

1 **The use of artificial intelligence in the diagnosis of carious lesions: Systematic review and meta-analysis**

2
3 Vanessa Gallego Arias Pecorari^{1,2}, Laís Renata Almeida Cezário², Caio Vieira de Barros Arato^{2*}, Tainá de Lima
4 Costa¹, Karine Laura Cortellazzi², Roberto Fiório Pecorari², José Erasmo Silva³

5
6 ¹Paulista University, UNIP, São Paulo, Brazil

7 ²School of Dentistry of Piracicaba, Universidade Estadual de Campinas, UNICAMP, São Paulo, Brazil

8 ³University of São Paulo, USP, São Paulo, Brazil

9
10 *Corresponding author: caio.arato@hotmail.com

11
12 [¶]These authors contributed equally to this work.

13 [&]These authors also contributed equally to this work.

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 Abstract

54 **Background:** The use of Artificial Intelligence (AI) has many applications in the healthcare field. Dental caries
55 is a disease with a prevalence rate of over 50% in Brazil. The diagnosis of caries is usually based on a clinical
56 examination and supplementary tests such as X-rays. The accuracy of a diagnostic test is evaluated by its
57 sensitivity, specificity, and accuracy. Various algorithms and neural network configurations are being used
58 for caries diagnosis. **Objective:** This systematic review evaluated the sensitivity, specificity, and accuracy of
59 using deep machine learning through a convolutional neural network in diagnosing dental caries. **Methods:**
60 This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic
61 review and Meta-Analyses (PRISMA) 2020 guidelines and registered with Prospero (ID CRD42024411477).
62 We used the PubMed, MEDLINE, and LILACS databases and MeSH and DECs descriptors in the search. **Results:**
63 After analyzing the eligibility of the articles, we selected 33 for full-text reading and included 13 in the meta-
64 analysis. We used the sensitivity, specificity, accuracy data, and the number of positive and negative tests to
65 generate a 2x2 table with TP, FP, FN, TN rates, and accuracy. We evaluated the heterogeneity of the SROC
66 curve using the Zhou & Dendurkuri I 2 approach. The results showed that the sensitivity and specificity of the
67 machine learning for detecting dental caries were 0.79 and 0.87, respectively, and the AUC of the SROC curve
68 was 0.885. **Conclusion:** The literature presented a variety of convolutional neural networks [CNN]
69 architecture, image acquisition methods, and training volumes, which could lead to heterogeneity. However,
70 the accuracy of using artificial intelligence for caries diagnosis was high, making it an essential tool for
71 dentistry.

72 **Keywords:** Deep learning, Systematic review, caries diagnosis, artificial intelligence.

73

74 Introduction

75 Dental caries is one of the most prevalent diseases (about 50%) in children worldwide and is
76 considered a public health problem (Bagramian, 2009). If not treated in time, it can affect not only chewing
77 function but also speech, smile and the psychosocial environment and quality of life of the child and family
78 (Losso, 2009). Dental caries is a multifactorial disease involving bacterial, dietary and host determinants
79 influenced by multiple sociological and environmental factors (Uribe, 2009; Leong *et al.*, 2013).

80 Diagnosis of the disease Caries is a complex process which involves the interpretation of a set of data
81 from clinical signs and symptoms and complementary exams (Nyvad, 2004; Pretty, 2006). The method for
82 detecting carious lesions must have some characteristics essential to be considered adequate. Be reliable,
83 non-invasive, capable of detecting caries lesions at an early stage and capable of differentiating lesions
84 reversible from irreversible. In addition to affordable cost, comfort for the patient, speed and ease of
85 execution, it must also be viable for all surfaces of the teeth with adequate accuracy (Marinho, 1998).

86 The radiographic detection of tooth decay is fundamentally based on the fact that with the
87 progression of a tooth decay lesion, the mineral content of tooth enamel and dentin decreases, resulting in
88 an attenuation of the X-ray beams when they pass through the tooth. These characteristics can be observed
89 in the image with increased radiographic density (Silva, 2008). The radiographic examination has high
90 sensitivity in detecting dentin caries lesions but low sensitivity for detecting enamel caries lesions (Soares *et al.*, 2012). The most recommended technique for the radiographic detection of caries is “bitewing,” also
91 known as interproximal (Braga, 2010; Wenzel, 2004). The interproximal radiography allows a better
92 estimation of the more sensitive depth of proximal and occlusal caries in dentin than clinical inspection alone
93 (Purger, 2011)

94
95 The artificial neural network [ANN] is the basis of machine learning [“machine learning - ML”] and
96 deep learning [“deep learning - DL”]. A *neural network* is an algorithm composed of layers connected that
97 processes input data, which were extracted from the X-ray image, using convulsive neural networks, and
98 through an activation function, it will classify patients as sick or not sick in the output layer. Artificial
99 intelligence [AI], specifically deep learning using convolutional neural networks [CNNs], has been suggested
100 to improve the reliability and validity of diagnosis by image analysis. CNNs allow mapping an input (image)
101 to an output (classification) based on a set of weights and learned data (LeCun *et al.*, 2015). Accuracy above
102 90% enables the development of software and applications that will assist the dentist in decision-making,
103 enabling the diagnosis of the disease for the population in censuses, public services and private practices.

