Systematic Review

SENSITIVITY EVALUATION OF DEEP LEARNING-BASED MODELS FOR DENTAL CARIES DETECTION IN BITEWING RADIOGRAPHS: SYSTEMATIC REVIEW AND META-ANALYSIS

ABSTRACT

Aims: To analyse the sensitivity of deep learning models in detecting carious lesions in bitewing radiographs through a systematic review and meta-analysis.

Study Design: Systematic review and meta-analysis.

Place and Duration of Study: The review was conducted using the Cochrane Library, LILACS, PubMed, Scielo, and the Virtual Health Library databases, covering articles published until March 2024.

Methodology: The research question was developed using the PICOT framework. Paired and independent searches with no filters applied were performed using the registered Medical Subject Headings "(machine learning) OR (deep learning) AND (dental caries) AND (radiography, bitewing)" and its related entry terms, covering articles relevant to the study's theme with no language or time restrictions. A total of 2,841 articles were initially identified, with 8 selected after title and abstract screening and removal of duplicates. After full-text review, 6 articles were included for analysis. The risk of bias was assessed using the QUADAS-2 tool, and the data were analysed with Review Manager 5.4 and R software.

Results: All included studies were classified as having a low risk of bias. The meta-analysis revealed that deep learning models demonstrated moderate sensitivity (0.77) in detecting carious lesions, with an overall performance AUC of 0.779. In comparison, dentists achieved higher performance with an AUC of 0.886. Significant heterogeneity was observed, largely due to variations in model architecture, training datasets, and possibly image preprocessing techniques. These factors impacted the models' diagnostic accuracy, emphasising the need for standardisation.

Conclusion: Deep learning models currently exhibit moderate sensitivity, and their variable performance underscores the importance of not relying on them in isolation for caries detection.

Instead, they should serve only as supplementary diagnostic tools to assist clinicians. Further research is necessary to enhance model reliability, address heterogeneity, and establish standardised evaluation frameworks.

Keywords: Dental Caries; Bitewing Radiography; Artificial Intelligence; Deep Learning.

INTRODUCTION

Dental practices have significantly evolved in tandem with modernity, driven by the direct influence of techno-scientific advances in healthcare. Despite sophisticated diagnostic and procedural improvements, dental caries remain the most common oral condition and non-communicable disease worldwide.[1,2]

Approximately 3.5 billion people globally suffer from carious lesions in their teeth, resulting from an oral dysbiotic imbalance caused by the interaction of primary factors: host, diet, microorganisms, and time. This disease is exacerbated by the interaction with a myriad of secondary modulating factors, notably standing out in the socioeconomic and demographic context of the individual and the quality not only of their saliva but of their entire immune system.[1,3,4,5]

The investigation of carious lesions should primarily be conducted through visual and tactile inspection, which can be enhanced by incorporating modern auxiliary tools such as transillumination and fluorescence techniques. Clinical confirmation holds paramount importance and is well supported by radiographs, especially bitewing radiographs, which may improve the detection of carious lesions on proximal surfaces that often go unnoticed when only visually examined.[6,7]

The detection of carious lesions is particularly challenging when located on proximal surfaces, making bitewing radiographs indispensable tools for ensuring accurate diagnosis.

Moreover, this imaging modality is a valuable resource not only for detection and assessment but also for monitoring the lesion following treatment (Schwendicke; Göstemeyer, 2020).

Advances in artificial intelligence (AI), a term coined in the 1950s to define the concept of machines capable of performing tasks typically done by humans, have permeated medical practices, including those of dental surgeons. The advent of deep learning (DL), an advanced form of machine learning (ML) that uses various types of artificial neural networks as its basic functional architecture, has facilitated its introduction into both medical and dental imaging centers.[8]

The functional architecture used in ML is generally limited by computational power and dependent on the quantity and quality of data used for training. It is simpler and represents some of the oldest approaches developed. In contrast, DL uses Convolutional Neural Networks (CNN), initially developed to mimic the functionality of human neurons, offering more modern and precise data analysis, whether abstract or complex.[9]

The increasing popularization and incorporation of AI into various aspects of modern life, such as speech recognition and object identification, have sparked interest in understanding, exploring, and expanding its functionality. Consequently, given the potential increase in the accuracy of detecting alterations in radiographs through AI modalities, it is evident that understanding their functionality is crucial for enhancing its applicability in dentistry, in the context of carious lesion detection, especially in bitewing radiographs.[10,11]

Although significant advances have been made in integrating AI with dental radiographic analysis, there remains a gap in understanding the real potential of its detection capabilities and its related parameters. Therefore, this study aims to analyze the sensitivity of AI in detecting carious lesions in bitewing radiographs through a systematic review and a meta-analysis.

METHODS

Research protocol registration

The present study was registered on the International Prospective Register of Systematic Reviews (PROSPERO) platform (https://www.crd.york.ac.uk/prospero) under the following identification code: CRD42023452759.

Research information and strategy

The research question was formulated based on the PICOT technique, where "P" refers to the population, "I" to the intervention, "C" to the comparison, "O" to the outcomes, and "T" to the types of study. In this context, "P" refers to patients with suspected caries, "I" to the use of artificial intelligence in detecting carious lesions, "C" to human radiographic analysis for detecting carious lesions, "O" to the efficacy of AI in detecting carious lesions, and "T" to the diagnostic accuracy studies included. Consequently, the research question posed was: "Are machine learning and deep learning modalities sensitive in detecting carious lesions in bitewing radiographs of patients with suspected caries?".

The searches were conducted in March 2024 across the following electronic databases: Cochrane Library, Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS), Public Medline (PubMed), Scientific Electronic Library Online (Scielo) and Virtual Health Library (VHL). Data collection was performed by two independent researchers. In cases of doubts or disagreements, a third researcher was consulted.

