REVIEW

# ARTIFICIAL INTELLIGENCE PLATFORMS IN DENTAL CARIES DETECTION: A SYSTEMATIC REVIEW AND META-ANALYSIS



LYNDON P ABBOTT[a], ANKITA SAIKIA[a], AND ROBERT P ANTHONAPPA[b]

[a]Paediatric Dentistry, UWA Dental School, The University of Western Australia, Perth, Australia
[b]Professor Paediatric Dentistry, UWA Dental School, The University of Western Australia, Perth, Australia.

## ABSTRACT

*Objectives*
To assess Artificial Intelligence (AI) platforms, machine learning methodologies and associated accuracies used in detecting dental caries from clinical images and dental radiographs.

*Methods*
A systematic search of 8 distinct electronic databases: Scopus, Web of Science, MEDLINE, Educational Resources Information Centre, Institute of Electrical and Electronics Engineers Explore, Science Direct, Directory of Open Access Journals and JSTOR, was conducted from January 2000 to March 2024. AI platforms, machine learning methodologies and associated accuracies of studies using AI for dental caries detection were extracted along with essential study characteristics. The quality of included studies was assessed using QUADAS-2 and the CLAIM checklist. Meta-analysis was performed to obtain a quantitative estimate of AI accuracy.

*Results*
Of the 2538 studies identified, 45 met the inclusion criteria and underwent qualitative synthesis. Of the 45 included studies, 33 used dental radiographs, and 12 used clinical images as datasets. A total of 21 different AI platforms were reported. The accuracy ranged from 41.5% to 98.6% across reported AI platforms. A quantitative meta-analysis across 7 studies reported a mean sensitivity of 76% [95% CI (65% - 85%)] and specificity of 91% [(95% CI (86% - 95%)]. The area under the curve (AUC) was 92% [95% CI (89% - 94%)], with high heterogeneity across included studies.

*Conclusion*
Significant variability exists in AI performance for detecting dental caries across different AI platforms. Meta-analysis demonstrates that AI has superior sensitivity and equal specificity of detecting dental caries from clinical images as compared to bitewing radiography. Although AI is promising for dental caries detection, further refinement is necessary to achieve consistent and reliable performance across varying imaging modalities.

## INTRODUCTION

Artificial intelligence (AI) is a device's ability to perform functions usually associated with human intelligence, such as reasoning, learning and self-improvement.[1] Machine Learning (ML) is a subfield of AI in which algorithms

*CORRESPONDING AUTHOR:*
*Lyndon P Abbott, Dental Registrar, UWA Dental School, The University of Western Australia, 17 Monash Ave, Perth 6009, Australia.*
*E-mail: lyndon.abbott@uwa.edu.au*

## KEYWORDS

*Artificial intelligence, Dental caries, Deep learning, Machine learning, Systematic review, Meta-analysis*

*Conflict of interest: The authors have no actual or potential conflicts of interest.*

*Received 16 August 2024; revised 25 November 2024; accepted 29 November 2024*

J Evid Base Dent Pract 2025: [102077]

1532-3382/$36.00



**doi: https://doi.org/10.1016/j.jebdp.2024.102077**

are applied to learn intrinsic statistical patterns and structures to make predictions on unseen data.[2] One use of ML in medicine is deep learning (DL), itself inspired by human brain neurons by feeding data through multiple layers and filters to identify hierarchical features, learn from each input, and ultimately optimise accuracy and performance.[3] Through recent advancements in computational power, data availability and the development of improved data processing software, AI applications have grown exponentially and guided current healthcare research. For example, preliminary data on AI has demonstrated performance metrics equal to or better than trained specialists in diagnosing lymph node metastases from tissue sections,[4] melanomas from clinical photographs,[5] and pneumonia from chest radiographs.[6] Such applications give promise that AI can transfer time-consuming tasks to machines, relieve burdened healthcare systems, make healthcare more affordable, and improve patient outcomes.

