNNs tools can increase the operator's performance when detecting early-stage carious lesions (enamel), not being the case in advanced lesions where dentin is also involved [Lee et al., 2021; Mertens et al., 2021].**

**Also, operators in this study performed a diagnostic error categorization due to the acknowledgement of repetitive mistakes conducted by the AI-based software. As mentioned, the most popular ones were related to presence of differences in radiographic density due to tooth fracture or other enamel defects (4.12%), overlapping of interproximal surfaces (4.69%) and presence of black spaces between teeth (19.39%). This was also noted by Lee et al. 2021, who reported that these limitations are due to lack of training data for cases of superposition with other unexpected features present in the image [Lee et al., 2021].**

Although the null hypothesis was rejected, we consider that the web-based software analyzed can be used as a reliable tool in the radiological diagnosis of interproximal caries. Model 2 deserves special consideration, since it offers the best results to differentiate between healthy teeth and decayed teeth. However, it should not be ignored the importance of analyzing the data set from a global point of view, in order to take advantage of the benefits offered by each model. Similarly, it must be beared in mind that this technology does not establish direct causal relationships, so the information provided must be interpreted with caution by users [Chen et al., 2020]. It is important to highlight that these programs must be developed and implemented in collaboration with experts from the different areas of health and computer engineering to guarantee the least possible number of errors [Mupparapu et al, 2018]. Also, it should be noted that the final responsibility for diagnosis and therapeutic intervention lies with the professional and never with technologies [Chen et al., 2020].

Concerning the methodology of dental caries scientific studies, it is important to establish an initial reference test to compare the results offered by the artificial intelligence program. The method considered as gold standard for the diagnosis of caries lesions is the histopathological examination [Downer, 1975]. Another reference test can be the one established in this work, the diagnosis of an experienced dentist based on the clinical examination according to the ICDAS criteria [Ekstrand et al., 2018], radiographic tests according to the ICCMS criteria [Ismail et al., 2015] and the subsequent cavity opening and obturation.  At the same time, the level of intra- and inter-observer agreement of the expert dentists in charge of evaluating radiographic images should be taken into consideration. The scientific literature reports that Kappa values between 0.81 and 1.00 indicate excellent concordance [Rigby, 2000].

As to the data set that is used as training of the artificial intelligence program, they are based on evaluations from human professionals, so the quality of the input information is critical for the correct development of technology [Prados-Privado et al., 2020]. **As the software used in this research was a pre-trained CNN launched in 2017, no previous training phase was carried through. Taking this into account, it is important to emphasize that overfitting was not analyzed in the present study, mainly because the detection module was carried out by means of bounding boxes, not anatomical delimitation of the structures. Also, it has to be noted that artificial intelligence applications are also being studied in areas such as breast cancer diagnosis** [McKinney et al., 2020] or skin cancer diagnosis [Haenssle at al., 2018], have showed better results than human professionals.

**There were some limitations in the present study. It has to be pointed out that a single radiodiagnosis device was used, so the diagnostic capacity of both the software and the human operators may have been influenced by the characteristics of the images recruited. Thus, generalizability might be compromised. [Yu et al., 2006]. Another limitation to bear in mind is the expert's clinical reference test, which was conducted only by one clinician [Mohammad-Rahimi et al. 2022]. Also, carious lesions were not differentiated according to their depth by the AI- based tool,** which restricts obtaining relevant data about the reliability of the artificial intelligence program and the stage of the detected disease. **Furthermore, the number of carious lesions included in this study were not homogeneously distributed in terms of severity nor localization. Thus, this may increase over-classification of the majority group due to class imbalance**. **A further study is necessary to train** the system with a much larger dataset, since the higher the number of subjects used to train the neural network, the higher the accuracy it will have for testing unknown objects. Deep learning techniques can be applied to numerous other tasks than carious detection. Several exciting prospects are currently being studied, for example, using machine learning technologies to detect pathologies, such as periodontitis and periapical cysts. Admittedly, to achieve expert-level results for such a task, these new systems would require larger datasets for training than used in this project.