The searches were carried out using a combination of base descriptors and their respective entry terms registered on the Medical Subject Headings (MeSH) platform: "(machine learning) OR (deep learning) AND (dental caries) AND (Radiography, Bitewing)" (Table 1). The configuration was adapted according to each tool used to find diagnostic accuracy studies, in any language, without a defined time criterion related to the topic under evaluation, and no filters were used.

Table 1. Descriptors and Entry Terms as recorded in the MeSH database, and the common research strategy format utilized for database searches.

Descriptors and basic research strategy	Entry Terms
#1	"Machine Learning" [MeSH terms] OR (Learning, Machine) OR (Transfer Learning) OR (Learning, Transfer)
#2	"Deep Learning" [MeSH terms] OR (Learning, Deep) OR (Hierarchical Learning) OR (Learning, Hierarchical)
#3	"Dental Caries" [MeSH terms] OR (Caries, Dental) OR (Dental Cavity) OR (Dental Decay) OR (Dental Cavities) OR (Cavities, Dental) OR (Cavity, Dental) OR (Carious Lesions) OR (Lesion, Carious) OR (Lesions, Carious) OR (Lesions, Carious) OR (Decay, Dental) OR (Carious Dentin) OR (Carious Dentins) OR (Dentins, Carious) OR (Dental White Spot) OR (Spot, Dental White) OR (Spots, Dental White) OR (White Spot, Dental) OR (White Spots, Dental) OR (Dental White Spots)
#4	"Radiography, Bitewing" [MeSH terms] OR (Bitewing Radiography) OR (Bitewing Radiographies) OR (Radiographies, Bitewing)
Common Research Strategy	#1 OR #2 AND #3 AND #4

Source: Authors (2024)

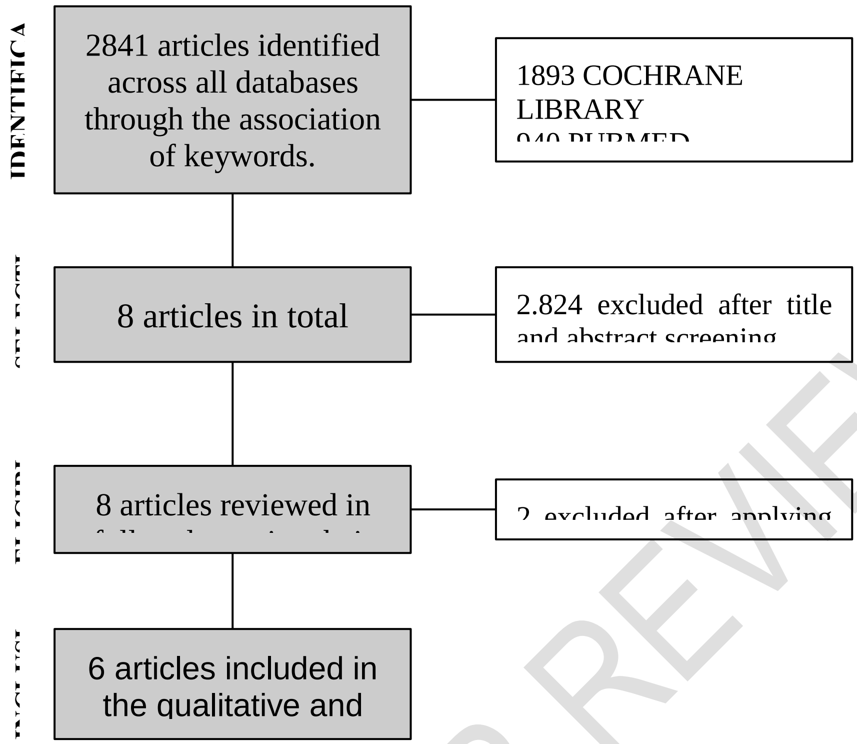
Eligibility criteria

Articles on the topic indexed in the consulted databases, regardless of the publication language and without predefined time criteria, were included if they evaluated ML- or DL-based models for detecting caries lesions in both analog and digital bitewing radiographs, explicitly reported sensitivity in their performance evaluation, and were accessible in full text. In contrast, articles indexed in the selected databases were excluded if they did not analyze or explicitly and objectively state the sensitivity of the proposed or evaluated model(s); studies that did not present an objective final average value quantifying sensitivity; studies associated with other types of tests or integrated classification methods; studies where ML and DL were applied to other types of radiographic images, such as panoramic and periapical; studies without a control group or with a control group not composed of dental surgeons; and other studies categorized as any other types of reviews on the topic.

Study selection

An initial screening was performed by two independent researchers for reading titles and abstracts, discarding those that explicitly did not relate to the core of the research. Then, the pre-selected studies were analyzed in full text, applying the inclusion and exclusion criteria, discarding those that did not align with the study's purpose (Figure 1).

Figure 1. Flowchart for identification and selection of studies.



Source: Authors (2024)

Data extraction

The included selected studies were tabulated by one researcher and carefully checked by two others in a Google Sheets spreadsheet and identified by authorship, year of publication, type of AI used, and a summary of their methods and main results. After completing the article selection stage and constructing the grouping table, these studies were individually analyzed qualitatively and quantitatively using ReviewManager 5.4 (RevMan 5.4).

Bias risk assessment

The risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool in RevMan 5.4 by a single researcher, following the software's standard protocol for diagnostic accuracy studies. This protocol was not altered and was completed after proper calibration by two researchers to address the objectives of this study. The parameters considered were patient selection, index test, reference standard, and flow and timing. The results were subsequently discussed with the research team after they reviewed the original articles, with no discrepancies identified.