Dentistry can also benefit from AI applications. Ongoing research demonstrates how AI can improve the diagnostic performance of dentists in the identification of tooth numbering,[7] dental anomalies,[8] periapical pathologies,[9] and dental caries.[3] Current estimates predict untreated primary dental caries affects 500 million children globally,[10] and 42% of children aged 5 to 10 in Australia.[11] This need, coupled with the rise of tele-dentistry requiring dentists to screen and diagnose dental caries from clinical images, increases the potential of AI in dentistry today.[12]

Despite the growing integration of AI in healthcare and dentistry, there remains a limited understanding of its diagnostic performance in detecting dental caries. Key elements such as AI methodologies, data preprocessing techniques, and the quality of training datasets are often poorly documented, resulting in inconsistent diagnostic accuracy, insufficient training data, and a lack of standardization.[3,13-16] These gaps raise significant regulatory and ethical concerns and hinder the full utilization of AI in enhancing diagnostic capabilities. Additionally, the rapid increase of AI-based studies in dentistry highlights the need for ongoing evaluation to better understand the current AI landscape and its future role in dental caries detection. The aim of this systematic review and meta-analysis is to evaluate different AI platforms and deep learning methods used for dental caries detection, using both clinical images and radiographs, and identify areas for future research.

## METHOD

Reporting of this systematic review followed the PRISMA - Diagnostic Test Accuracy (DTA) Checklist,[17] and was registered at PROSPERO (CRD42022337833).

### Eligibility Criteria

#### *Inclusion Criteria*

Published studies that (i) involved AI platforms to detect dental caries, and (ii) used radiographs or clinical images as datasets were included. These studies were published between January 2000 and March 2024.

#### *Exclusion Criteria*

Studies that (i) did not report on the AI platform, (ii) published in a language other than English and, (iii) were unavailable in full text were excluded. In addition, studies using AI for purposes other than diagnosing dental caries have been excluded.

### Information Sources and Search

Electronic searches were conducted from January 2000 to March 2024 using the following 8 databases: Scopus, Web of Science, MEDLINE (PubMed), Educational Resources Information Centre (ERIC), Institute of Electrical and Electronics Engineers (IEEE) Explore, ScienceDirect, Directory of Open Access Journals and JSTOR. The search strategy is detailed in Table 1, with MeSH terms including AI, dental caries, DL, ML, systematic review and meta-analysis.

### Study Selection

Results from the search were input into EndNote 20, where 2 reviewers (LA, AS) screened the title and abstract independently for relevance and removed duplicate results. Subsequently, the full-text articles were assessed against the inclusion and exclusion criteria to finalise studies for qualitative and quantitative synthesis.

### Data Collection and Extraction

The following data, namely, bibliographic details, data modality, dataset size (total, training, validity, test), the definition of dental caries, labelling procedure, use of annotation tool, exclusion criteria, dentition, image augmentation, AI platform and performance metrics were extracted. In cases where more than one AI platform was used, results from both platforms were extracted and included as a range. Performance of the included studies was reported through each study using a test dataset to assess outcomes including sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), F1-score and Area Under the Receiver Operating Characteristic Curve (AUC). Evaluation of the different definitions of dental caries was performed by assessing the number of divisions between caries-free and dental caries encroaching the pulp. This included the following scales: 2-point scale ("sound" vs "carious"), 3-point scale (normal to advanced), 5-point scale,[18] and 7-point scale.[19]

Table 1. Search strategy.

| Search | Topic and terms |
|---|---|
| #1 | Artificial Intelligence: "artificial intelligence" OR "deep learning" OR "neural networks, computer" OR "machine learning" OR "fuzzy logic" |
| #2 | Dentistry: "dentistry" OR "tooth" OR "dental caries" OR "radiography, dental" OR "paediatric dentistry" |
| #3 | Search #1 and #2 |

### Reporting Standards Assessment

The reporting standard of each study was assessed through the Checklist for Artificial Intelligence in Medical Imaging (CLAIM).[20] This best practice checklist was selected because it was developed by a group of experts through a Delphi consensus process. It promotes transparency and reproducibility in research by assessing 42 important factors related to AI applications. These factors include study design, data sources, data processing, reference standards, data splitting, AI model structure, and the use of training, validation and testing datasets. In addition, the CLAIM checklist assesses the robustness of the statistical analysis, as well as limitations and implications for clinical practice. By evaluating each study in this way, the quality of AI-based studies on dental caries detection can be assessed, proving useful recommendations for future research.