The purpose of artificial intelligence is not to exclude the input of the clinician from the process of radiographic dental caries diagnosis, but to highlight suspicious areas, increasing clinical efficiency by focusing human effort towards specific sites that need further evaluation. The deep convolutional neural network-based software tested in this study can be used as a reliable tool in the radiological diagnosis of interproximal caries, showing model 2 the best results to differentiate between healthy teeth and decayed teeth. **Also, our results provide a general caries severity detection reliability and** an error analysis that were not included in most of the published articles concerning this topic – only studies concerning the CNNs capacity to detect and classify dental caries on behalf of their severity based on the affected dental tissue have been published. This way, our findings may help other researchers to develop similar studies where these problematics might be dealt with and, eventually, overcome.

## Statement of Ethics

This study protocol was reviewed and approved by the Ethics Committee of the Complutense University of Madrid, approval number 21/303-E. All patients provided written informed consent to participate in the study according to the Declaration of Helsinki.

## Conflict of Interest Statement

The authors have no conflicts of interest to declare.


## Funding Sources

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


## Author Contributions

G.P. and S.P.-M conceived and designed the study. A.G.-C. and M.B.-G. performed the study. G.P., S.P.-M and F.M.-R. analyzed and interpretated the data. A.G.-C., M.B.-G, and F.M.-R wrote the paper. All of the authors approved the final version of this paper and are accountable for all aspects of this work.

## Data Availability Statement

The data that support the findings of this study are not publicly available due to their containing information that could compromise the privacy of research participants, but are available from the corresponding author [F.M.-R.] on reasonable request.

## References

Albahbah AA, El-Bakry HM, Abd-Elgahany S. Detection of caries in panoramic dental X-ray images using back-propagation neural network. Int J Electron Commun Comput Eng. 2016;7(5):250-6.

Baykratar Y, Ayan E. Diagnosis of interproximal caries lesions with deeo convolutiooonal neural nnetwork inn diigital bitewing radioigraphs. Clin Oral Investig. 2021 June.

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HC, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015 Oct 28;351:h5527.

Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. J Dent. 2020 Sep;100:103425.

Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. Quintessence Int. 2020;51(3):248-57.

Downer MC. Concurrent validity of an epidemiological diagnostic system for caries with the histological appearance of extracted teeth as validating criterion. Caries Res. 1975;9(3):231-46.

Ekstrand KR, Gimenez T, Ferreira FR, Mendes FM, Braga MM. The International Caries Detection and Assessment System - ICDAS: A Systematic Review. Caries Res. 2018;52(5):406-419.

Gansky SA. Dental data mining: potential pitfalls and practical issues. Adv Dent Res. 2003 Dec;17:109-14.

Geetha V, Aprameya KS, Hinduja DM. Dental caries diagnosis in digital radiographs using back-propagation neural network. Health Inf Sci Syst. 2020 Jan 3;8(1):8.

Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018 Aug 1;29(8):1836-42.

Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017 Apr;69S:S36-S40.

Hayden SR, Brown MD. Likelihood ratio: A powerful tool for incorporating the results of a diagnostic test into clinical decisionmaking. Ann Emerg Med. 1999 May;33(5):575-80.

Diagnosis of caries using artificial intelligence.

Ismail AI, Pitts NB, Tellez M, Authors of International Caries Classification and Management System (ICCMS). The International Caries Classification and Management System (ICCMS™) An Example of a Caries Management Pathway. BMC Oral Health. 2015;15 Suppl 1(Suppl 1):S9.

Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 2018 Oct;77:106-111.

Lee S, Oh S-I, Jo J, Kang S, Shin Y, Park J-W. Deep learning for early dental caries detection in bitewing radioigraphs. Sci Rep. 2021; 11:16807.

Li W, Kuang W, Li Y, Li Y, Ye W. Clinical X-Ray Image Based Tooth Decay Diagnosis using SVM. In: 6th International Conference on Machine Learning and Cybernetics. Hong Kong; 2007; p. 1616-19.

Mazurowski MA, Buda M, Saha A, Bashir MR. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. J Magn Reson Imaging. 2019 Apr;49(4):939-954.

McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020 Jan;577(7788):89-94.