Meta-analysis

The sensitivity, specificity, and accuracy were assessed. The data were tabulated using Google Sheets and subsequently exported to RevMan 5.4 and R software for analysis. A random-effects model was employed to synthesize the data. Heterogeneity among the included studies was evaluated using the I^2 and Tau^2 statistics. Additionally, the risk of publication bias was examined using Begg's and Egger's tests. To further assess the impact of individual studies on the overall meta-analytic outcomes, a leave-one-out analysis was conducted. Moreover, a Receiver Operating Characteristic (ROC) curve was generated to compare the performance of the DL-based model groups with that of dentists without assistance via Area Under the Curve (AUC) analysis.

Final article development

Finally, this article was written following the guidelines of the Preferred Reporting Items for Systematic reviews and Meta-Analyses literature searching extension (PRISMA-S) protocol.[12]

RESULTS

Included studies

The searches conducted in the selected databases resulted in a total of 2841 studies. Following title and abstract screening by two independent researchers, 17 articles were initially selected. After removing duplicates, 8 remained. However, 2 were excluded upon full-text analysis and application of eligibility criteria, leaving 6 articles to be included in this review (Table 2).

Table 2. Characterization of selected studies regarding authorship, year of publication, study country of origin, objective, AI-type employed, radiographs used for testing and validation, reported overall sensitivity percentage, and their main findings.

Authorship, Year, and Country of Origin	Objective	ALG	AG (n)	T (n)	TR (n)	S (%)	Main Findings
Bayraktar; Ayan, 2021, Turkey[13]	To investigate the effectiveness of CNNs in diagnosing interproximal carious lesions in digital bitewing radiographs.	YOLO	2	800	200	72.26	The model can be employed for carious lesion detection in bitewing radiographs, demonstrating robust performance and high accuracy.
Lee *et al.*, 2021, South Korea[14]	To develop a CNN model for detecting carious lesions in bitewing radiographs and investigate how this model can enhance clinical performance.	U-NET	3	304	50	65.02	The model can assist clinicians in diagnosing carious lesions more accurately as a second opinion. However, for more stable and precise results, additional data is needed for training.
Bayrakdar *et al.*, 2022, Turkey[15]	To recommend a model for automatic detection and segmentation of caries based on CNN algorithms and evaluate the clinical performance of the model compared to human analysts.	VGG	2	518	50	77	The CNNs have the potential to accurately and effectively detect and segment dental caries in bitewing radiographs. They also have the potential to assist professionals in clinical practice by providing quick and reliable confirmation of caries, making them a beneficial support for dentistry.
Baydar *et al.*, 2023, Turkey[16]	To perform a diagnostic evaluation of an AI-supported model based on CNNs for assessing bitewing radiographs.	U-NET	2	1052	132	84	The model can be used to automatically assess bitewing radiographs, and the results are promising. This allows professionals in a busy clinical setting to work more efficiently and quickly.

Authorship, Year, and Country of Origin	Objective	ALG	AG (n)	T (n)	TR (n)	S (%)	Main Findings
Chen *et al.*, 2023, China[17]	To evaluate the validity of CNNs based on DL for detecting carious lesions in bitewing radiographs.	Faster R-CNN	3	818	160	72	CNNs can assist in detecting proximal carious lesions in bitewing radiographs.
Ayhan; Ayan; Bayraktar, 2024, Turkey[18]	To automatically detect and number teeth in digital bitewing radiographs obtained from patients, and assess the real-time diagnostic efficiency of carious teeth using DL algorithms.	YOLO	2	1000	100	83.3	The proposed model showed promising results, highlighting the potential use of CNNs for tooth numbering and simultaneous detection of teeth and carious lesions.

ALG – algorithm used for detection task; AG – annotation group; AI-type – type of Artificial Intelligence; DL – deep learning; T – training radiographs; TR – test radiographs; (n) – numerical value; S% – overall sensitivity percentage for dental caries detection; CNNs – Convolutional Neural Networks.

Source: Authors (2024)

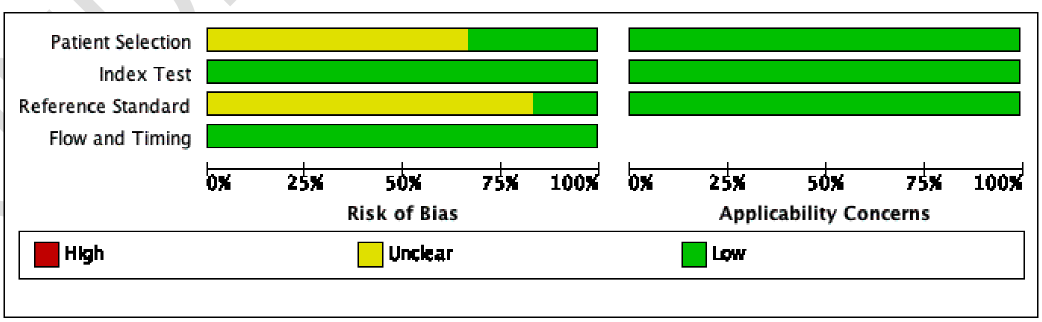
Descriptive analysis

The included studies explore the detection of dental carious lesions in bitewing radiographs using DL-based models (Table 2), summarizing the construction, training, validation, and testing of various models for this purpose. Overall, these studies, while differing in minor details such as the numbers of radiographs used for each stage, professionals in the control group, and the type of algorithm employed, showed some similarities from the step-by-step process taken for database construction to the final performance evaluation.