104 This work aims to evaluate the sensitivity, specificity and accuracy of using deep machine learning
105 through CNNs in diagnosing dental caries in periapical radiographs of adults through a systematic review.

106

107 **Material and methods**

108 Our study protocol was registered on PROSPERO (CRD42024411477). This systematic review and
109 meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)
110 standards. Therefore, prior to the beginning of the research, a search was carried out in the literature to find
111 out whether there was a need for a systematic review or whether other reviews that addressed the proposed
112 topic already existed. As we did not find matching work, we drew the protocol for the Systematic registration,
113 and the project followed the criteria recommended by PRISMA and the Minimum information about the
114 clinical artificial intelligence modelling checklist (MI-CLAIM).

115

116 **Search strategy**

117 Two authors (VGAP and RFP) separately performed PubMed, MEDLINE, LILACS and Cochrane Library
118 database search using standard search formulas. To search for gray literature (unpublished studies), we used
119 the following sources: ISI Web of Knowledge, British Library Inside, BMC Meeting Abstracts, OpenSIGLE,
120 Clinical Trials database and REBEC. Additionally, we researched the thesis database, government publications
121 and CAPES.

122 We created the search strategy using subject descriptors (index) for MeSH (Medical Subject
123 Headings) and DECS (Health Science Descriptors). To search for works, we selected terms, the increase in
124 uncontrolled vocabulary with synonymous words, acronyms, related terms and spelling variations ("entry
125 terms"), as shown in tables 1 to 3 in the different bases used. The search rule and Syntax were through
126 Boolean logic that uses logical operators OR, AND and NOT, as shown below:

127

128 **VHL:** MH:" Dental Caries " OR (Tooth Decay) OR (Tooth Decay) OR (Cavities) OR (Dental Cavities) OR (Dental Cavities) OR (Decayed Tooth) OR (Carious Lesions) OR (White Spots) OR (Dental Caries) OR (Caries) OR (Caries Dentales) OR (Carious Lesions) OR (White Spots Dentales) OR (Dental Caries) OR (Caries, Dental) OR (Carious Dentin) OR (Carious Dentins) OR (Carious Injury) OR (Carious Injuries) OR (Decay , Dental) OR (Dental Decay) OR (Dental White Spot) OR (Dental White Spots) OR (Dentin , Carious) OR (Dentins , Carious) OR (Lesion , Carious) OR (Lesions , Carious) OR (Spot, Dental White) OR (Spots, Dental White) OR (White Spot, Dental) OR (White Spots, Dental) AND MH:" Machine Learning" OR (Machine Learning) OR (Machine Learning) OR (Transfer Learning) OR (Machine Learning) OR (Machine Learning) OR (Transfer Learning) OR (Automatic Learning) OR (Transfer Learning) _ OR (Machine Learning) OR (Learning, Machine) OR (Learning, Transfer) OR (Transfer Learning)

138 **PUBMED:** "Dental Caries"[Mesh] OR (Dental Caries) OR (Caries, Dental) OR (Carious Dentin) OR (Carious Dentins) OR (Carious Lesion) OR (Carious Lesions) OR (Decay, Dental) OR (Dental Decay) OR (Dental White Spot) OR (Dental White Spots) OR (Dentin, Carious) OR (Dentins, Carious) OR (Lesion, Carious) OR (Lesions, Carious) OR (Spot, Dental White) OR (Spots, Dental White) OR (White Spot, Dental) OR (White Spots, Dental) AND "Machine Learning"[Mesh] OR (Machine Learning) OR (Learning, Machine) OR (Learning, Transfer) OR (Transfer Learning)

144

145 **Eligibility Criteria**

146 We included the literature published in PubMed, MEDLINE, LILACS, and the Cochrane Library
147 database and gray literature on diagnosing dental caries in adults using AI in radiographic images.

148 The inclusion criteria was as follows: Participants: patients with occlusal and/or interproximal dental
149 caries of both sexes in permanent dentition who were diagnosed using interproximal or periapical X-ray
150 imaging; Index test: studies that used the use of artificial intelligence as an index test; Reference test: Visual
151 diagnosis by specialists in the field; Outcomes: sensitivity, specificity and accuracy and studies that evaluated
152 caries in percentage or qualitatively classifying in presence and absence; Types of studies: observational
153 cross-sectional studies, cohort, case-control and clinical trials.