### Risk of Bias (RoB) Assessment

Two reviewers (LR, AS) assessed each study for RoB using the Quality Assessment tool for Diagnostic Accuracy Studies (QUADAS-2).[21] The QUADAS-2 tool consists of 4 domains: patient selection, index test, reference standard, flow, and timing. In addition, it contains 3 domains that evaluate the applicability of each study: patient selection, index test, and reference standard. Studies were marked as 'high RoB' if any of the 4 domains was rated as 'no'; 'low RoB' if all 4 questions were rated as 'yes'; and 'unclear RoB' if any domains presented unclear reporting.

### Quantitative Synthesis

Data on the diagnostic accuracy of AI to detect dental caries was extracted from individual studies, and statistical analysis was conducted using STATA 17 software (STATA Corporation, College Station, Texas, USA) via MIDAS v.3.0 package and METANDI module. Forest plots were generated to display the summary estimates for sensitivity and specificity of AI in dental caries detection. The hierarchical summary receiver operating characteristic (HSROC) curve along with its 95% confidence and predictive region was generated in the SROC space by adopting the bivariate hierarchical model proposed by Rutter & Gatsonis.[22] The impact of unobserved heterogeneity across the studies was assessed using the $I^2$ statistic.[23]

## RESULTS

### Study Selection and Characteristics

Initially, 2538 articles were identified. After 168 duplicate articles were removed, 2370 articles underwent title and abstract screening, 2286 articles failed to meet the inclusion criteria, and 39 articles were excluded following full-text assessment. This resulted in 45 articles in the final analysis Figure 1, with a detailed breakdown of each study outlined in Table 2 ,[24-56] and Table 3.[57-68] The final 45 studies were published between 2008 and 2024, with 41 (91%) published between 2020 and 2024. Of the 33 studies examining dental radiographs, 17 used bitewing, 7 periapical, 4 panoramic, 1 cone beam computed tomography and 4 combined dental radiographs. Conversely, for the 12 studies examining clinical images, 4 used Digital Single Lens Reflex (DSLR) images, 4 images from mobile phones, 3 images from intraoral cameras, and 1 used a combination of different clinical photos. Annotation tools were used in 47% of all studies, with 85% of included studies using preprocessing augmentation to increase the dataset size.

### Datasets Used Across Included Studies

A significant range in the dataset size input into each study was observed, with the dental radiograph studies comprising 40 to 13,887 dental radiographs and the clinical images studies comprising 88 to 24,578 clinical images. Coupled with this is the variety in how the data was annotated. The included studies employed between one and twenty human annotators with and without different annotation software, 2 studies used histological analysis, and 1 study used image editing software to artificially create dental caries. The definition and assessment of dental caries exhibited differences with a 2-point scale used in 19 studies, a 3-point scale in 9 studies, a 4-point scale in 3 studies, a 5-point scale in 7 studies, a 6-point scale in 3 studies and a 7-point scale in 4 studies. Pooling studies that classified dental caries using the same number of divisions demonstrated the following mean accuracies: 2-point scale 86.1% [91% CI 81.4% - 90.8%],

Table 2. Details of the studies using artificial intelligence to detect dental caries from dental radiographs.

| Author year | Data modality | Dataset (train/ validation/ test) | Dental caries classifi-cation scale | Labelling procedure | Annot-ation tool | Exclusion criteria | Dentition | Preproc essing aug-mentation | AI platform | Results (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ahmed et al.[24] | BW | 554 (443, 55, 55) | 5 | 2 dental specialists | Y | Overlapping images, images with distortion and shadows, indirect restorations, | PER | N/A | Resnet50, ResNext101, Inception- V2 | IoU<br>F1-score | 55.1<br>53.5 |
| Amasya et al.[25] | CBCT | 500 | 2 | 3 dento-maxillofacial radiologists | N | Edentulous patients, images with exceeding artifact | PER | Cropping and resizing | U-Net | Accuracy<br>Sensitivity<br>Specificity | 92.9<br>87.4<br>95.2 |
| Ayhan et al.[26] | BW | 1170 (1000, -, 170) | 2 | 1 dental specialist | Y | Positioning/ technical errors, fixed prosthesis, braces, restorations | PER | N/A | YOLOv7 | Precision<br>F1-score | 86.6<br>84.9 |
| Azhari et al.[27] | BW | 771 | 5 | 2 examiners | N | Poor quality, identifiable images, duplicates, restorations, braces | PRI + PER | N/A | Resnet50, ResNext101, Inception- V2 | IoU | 37.7-47.8 |
| Bayrakdar et al.[28] | BW | 621 (518, 50, 53) | 2 | 2 dental specialists | Y | N/A | PER | Resizing, horizontal and vertical flipping | U-Net and VGG-16 | Sensitivity<br>Precision<br>F1-score | 84.0<br>81.0<br>84.0 |
| Bayraktar and Ayan[29] | BW | 1000 (800, -, 200) | 2 | 2 experienced dentists | Y | Restorations, 3rd permanent molars | N/A | Rotation, scaling, zooming and cropping | YOLO | Accuracy<br>Sensitivity<br>Specificity | 94.2<br>72.3<br>98.2 |
| Cantu et al.[30] | BW | 3686 (3293, 252, 141) | 6 | 4 trained dentists | N | Primary dentition, images of poor quality | PER | Resizing, random transforma-tion | U-Net | Accuracy<br>Sensitivity<br>Specificity | 80.0<br>75.0<br>83.0 |
| Chaves[31] | BW | 425 (340, 42, 42) | 3 | 2 researchers with caries experience and PhD student | Y | Image distortion or poor angulation | PER | Flipping, resizing and cropping | Mask-RCNN | Accuracy<br>Sensitivity | 75.4<br>71.6 |