Bias risk assessment

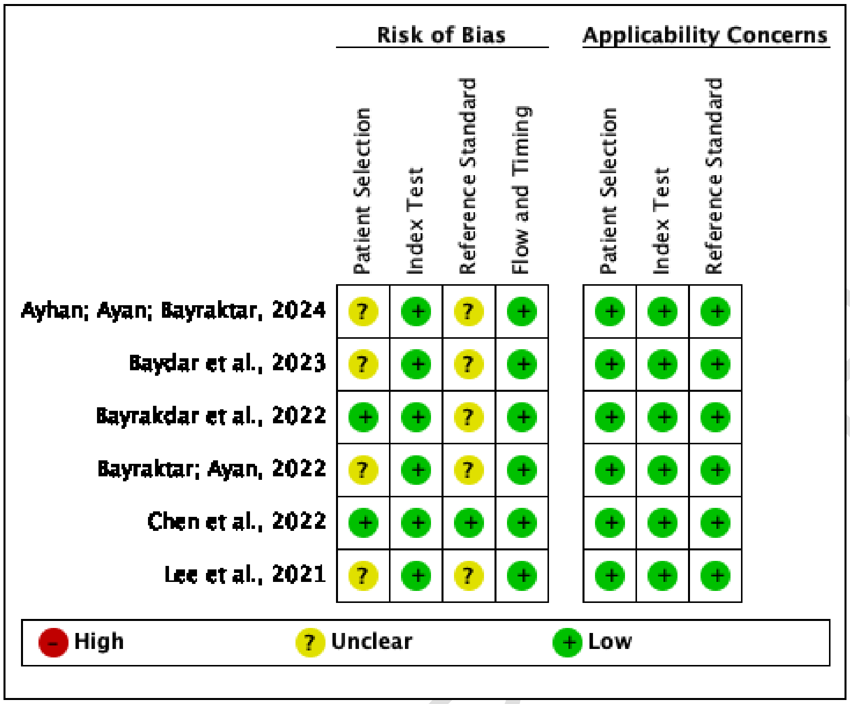
These studies were assessed for bias using RevMan 5.4 (Figures 1 and 2), demonstrating low risk. Risks related to patient selection and the construction of the reference group were considered uncertain, as some inherent steps in these processes were not adequately described. Nonetheless, these uncertainties did not seem to compromise the quality of the DL evaluation tests or its functionality. Thus, none of the evaluated studies were excluded; all were included in both qualitative and quantitative analyses.

Figure 2. Methodological quality assessment graph of selected studies.



Source: Authors (2024)

Figure 3. Summary of methodological quality assessment.

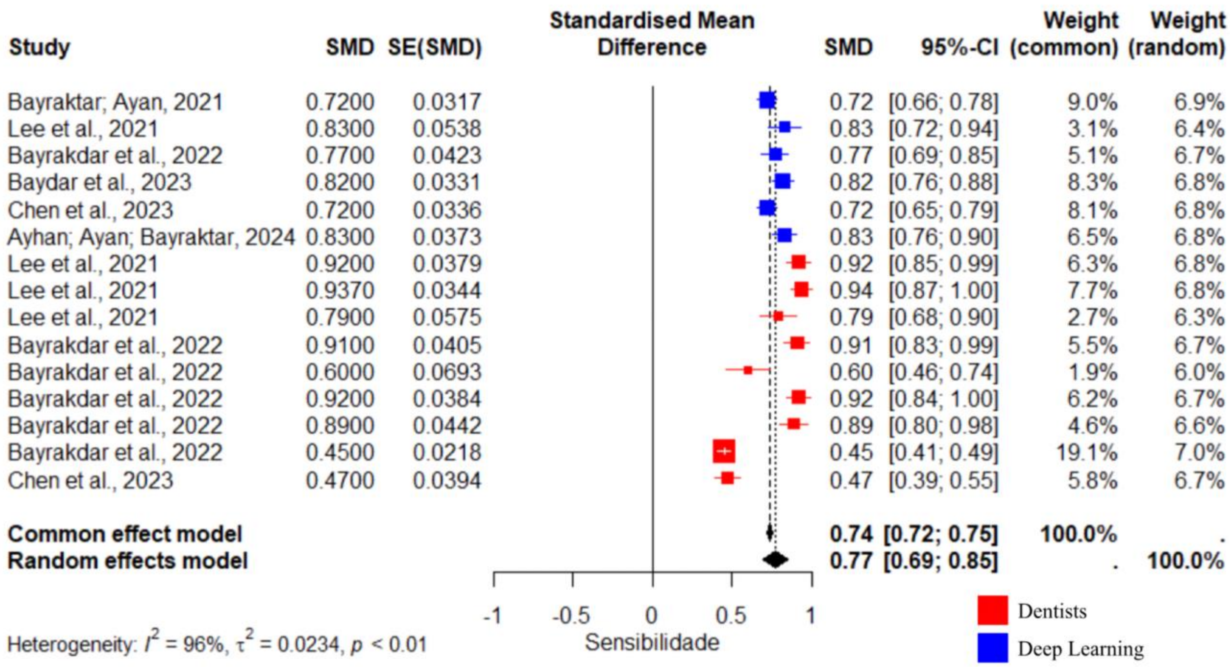


Source: Authors (2024)