154 Excluded criteria were case reports and animal studies, decayed primary teeth, studies that analyzed
155 artificial intelligence for the diagnosis of diseases other than dental caries in the field of dentistry knowledge,

156 studies that the response variable (outcome) or parameters have not been clearly reported, impossibility of
157 extracting study data and lack of response after contact with the study authors.

158

159 **Study selection**

160 Before searching for articles, we calibrated reviewers to ensure high reliability in selecting works.
161 The references of the selected articles were listed using the Rayyan reference manager
162 <<http://rayyan.qcri.org/reviews/5>>. We removed the duplicate articles after adding the search results from
163 all databases. Two reviewers (VGAP and RFP) selected the articles, analyzing them by titles and abstracts and
164 reached a consensus on the disagreements between reviewers. We created a clinical form containing the
165 eligibility criteria to record the reasons for exclusion from each study using the flowchart. We excluded
166 ineligible studies based on the eligibility criteria. After selecting the works, we obtained the full articles.

167

168 **Data extraction**

169 The following data were extracted from the included literature: reference with the first Author and
170 year of publication, type of study, classification of study quality (risk of bias), number of images used in
171 articles, type of neural network used (architecture, number of neurons), hyperparameters adopted, dental
172 caries was categorized in presence or absence, sensitivity of the diagnostic method, specificity of the
173 diagnostic method, accuracy of the method or area on the ROC curve.

174

175 **Quality of articles**

176 Two independent reviewers (VGAP and RFP) classified the articles by the instrument of critical
177 evaluation and their risk of bias using the tool Quadas 2 (“Quality Assessment of Diagnostic Accuracy Studies-
178 2”). Reviewers rated each work by consensus based on this tool to qualify the studies.

179

180 **Statistical analysis**

181 We used the sensitivity and specificity data, with the number of positive tests obtained by the “gold”
182 standard and the number of tests to generate a 2x2 table with the TP, FP, FN and TN rates. With these rates,
183 we recalculated the sensitivity and specificity values with their respective 95% confidence intervals and
184 presented them in forest graphs (Forest Plots). As it is common that the presence of heterogeneity in
185 diagnostic accuracy studies is not only by chance but also by the implicit variation of the cutoff point
186 (threshold effect), the summary point that represents the sensitivity and rate of false positives (1-specificity)
187 were combined using a bivariate model of random effects (Reitsma et al., 2005). We chose the bivariate
188 model of random effects because it considers the correlation between the rates of sensitivity and false
189 positives of the included studies. We evaluated the heterogeneity of the summarized ROC curve [SROC] by
190 the I2 approach of Zhou & Dendurkuri, implemented in the Mada package. In addition to the summary point
191 of sensitivity and rate of false positives (1-specificity), we also calculated the Positive Likelihood Ratio and
192 Negative (RV+ and RV-) together with the estimate of the Diagnostic Odds Ratio [DOR] through the MCMC
193 procedure of Zwindermann et al. (2008) parameters obtained by the bivariate model. We performed all
194 analyses in R language (version 4.2 for Mac iOS) with the help of Mada (version 0.5.11) and DTAPlots (version
195 1.0.2.5) packages.

196

197 **Quality assessment and publication bias**

198 To evaluate the quality of the studies, we applied the QUADAS-2 risk checklist to test the bias risk in
199 each study (Table 2). The articles selected presented a low risk of bias. Most of the authors followed the
200 STARD and/or CLAIM standards. QUADAS2 presents four domains (patient selection, Index test, reference
201 test and follow-up over time).

202

203 **Results**

204 Based on the search strategy, 225 articles were selected from the databases. Among these, 78
205 duplicates were removed, and 57 were excluded for other reasons, resulting in 90 articles. Upon reviewing
206 the title and abstract, 57 papers were excluded for not meeting the eligibility criteria, 2 did not receive

207 responses from the authors, leaving 31 for full-text reading. Of these 31, 13 articles were deemed eligible for
208 meta-analysis (Figure 1).

209

210 **Fig 1.** Flowchart.

211

212 In total, true positive (TP), false positive (FP), false negative (FN), and true negative (TN) rates from
213 13 studies were utilized to generate the Forest Plot illustrating sensitivity and specificity. The sensitivity of
214 machine learning for caries detection ranged from 0.58 to 0.90 across studies, and specificity ranged from
215 0.68 to 0.95. The test for equality was statistically significant for both sensitivity ($\chi^2=188.34$; $p<0.0001$) and
216 specificity ($\chi^2=332.37$; $p<0.0001$) with a weak negative correlation ($\rho=-0.071$) between sensitivity rates and
217 false positives (1-specificity), as demonstrated in Figure 2.

218 The result obtained by the bivariate random-effects model revealed that the sensitivity and
219 specificity of machine learning for detecting dental caries were 0.79 (95% CI: 0.73; 0.84) and 0.87 (95% CI:
220 0.81; 0.92), respectively. The summary receiver operating characteristic (SROC) curve's area under the curve
221 (AUC) was 0.885, and the heterogeneity, estimated by the I² statistic using Zhou & Dendukuri's approach,
222 was I² = 45.3% (Figure 2).