Mertens S, Krois J, Garcia Cantu A, Asiwala LT. Artificial Intelligence for caries detectoin: Rodomiized Trial. J Dent. 2021 Aug; 115:1-6.

Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiol Artif Intell. 2020 Mar 25;2(2):e200029.

Mohammad-Rahimi H, Motamedian SR, Rohban MH, Krois J, Uribe S, Mahmoudinia E, et al. Deep learning foor caries detection: A systematic Review. J Dent. 2022 Jan;122:1-16.

Mupparapu M, Wu CW, Chen YC. Artificial intelligence, machine learning, neural networks, and deep learning: Futuristic concepts for new dental diagnosis. Quintessence Int. 2018;49(9):687-8.

Prados-Privado M, García Villalón J, Martínez-Martínez CH, Ivorra C, Prados-Frutos JC. Dental Caries Diagnosis and Detection Using Neural Networks: A Systematic Review. J Clin Med. 2020 Nov 6;9(11):3579.

Prajapati SA, Nagaraj R, Mitra S. Classification of dental diseases using CNN and transfer learning. In: 5th Int. Symp. Comput. Bus. Intell. ISCBI 2017. Institute of Electrical and Electronics Engineers Inc. New York; 2017; p. 70-4.

Diagnosis of caries using artificial intelligence.

Rigby AS. Statistical methods in epidemiology. v. Towards an understanding of the kappa coefficient. Disabil Rehabil. 2000 May 20;22(8):339-44.

Schwendicke F, Tzschoppe M, Paris S. Radiographic caries detection: A systematic review and meta-analysis. J Dent. 2015 Aug;43(8):924-33.

Schwendicke F, Splieth C, Breschi L, Banerjee A, Fontana M, Paris S, Burrow MF, et al. When to intervene in the caries process? An expert Delphi consensus statement. Clin Oral Investig. 2019 Oct;23(10):3691-3703.

Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, Uribe S, Krois J; IADR e-oral health network and the ITU WHO focus group AI for Health. Artificial intelligence in dental research: Checklist for authors, reviewers, readers. J Dent. 2021 Apr;107:103610.

Selwitz RH, Ismail AI, Pitts NB. Dental caries. Lancet. 2007 Jan;369(9555):51-9.

Srivastava MM, Kumar P, Pradhan L, Varadarajan S. Detection of tooth caries in bitewing radiographs using deep learning, ArXiv. DOI: 10.48550/arXiv.1711.07312.

Tuzoff DV, Tuzova LN, Bornstein MM, Krasnov AS, Kharchenko MA, Nikolenko SI, Sveshnikov MM, Bednenko GB. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofac Radiol. 2019 May;48(4):20180051.

Yu Y, Li Y, Li Y, Wang J, Lin F, Ye W, Tooth Decay Diagnosis using Back Propagation Neural Network. 5th International Conference on Machine Learning and Cybernetics. Dalian;  2006, p. 3956-59.

## Figure Legends

Fig. 1. Flowchart of the detection of dental caries in the deep convolutional neural network-based software: First, the radiological images were uploaded to the platform, which were immediately anonymized. The identification number of the patient, the gender, the date of the image and the date of birth of the patient were selected. Once all the fields were completed, the save option was clicked. Although the program offers the possibility of rotating the image, modifying brightness and contrast, and applying a negative or sharpening filter, none of these parameters were modified for the study. Once the definitive image was obtained, the analyze option was clicked and in a total of four seconds the program analyzed the radiological image, identifying the teeth present, calculating the distance from the alveolar bone to the to the amelocemental junction, detecting previous treatments and pathologies by offering a percentage that indicated the confidence threshold, and marking the findings in the image through green and red dashed lines. Finally, the program produced a written report that contained the data resulting from the process.

Fig. 2. Comparison of the ROC curves of models 1, 2, 3 and 4 of the deep convolutional neural network-based software.

## Table Legends

Table 1. Inclusion and exclusion criteria employed in this study.

Table 2. Distribution of the analyzed teeth.

Table 3. Values of accuracy (A), sensitivity (S), specificity (E), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR) showed by the four models of machine learning software.