Quantitative results

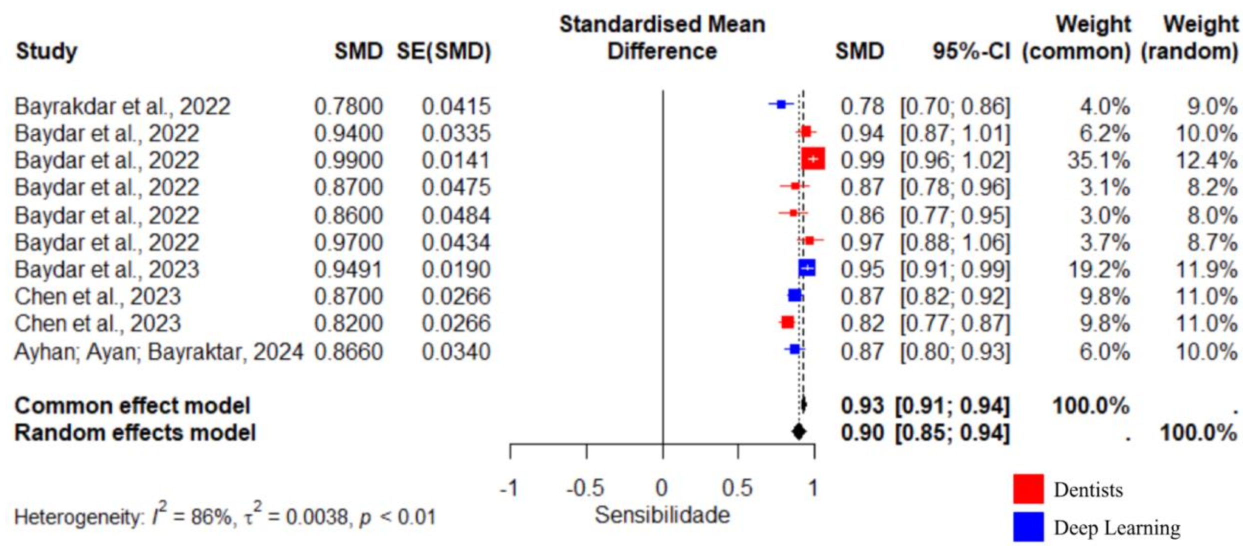
All articles were included in the meta-analysis. Results from parameters related to carious lesion detection in bitewing radiographs by DL-based models showed high heterogeneity in sensitivity, the ability to detect caries presence (Figure 4), and accuracy, the ability to correctly predict caries presence or absence (Figure 5). The specificity, the ability to detect caries absence, showed low performance (Figure 6). Detection capability evaluation for both groups was expressed by the AUC metric and expressed in a ROC curve (Figure 8), a decrease in DL model performance (0.779) was observed compared to dental surgeons (0.886).

Figure 4. Meta-analysis assessment of sensitivity for carious lesion detection in bitewing radiographs by dentists and DL-based models.



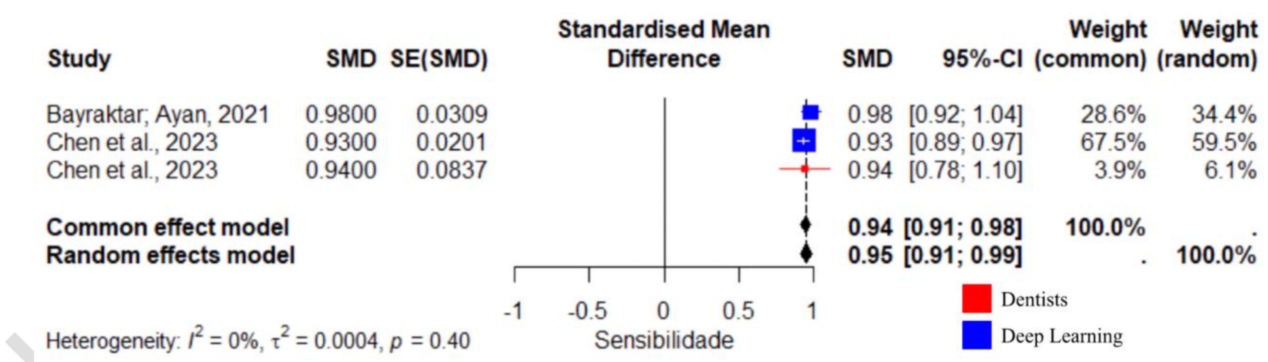
Source: Authors (2024)

Figure 5. Meta-analysis assessment of accuracy for carious lesion detection in bitewing radiographs by dentists and DL-based models.



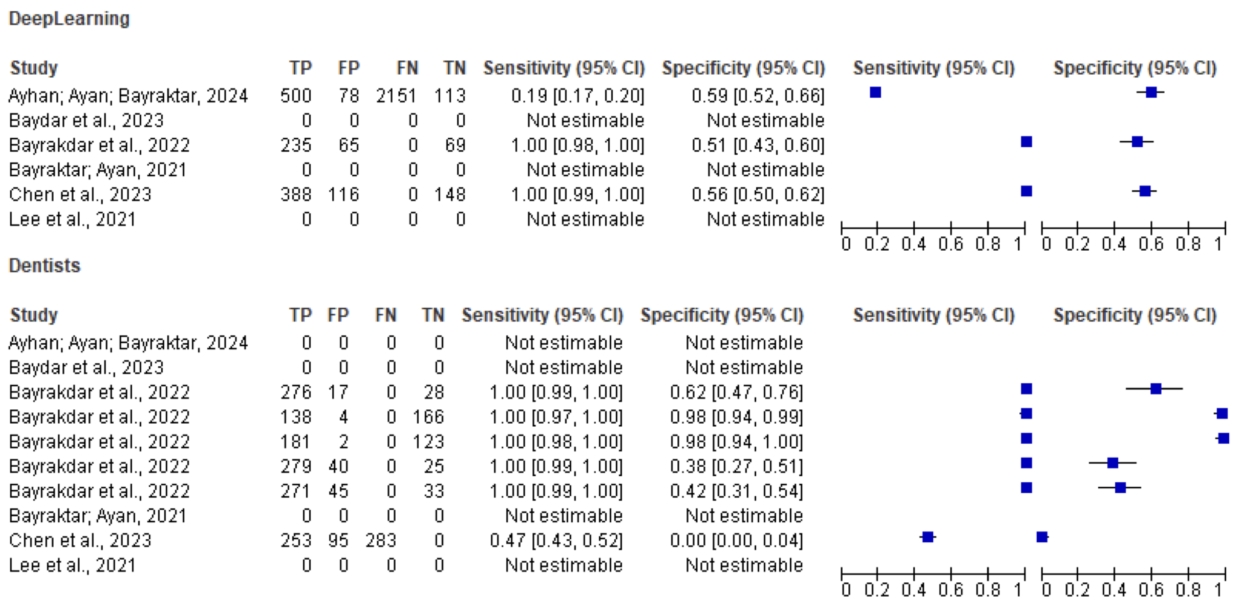
Source: Authors (2024)

Figure 6. Meta-analysis assessment of specificity for carious lesion detection in bitewing radiographs by dentists and DL-based models.