223

224 **Fig 2.** Forest Plot of machine learning accuracy measures in dental caries detection.

225

226 Figure 3 depicts the SROC, with the summary point of sensitivity and false positive rate (1-specificity)
227 of machine learning for detecting dental caries. The SROC demonstrated an accuracy of 0.885 encompassing
228 all studies within the prediction region. We calculated the Positive and Negative Likelihood Ratios along with
229 the Diagnostic Odds Ratio estimate as presented in Table 3.

230

231 **Fig 3.** SROC (“Summary Receiver-Operating Characteristic”).

232

233 Additionally, based on the data, we constructed the Fagan's nomogram, as shown in Figure 4. The
234 Fagan's nomogram illustrated that for a positive caries test diagnosed by convolutional neural networks, the
235 probability of the tooth having caries is 86.01%, and the probability of it being a false negative when caries
236 is not detected is 20%.

237

238 **Fig 4.** Fagan's nomogram

239

240 Discussion

241 Among the machine learning techniques until 2018, the supervised vector method [SVM] was the
242 most used classification method. According to Santana et al. (2022) SVM have classification algorithms with
243 better sensitivity values when compared to ANN. However, SVM and CNNs result in similar conclusions. CNNs
244 are a class of deep learning algorithms that have been widely used in computer vision tasks, including caries
245 detection. One of the main advantages of CNNs is the ability to detect relevant characteristics at different
246 spatial scales, which is especially important for detecting caries, such as in X-ray images or smartphone
247 photos.

248 SROC demonstrated an accuracy of 0.885, containing all studies within the prediction region and a
249 summarized sensitivity of 0.79 (95%CI 0.73 to 0.84) and summarized specificity of 0.87 (95%CI 0.81 to 0.92)
250 and heterogeneity of 45.3%. Of the articles selected for the meta-analysis, five authors worked with regions
251 of interest [ROIs] to reduce dimensionality and make the diagnosis more accurate. However, sensitivity
252 ranged from 0.58 to 0.80 and specificity from 0.77 to 0.98, not the only way to improve accuracy.

253 The sensitivity results among the authors ranged from 0.68 to 0.95. The two authors with the lowest
254 sensitivity values were Askar et al. (2021) and Scwendicke et al. (2023), and the authors with the highest
255 sensitivity values were Kühnisch et al. (2022) and Duong et al. (2021). The authors who presented the lowest
256 specificity were Duong et al. (2021) and Park et al. (2022), and the highest specificity values were the works

257 of Estai et al. (2022) and Kim et al. (2022). There is an inverse relationship between sensitivity and specificity.
258 Our review observed this relationship with the authors Duong et al. (2021) and Askar et al. (2021).

259 The clinical validity of caries diagnosis is complex because it requires a clinical diagnosis in which it
260 needs to present clean, dry and well-lit teeth. Rx is considered a complementary examination to the clinical
261 examination and requires experience from the dentist. Thus, even being considered the gold standard is not
262 without error. The X-ray quality, as contrast and proper positioning, is another preponderant factor for a
263 correct diagnosis. In all articles, dentists diagnosed from photos or X-rays, making the diagnosis less accurate
264 in more subtle cases, such as white spots and caries without cavities, regardless of the evaluator's expertise,
265 which can also impact machine learning.

266 The type of image in our study consisted of both photos of teeth of smartphones and images of X-
267 ray (panoramic or interproximal). They are different images because the photos of smartphones and high-
268 resolution cameras are coloured, while those from X-ray have grayscale. The authors who used camera
269 photos were Askar et al. (2021), Schwendicke et al. (2023), Duong et al. (2021), Kühnisch et al. (2022) and
270 Park et al. (2022). The other authors used X-ray (bitewing or panoramic) or CT scans. This diversity of images
271 may be one of the explanations for the heterogeneity among the authors. The hyperparameters the authors
272 used showed similarities in the learning rate (between 0.001 and 0.00001). The number of times varied
273 considerably among the authors (10 to 10,000). There were also variations between the number of batches
274 (2 to 150).

275 Most authors used Transfer Learning and cross-entropy as a method of deep learning. Another
276 important factor to highlight and may be related to the differences in sensitivities and specificity of AIs
277 learning is the network architecture adopted in each study. The high number of layers with fewer images can
278 cause overfitting. Of the selected studies, 23.08% (3 articles) used 50 layers, 30.77% (4 articles) used 18
279 layers, and 15.38% (2 articles) used 22-25 layers of artificial neurons. The author who presented high
280 sensitivity and specificity values was Kühnisch et al. (2022), who used 1891 images of all types of teeth in
281 training the network architecture of 18 layers and the type of learning of Transfer Learning and cross-entropy.