Source: Authors (2024)

Figure 7. Meta-analysis of sensitivity and specificity for carious lesion detection in bitewing radiographs by dentists and DL-based models.



DeepLearning

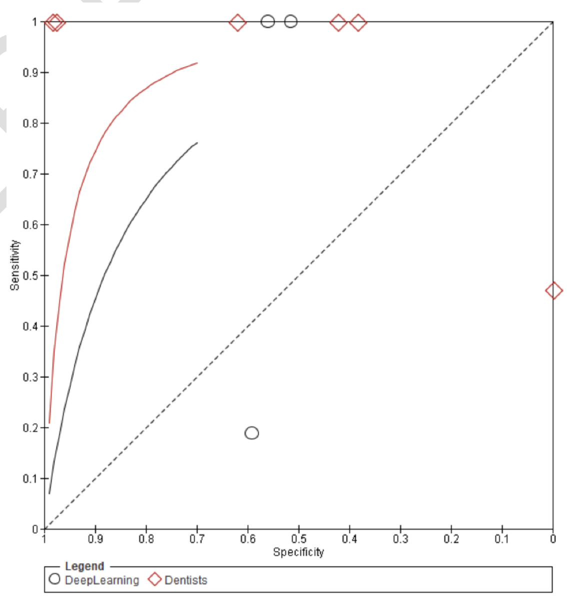
Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ayhan; Ayan; Bayraktar, 2024	500	78	2151	113	0.19 [0.17, 0.20]	0.59 [0.52, 0.66]
Baydar et al., 2023	0	0	0	0	Not estimable	Not estimable
Bayrakdar et al., 2022	235	65	0	69	1.00 [0.98, 1.00]	0.51 [0.43, 0.60]
Bayraktar; Ayan, 2021	0	0	0	0	Not estimable	Not estimable
Chen et al., 2023	388	116	0	148	1.00 [0.99, 1.00]	0.56 [0.50, 0.62]
Lee et al., 2021	0	0	0	0	Not estimable	Not estimable

Dentists

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ayhan; Ayan; Bayraktar, 2024	0	0	0	0	Not estimable	Not estimable
Baydar et al., 2023	0	0	0	0	Not estimable	Not estimable
Bayrakdar et al., 2022	276	17	0	28	1.00 [0.99, 1.00]	0.62 [0.47, 0.76]
Bayrakdar et al., 2022	138	4	0	166	1.00 [0.97, 1.00]	0.98 [0.94, 0.99]
Bayrakdar et al., 2022	181	2	0	123	1.00 [0.98, 1.00]	0.98 [0.94, 1.00]
Bayrakdar et al., 2022	279	40	0	25	1.00 [0.99, 1.00]	0.38 [0.27, 0.51]
Bayrakdar et al., 2022	271	45	0	33	1.00 [0.99, 1.00]	0.42 [0.31, 0.54]
Bayraktar; Ayan, 2021	0	0	0	0	Not estimable	Not estimable
Chen et al., 2023	253	95	283	0	0.47 [0.43, 0.52]	0.00 [0.00, 0.04]
Lee et al., 2021	0	0	0	0	Not estimable	Not estimable

Source: Authors (2024)

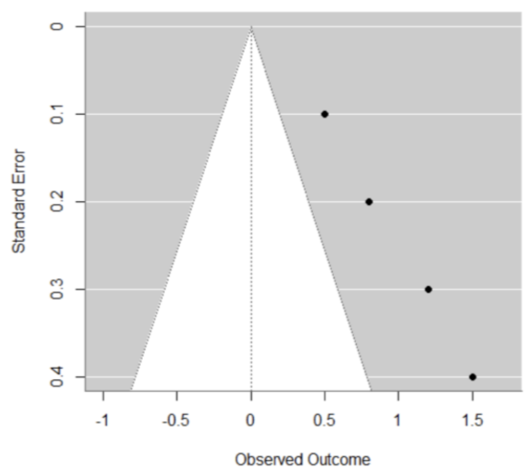
Figure 8. ROC curve demonstrating the AUC relationship in classification performance between DL models and dentists.



Source: Authors (2024)

Beggs tests (S = 2, P-value = 0.3333, τ = 0.8) and Egger tests (T = 5, P-value = 0.3333, τ = 0.6666667) were also conducted, showing no significant evidence of publication bias (Figure 9).

Figure 9. Publication bias risk assessment graph of included studies.



Source: Authors (2024)

DISCUSSION

Detection of dental carious lesions by DL-based models can streamline dental surgeons' clinical routines; however, such mechanisms are prone to errors that demand correct development, training, validation, and adequate calibration of the machinery to be minimized. All stages are interdependent, and the beginning of model development requires good data feeding, identification, and marking of structures visible in bitewing radiographs to correctly identify and distinguish carious lesions under various dental conditions.[19]

The selection of radiographs for database construction aimed at model training is a crucial step that requires careful attention from the professionals behind these systems. Lee et al.[14] emphasized the importance of including cases similar to those expected in clinical practice and excluding radiographs with low image quality, excessive distortion, or severe overlap of proximal surfaces due to the anatomical arrangement of specific teeth, as these characteristics interfere with caries diagnostic accuracy.