282 The sample size is a crucial point for machine learning. The sample size and proportion of the disease
283 are parameters that alter diagnostic test sensitivity and specificity values (Linnet et al. 2012). The volume of
284 images in training is one of the main factors for adequate learning and, consequently, better accuracy values.
285 However, having a database with large volume, veracity, and image quality is the most significant difficulty
286 encountered in studies. There were large variations between authors regarding the number of images
287 obtained and augmented for training. The images were enlarged by delimiting regions of interest, rotating
288 among other techniques, and may have improved sensitivity and/or specificity. However, new studies with
289 different severities of caries disease need to be carried out to improve the accuracy of CNN, especially in
290 initial caries where the diagnosis is difficult.

291 The articles selected presented a low risk of bias since most of the authors followed the STARD
292 and/or CLAIM standards. QUADAS2 presents four domains (patient selection, Index test, reference test and
293 follow-up over time). In the case of artificial intelligence use, several domains receive low risk because, for
294 the first domain, a large number of patients and images are obtained from several sources randomly. As for
295 the index and pattern domains, all images pass the pattern for the index test to learn to diagnose. Thus, a
296 low risk of bias was attributed. However, it is worth remembering that QUADAS2 is a tool to assess the risk
297 of bias from conventional diagnostic tests and is not being developed for diagnostic test analysis for AIs. With
298 the increasing use of IA in diagnostic tests, there is a need to develop a specific tool for these cases in
299 systematic reviews. Only the author Kim et al. (2022) classified it as uncertain due to the incomplete
300 description of the network architecture of the index test and the network training method. However, it is
301 worth remembering that the tool still needs to be specific to evaluate articles that use artificial intelligence
302 as an index test.

303 The DOR measures the effectiveness of the diagnostic test analyzed and is independent of the
304 disease's prevalence, unlike the test's accuracy. The AIs presented a DOR of 25.11 times of the test when
305 giving positive results of the individual with the disease concerning those without the disease and a 95%
306 confidence interval ranging from 13.96 to 45.20. We can also verify through the Fagan nomogram that the
307 CNNs presented the probability of 86.1% of diagnosing existing caries and only 20% of false negatives.

308 Despite the values obtained in our study, the improvement of the sensitivity, specificity and accuracy
309 of CNNs can still be explored in future studies with the impact analysis of the number of training images,
310 types of images, prevalence of disease and different architectures in the diagnosis of caries, enabling its use
311 in modern dentistry with greater safety.

312

313 Conclusion

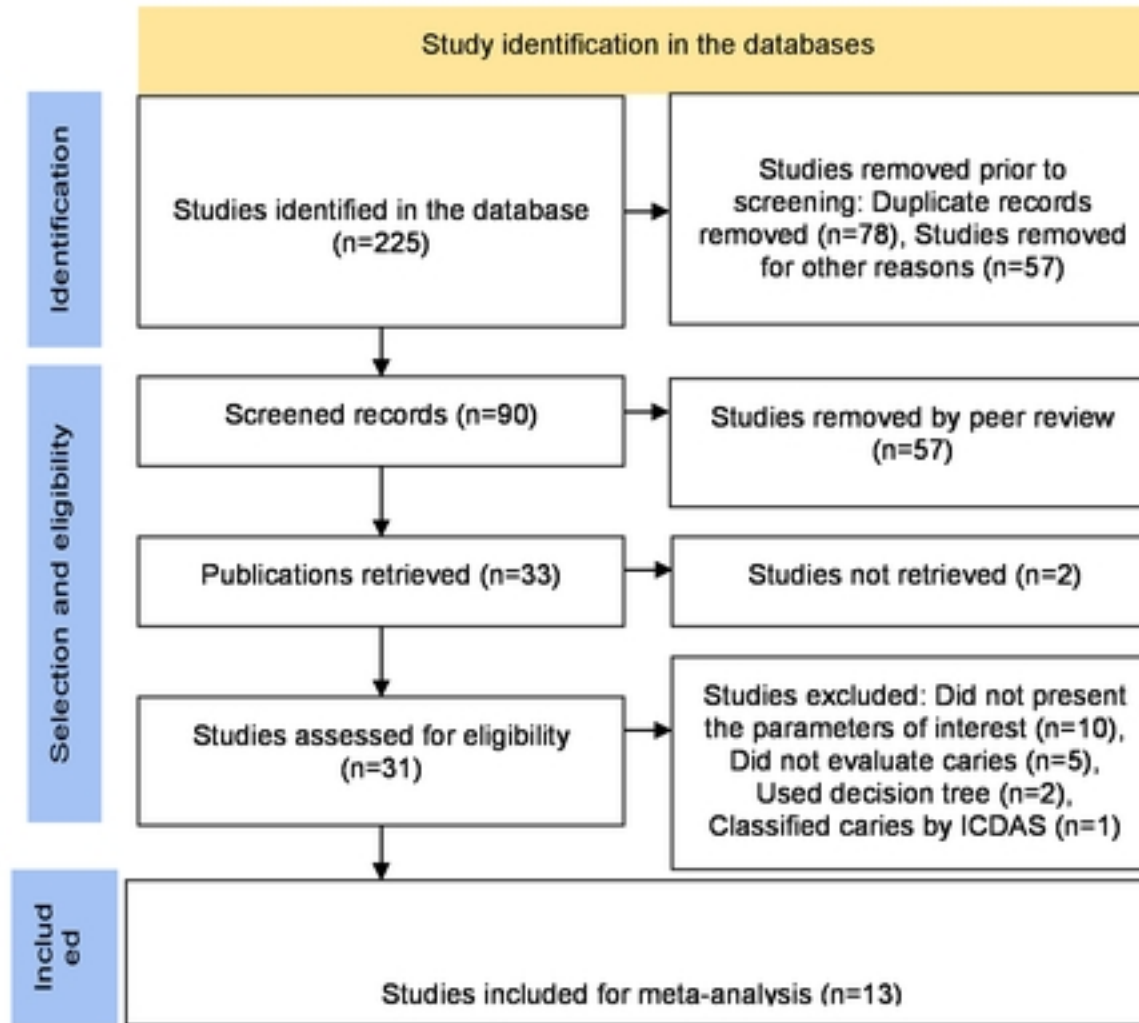
314 Using artificial intelligence through convolutional neural networks to diagnose caries showed global
315 accuracy (0.88), and it can be trained to detect visual patterns that indicate the presence of caries in both X-
316 ray images and photos and can be used in new devices to assist the dentist in diagnosing caries disease.

317

318 References

- 319 1. Bagramian R.A.; Garcia-Godoy F.; Volpe A.R. 2009. The global increase in dental caries. A pending
320 public health crisis. *Am J Dent* 22(1):3–8.
- 321 2. Braga M.M.; Mendes F.M.; Ekstrand K.R. 2010. Detection activity assessment and diagnosis of dental
322 caries lesions. *Dent Clin North Am* 54 (3): 479-93.
- 323 3. LeCun Y.; Bengio Y.; Hinton G. 2015. Deep learning. *Nature* 521(7553): 436–444.
- 324 4. Leong P.M.; Gussy M.G.; Barrow S.Y.; de Silva-Sanigorski A.; Waters E. 2013. A systematic review of
325 risk factors during first year of life for early childhood caries. *Int J Paediatr Dent* 23(4):235–250.
- 326 5. Losso E.M.; Tavares M.C.R.; Silva J.Y.B.; Urban C.A. 2009. Cárie precoce e severa na infância: uma
327 abordagem integral. *J Pediatr* 85(4):295-300.
- 328 6. Marinho V.A., Pereira G.M. 1998. Revisão de literatura cárie: diag- nóstico e plano de tratamento.
329 *Rev Un Alfenas* 4: 27-37.
- 330 7. Nyvad D, B. 2004. Diagnosis versus detection of caries. *Caries Res* 38(3): 192-8.
- 331 8. Pretty I.A. 2006. Caries detection and diagnosis: Novel Technologies. *J Dent* 34 (10): 727-39.
- 332 9. Purger F.P.; Oliveira P.R.A.; Vasconcellos A. et al. 2011. Relative importance of radiographs in
333 diagnosing primary molar’s proximal caries. *Journal Dental Research* 90, IADR/AADR General Session
334 & Exhibition Abstracts.
- 335 10. Reitsma J.B.; Glas A.S.; Rutjes W.S.; Scholten R.J.P.M.; Bossuyt P.M.; Zwinderman A.H. 2005. Bivariate
336 analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.
337 *Journal of Clinical Epidemiology* 58: 982–990.
- 338 11. Santana C.P.; Carvalho E.A.; Rodrigues I.D.; Bastos G.S.; Souza A.D.; Brito L.L. 2022. Rs-fMRI and
339 Machine learning for ASD diagnosis: A systematic review and meta-analysis. *Scientific Reports* 12:
340 6030.
- 341 12. Silva P.R.D.S. 2008. Avaliação da acurácia dos métodos radiográficos convencional e digital direto na
342 análise da extensão de lesões de cárie oclusais em molares decíduos. *Estudo In Vitro. São Paulo tese*
343 (doutorado) – FO/USP.
- 344 13. Soares G.G.; Souza P.R.; Purger F.P.C.; Vasconcellos A.B.; Ribeiro AA. 2012. Métodos de detecção de
345 Cárie. *Rev bras odontol* 69 (1):84-9.
- 346 14. Uribe S. 2009. Early childhood caries–risk factors. *Evid Based Dent* 10(2):37-38.
- 347 15. Wenzel A. 2004. Bitewing and digital bitewing radiography for detec- tion of caries lesions. *J Dent*
348 *Res* 83: C72-5.
- 349 16. Zhou Y. & Dendukuri N. 2014. Statistics for quantifying heterogeneity in univariate and bivariate
350 meta-analyses of binary data: the case of meta-analyses of diagnostic accuracy. *Stat Med* 33(16):
351 2701-17.
- 352 17. Zwinderman A.H.; Glas A.S.; Bossuyt P.M. 2008. Statistical models for quantifying diagnostic accuracy
353 with multiple lesions per patient. *Biostatistics* 9(3):513–522.