Studies employing a larger number of radiographs for training, such as Ayhan, Ayan, and Bayraktar[18] with 1,000 and Baydar et al.[16] with 1,052, demonstrated higher sensitivity. Thus, it is confirmed that the greater the number of radiographs used for training, the better the results obtained by these models.

A complete evaluation of DL-based models requires consideration of multiple performance metrics, including sensitivity, specificity, and precision. This study focused on sensitivity, also referred to in the literature as recall, defined in the current context by ForouzeshFar et al.[19] as the probability of correctly classifying a carious tooth. Individual findings on the sensitivity of the model from each study were evaluated through meta-analysis (Figure 4) and, when possible, also assessed in conjunction with other parameters (Figure 7).

In addition to sensitivity, this study analyzed accuracy, which represents the proportion of correct predictions made by the model (Figure 5), as well as specificity, reflecting the model's overall ability to perform its detection function (Figure 6). These results were summarized in an AUC relationship and illustrated in a ROC curve (Figure 8). Notably, dental surgeons achieved a higher score (0.886) compared to DL models (0.779), although both demonstrated acceptable performance.

Interestingly, the slight overall superiority of dental surgeons in the evaluated studies reinforces the complexity of diagnosing carious lesions in radiographs, a task that can be challenging even for professionals with decades of experience depending on the case. Therefore, radiographs are still considered a secondary examination in the face of suspicions, and lesion evaluation and confirmation should primarily be done clinically. However, this does not diminish the importance of DL-based models, which can contribute to radiograph evaluation and consequently result in clinical procedural time savings.[13–18]

Regarding the bias assessment in the included articles, although the selection of bitewing radiographs was generally random, some steps in the selection process were poorly described by the authors. The markings made prior to DL model training by dental surgeons were systematically performed in only one study[17], while others used simple consensus methods or did not mention them at all, which could potentially affect the critical evaluation by reviewers at this stage. However, as previously stated, this did not appear to impact the reported results of the evaluated studies, and all the studies were considered to have a low risk of bias. In addition, Beggs and Egger tests (Figure 9) concluded the absence of significant evidence of publication bias.

The high heterogeneity arising from the inherent methodological differences among the studies does not compromise the interpretation of the results of this meta-analysis. It can be explained by differences in the data used, their processing, the DL architecture employed and its number of layers, and any other variables that may influence the construction of the system. The studies converge on the conclusion that DL-based models serve as effective support for detecting carious lesions in bitewing radiographs. However, the findings from these models should not be regarded as definitive, as there is still insufficient evidence to justify disregarding the confirmation of these findings by a qualified professional.[13–18]

Understanding parameters related to detection capability is crucial, as many studies in this field focus solely on overall performance. The lack of specific data, such as sensitivity and precision, limits the selection of studies for inclusion, as some do not report final sensitivity values. This highlights a significant failure in the reporting of the performance of a developed and formally published product.

The models could potentially achieve greater sensitivity with an increased number of training radiographs. The use of Data Augmentation (DA) is recognized as a valuable technique that ensures robust data diversity for the algorithm to learn from, thus addressing this limitation. Even large datasets can benefit from DA, as the model is exposed to a significantly greater volume of data for training, thereby improving outcomes.[13, 15, 17]

For effective model development, it is crucial that the selected reviewers are properly calibrated regarding the criteria for identifying these lesions. It is also important to emphasize the need to provide a thorough description of the professionals involved in the labeling group for the radiographs included in the training. Varying their levels of professional experience will better reflect the true capacity for human detection, as noted by Bayrakdar et al.,[15] who found that one of the professionals in their group achieved lower scores than the model. Furthermore, the labeling process must be clearly described in the text to minimize the risk of bias during classification.

During the literature search for study selection, despite including descriptors and entry terms related to ML, no studies were identified using this AI modality for carious lesion detection in bitewing radiographs. This is believed to be due to the fact that the machinery of ML-based models is not ideal for supporting tasks as complex as image interpretation compared to DL, which has the potential to achieve very satisfactory results with proper development.[10]

Additionally, this research identified challenges in evaluating studies that present their results using combined metrics, which directly impacts the understanding of how the described product functions. As a result, such studies were excluded for not fully meeting the eligibility criteria. This practice, stemming from a lack of crucial information, may reflect failures in the quality of reporting findings and suggest potential bias in the authors' results, which may be incomplete or obscured by other factors.

According to Ver Berne et al.[20], traditional metrics like accuracy or sensitivity may not fully capture the clinical usability of DL models, particularly in complex fields like radiology. Future evaluations should include external dataset validation and real-world testing in clinical settings to assess the practical benefits and limitations of these systems in assisting dental professionals, ensuring that these tools provide actionable and reliable support in practice.

Further research on this topic is recommended to gain a better understanding of it and its associated factors. This is crucial for effectively disseminating information to the dental community and encouraging improvements in performance. The growing popularity of the subject also indicates a smoother integration of AI and its various applications into routine dental practice, extending beyond the confines of large imaging centers and technological hubs, thereby significantly contributing to its true democratization within real-world dental clinics.

Finally, this study acknowledges the current superiority of humans over DL models in terms of sensitivity, which possibly reflects the contemporary incipiency of these prototypes. The potential of these advancements remains unknown, emphasizing the ongoing need for further research and discussions to substantiate future projections on the topic.

CONCLUSION

The evaluated DL-based models have shown moderate sensitivity and acceptable overall performance in detecting caries lesions in bitewing radiographs. However, their results cannot be considered in isolation, as there is still insufficient basis to justify dispensing with the need for a dentist to confirm the system's findings. Therefore, these models continue to serve merely as tools to enhance diagnostic accuracy and reduce the time spent identifying these alterations.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

The authors declare that the AI technology, GPT-4o, was used fairly and solely for rewriting and editing this manuscript, with the specific purpose of correcting the English grammar of the translated text, which was originally written in Brazilian Portuguese. No additional information was inserted into the text; the AI's role was limited to verifying and refining the accuracy of the translation. Details of the AI usage are as follows:

1. The original manuscript, written in Brazilian Portuguese, was translated into English, and the AI was employed to ensure the final text adhered to the grammatical standards of academic English.
2. Specific prompts were designed to guide the AI in providing grammatical corrections and verifying the translation's alignment with academic conventions.