1

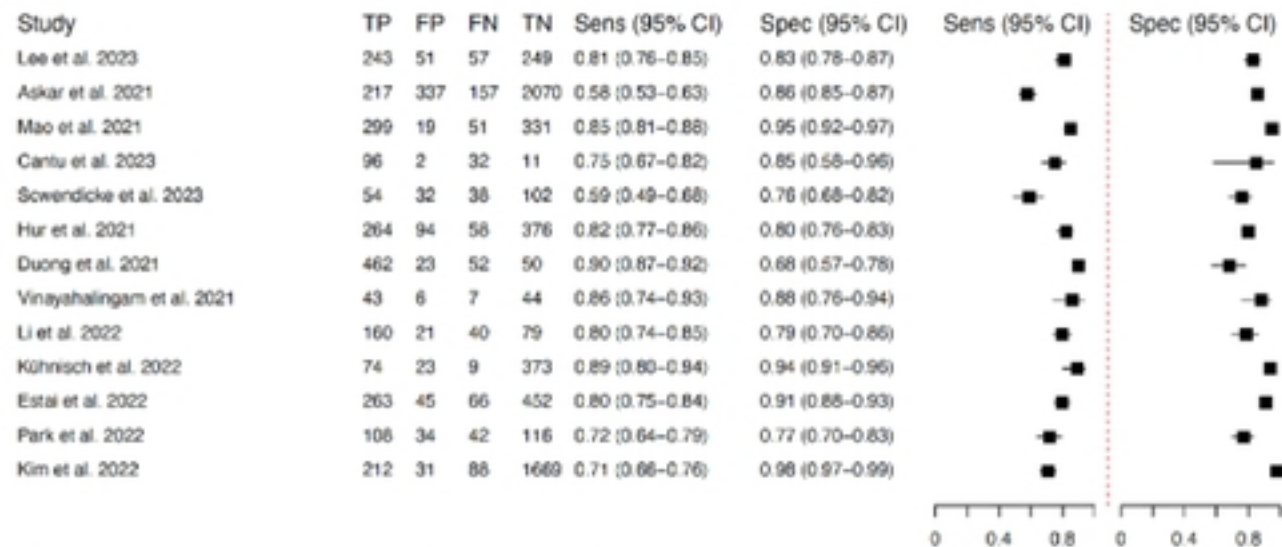


2

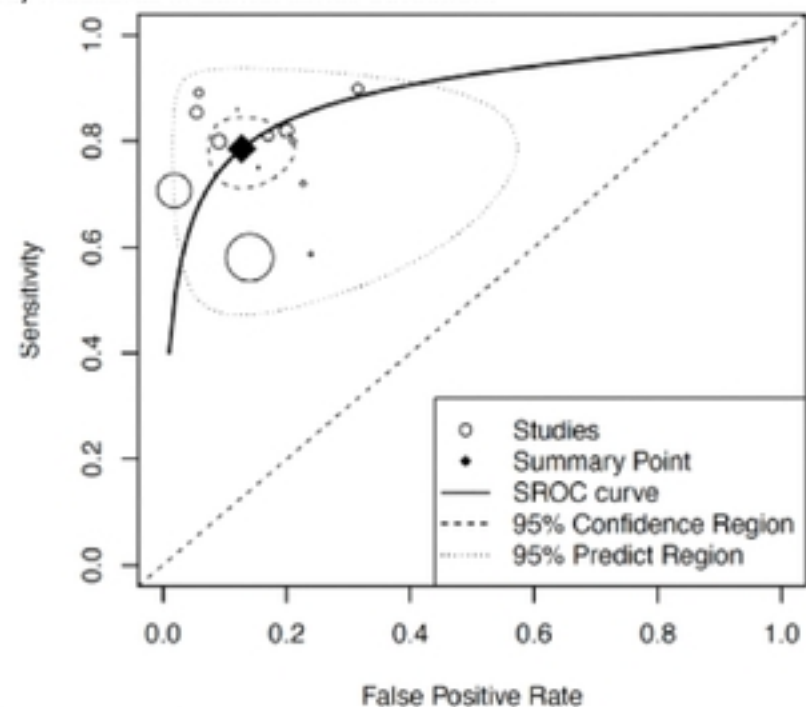
3

4

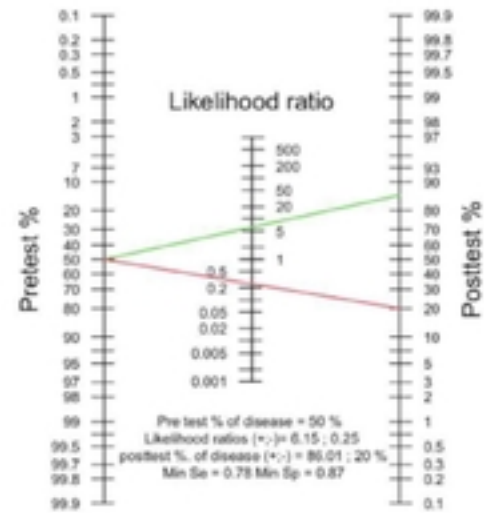
Fig 1. Flowchart.



5
6 **Fig 2.** Forest Plot of machine learning accuracy measures in dental caries detection.



7
8 **Fig 3.** SROC ("Summary Receiver-Operating Characteristic").



##	pre	post.pos	post.neg	min.LRpos	max.L
Rneg	minSe	minSp			
##	0.5000000	0.8601399	0.2000000	6.1500000	0.250
0000	0.7817797	0.8728814			

10
 11 **Fig 4.** Fagan's nomogram

12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25

26 **Table 1.** Characteristics of selected articles
27

Author/ Year	Positive gold standard image (test bench)	Total image (Test bench)	Tooth type	Image acquisition method	Image area	Total images	Total training	Training with cavities	Learning type	Number of layers	Batches	Seasons	Momentum	Learning rate
Lee et al. 2023	300	600	Molar and/or premolar	Periapical X-Ray	Tooth	3000	2400	1200	Transferring learning	22	32	1000	NI	0.01
Askar et al. 2021	374	2781	Incisors and canines	Pictures	ROI	2781	.	374	.	8	150	.	0.9	0.00006
Mao et al. 2021	350	700	Molar and/or premolar	Bitewing X-Ray	Tooth labeled	3716	700	350	Transferring learning	25	64	100	0.9	0.00006
Cantu et al. 2023	128	141	Molar and/or premolar	Bitewing X-Ray	Tooth Labeled	3686	3293	.	Adaptive moment estimation	NI	2	190	NI	0.00005
Schwendicke et al. 2023	92	226	Extracted molar and/or premolar	Intraoral Picture	Tooth	226	226	92	Transferring learning	50	NI	NI	NI	0.00001
Hur et al. 2021	322	792	Molar	Panoramic X-Ray and tomography	Tooth	2642	2400	NI	Transferring learning	NI	NI	NI	NI	0.0001
Duong et al. 2021	514	587	Molar and/or premolar	Smartphones Pictures	Tooth	587	NI	293	Transferring learning	50	NI	10	NI	0.0001
Vinayahalingam et al. 2021	50	100	Third molar	Panoramic X-Ray	Tooth	500	320	200	Transferring learning	18	32	10	NI	0.0001