REFERENCES

1. Zewdu T, Abu D, Agajie M, Sahilu T. Dental caries and associated factors in Ethiopia: systematic review and meta-analysis. Environ Health Prev Med. 2021;26(1):21. https://doi.org/10.1186/s12199-021-00943-3
2. Zhang JS, Chu CH, Yu OY. Oral Microbiome and Dental Caries Development. Dent J (Basel). 2022;10(10):184. https://doi.org/10.3390/dj10100184
3. Martignon S, Roncalli AG, Alvarez E, Aránguiz V, Feldens CA, Buzalaf MAR. Risk factors for dental caries in Latin American and Caribbean countries [published correction appears in Braz Oral Res. 2021 Jul 19;35(Supp 1):e053err. doi: 10.1590/1807-3107bor-2021.vol35.0053.erratum] [published correction appears in Braz Oral Res. 2021 Sep 06;35(Supp 1):e053err2. doi: 10.1590/1807-3107bor-2021.vol35.0053erratum2]. Braz Oral Res. 2021;35(suppl 01):e053. https://doi.org/10.1590/1807-3107bor-2021.vol35.0053
4. Giacaman RA, Fernández CE, Muñoz-Sandoval C, *et al.* Understanding dental caries as a non-communicable and behavioral disease: Management implications. Front Oral Health. 2022;3:764479. https://doi.org/10.3389/froh.2022.764479
5. Anil S, Porwal P, Porwal A. Transforming Dental Caries Diagnosis Through Artificial Intelligence-Based Techniques. Cureus. 2023;15(7):e41694. https://doi.org/10.7759/cureus.41694
6. Al Saffan AD. Current Approaches to Diagnosis of Early Proximal Carious Lesion: A Literature Review. Cureus. 2023;15(8):e43489. https://doi.org/10.7759/cureus.43489

7. Panyarak W, Suttapak W, Wantanajittikul K, Charuakkra A, Prapayasatok S. Assessment of YOLOv3 for caries detection in bitewing radiographs based on the ICCMS™ radiographic scoring system [published correction appears in Clin Oral Investig. 2023 Apr;27(4):1743. doi: 10.1007/s00784-023-04865-y]. Clin Oral Investig. 2023;27(4):1731-1742. https://doi.org/10.1007/s00784-022-04801-6
8. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. J Dent Res. 2020;99(7):769-774. https://doi.org/10.1177/0022034520915714
9. Hwang JJ, Jung YH, Cho BH, Heo MS. An overview of deep learning in the field of dentistry. Imaging Sci Dent. 2019;49(1):1-7. https://doi.org/10.5624/isd.2019.49.1.1
10. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep Learning in Medical Image Analysis. Adv Exp Med Biol. 2020;1213:3-21. https://doi.org/10.1007/978-3-030-33128-3_1
11. Mao YC, Chen TY, Chou HS, *et al.* Caries and Restoration Detection Using Bitewing Film Based on Transfer Learning with CNNs. Sensors (Basel). 2021;21(13):4613. https://doi.org/10.3390/s21134613
12. Rethlefsen ML, Kirtley S, Waffenschmidt S, *et al.* PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev. 2021;10(1):39. https://doi.org/10.1186/s13643-020-01542-z
13. Bayraktar Y, Ayan E. Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. Clin Oral Investig. 2022;26(1):623-632. https://doi.org/10.1007/s00784-021-04040-1
14. Lee S, Oh SI, Jo J, Kang S, Shin Y, Park JW. Deep learning for early dental caries detection in bitewing radiographs. Sci Rep. 2021;11(1):16807. https://doi.org/10.1038/s41598-021-96368-7

15. Bayrakdar IS, Orhan K, Akarsu S, *et al.* Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. Oral Radiol. 2022;38(4):468-479. https://doi.org/10.1007/s11282-021-00577-9

16. Baydar O, Różyło-Kalinowska I, Futyma-Gąbka K, Sağlam H. The U-Net Approaches to Evaluation of Dental Bite-Wing Radiographs: An Artificial Intelligence Study. Diagnostics (Basel). 2023;13(3):453. https://doi.org/10.3390/diagnostics13030453

17. Chen X, Guo J, Ye J, Zhang M, Liang Y. Detection of Proximal Caries Lesions on Bitewing Radiographs Using Deep Learning Method. Caries Res. 2023;56(5-6):455-463. https://doi.org/10.1159/000527418

18. Ayhan B, Ayan E, Bayraktar Y. A novel deep learning-based perspective for tooth numbering and caries detection. Clin Oral Investig. 2024;28(3):178. https://doi.org/10.1007/s00784-024-05566-w

19. ForouzeshFar P, Safaei AA, Ghaderi F, Hashemikamangar SS. Dental Caries diagnosis from bitewing images using convolutional neural networks. BMC Oral Health. 2024;24(1):211. https://doi.org/10.1186/s12903-024-03973-

20. Ver Berne J, Saadi SB, Politis C, Jacobs R. A deep learning approach for radiological detection and classification of radicular cysts and periapical granulomas. J Dent. 2023;135:104581. https://doi.org/10.1016/j.jdent.2023.104581

21. Schwendicke F, Göstemeyer G. Conventional bitewing radiography. Clin. Dent. Rev. 2020;24(1):3981–3995. https://doi.org/10.1007/s41894-020-00086-8