Li et al. 2022	200	300	All types	Periapical X-Ray	ROI	4129	3829	2062	Cross-entropy	18	50	200	NI	0.0001
Kühnisch et al. 2022	83	479	All types	Nikon Pictures	Tooth	2417	1891	348	Transferring learningn / cross-entropy	18	16	50	NI	0.001
Estai et al. 2022	329	862	Molar and/or premolar	Bitewing X-Ray	ROI	8618	7756	2967	Transferring learning	50	64	10	NI	0.001
Park et al. 2022	150	300	Molar and/or premolar	Intraoral Picture	ROI	1938	1638	678	Cross-entropy	18	NI	20	NI	0.001
Kim et al. 2022	300	2000	Molar and/or premolar	Panoramic X-Ray	ROI	10	NI	NI	Softmax	NI	NI	10	NI	0.001

28 ROI regions of interest. NI not informed.

29 **Table 2.** Risk of bias of selected studies

Author	Year	Domain 1		Domain 2		Domain 3		Domain 4		Bias Risk
		A	B	A	B	A	B	A		
Lee et al.	2023	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Askar et al.	2021	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Mao et al.	2021	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Cantu et al.	2023	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Schwendicke et al.	2023	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Hur et al.	2021	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Duong et al.	2021	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Vinayahalingam et al.	2021	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Li et al.	2022	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Kühnisch et al.	2022	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Estai et al.	2022	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Park et al.	2022	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Kim et al.	2022	LOW	LOW	Unclear	Unclear	LOW	LOW	LOW	LOW	Unclear

30

31 **Table 3.** Estimates of diagnostic accuracy of machine learning to detect dental caries.

Estimators	Median	CI inferior	CI superior
Positive Likelihood	6,15	4,03	9,63
Negative Likelihood	0,25	0,19	0,32
Odds Ratio Diagnostic	25,11	13,96	45,2

32