(continued on next page)

Table 2 (*continued*)

| Author year | Data modality | Dataset (train/ validation/ test) | Dental caries classifi-cation scale | Labelling procedure | Annot-ation tool | Exclusion criteria | Dentition | Preproc essing aug-mentation | AI platform | Results (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Chen et al.[32] | PA | 2900 | 3 | 1 dentist | N | Primary dentition, image distortion | PER | N/A | R-CNN | IoU<br>Precision | 72.0<br>62.0 |
| Devito et al.[33] | BW | 80<br>(40, -, 40) | 5 | Optical microscope | N | Restorations | PER | N/A | CNN | ROC | 88.4 |
| Devlin et al.[34] | BW | 1446<br>(1343, 24, 103) | 2 | 6 dento-maxillofacial radiologists | N | Poor quality images, dentine caries | N/A | N/A | AssistDent | Sensitivity<br>Specificity | 75.8<br>85.4 |
| Estai et al.[35] | BW | 2468<br>(2221, -, 247) | 5 | 3 dentists | N | Restorations, primary teeth, overlapping, poor image quality, retained roots | PER | Rotations, horizontal shifts, scaling, horizontal and vertical flipping | ResNet-v2<br>Inception-v3<br>ResNet-50 | Precision<br>Specificity<br>F1-score | 87.0<br>87.7<br>85.3 |
| Frutos et al.[36] | BW | 13887<br>(4565, -, 197) | 7 | 6 dental specialists | Y | N/A | PER | Horizontal and vertical flipping | ResNet50<br>YOLOv5<br>EfficientDet | Precision<br>F1-score | 41.5<br>45.1 |
| Geetha et al.[37] | N/A | 105 | 2 | 1 dentist | N | N/A | N/A | Resizing | CNN | Accuracy | 97.1 |
| Gunec et al.[38] | OPG | 500 | 2 | 2 dental specialists | N | N/A | PER | N/A | CNN | Sensitivity<br>Specificity<br>F1-score | 90.7<br>76.0<br>78.6 |
| Imak et al.[39] | PA | 340 | 2 | 1 dentist | N | image distortion, fractures, cysts and infections | N/A | Contrast, sharpening filter | AlexNet | Accuracy | 87.6 |
| Kawazu et al.[40] | PA | 300 | 2 | Photoshop artificial caries on image | N | Dental caries, restorations | PER | Resizing and flipping | CNN | Accuracy | 78.0 |
| Kim et al.[41] | OPG | 10,000<br>(6000, 2000, 2000) | 5 | N/A | Y | N/A | N/A | Image cropping | Fast R-CNN | Precision<br>Sensitivity<br>Specificity | 63.5<br>75.5<br>95.0 |

(*continued on next page*)

Table 2 (*continued*)

| Author year | Data modality | Dataset (train/ validation/ test) | Dental caries classifi-cation scale | Labelling procedure | Annot-ation tool | Exclusion criteria | Dentition | Preproc essing aug-mentation | AI platform | Results (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Lee et al.[42] | PA | 3000 (2400, -, 600) | 3 | 4 dentists | N | Image distortion, restorations, primary teeth, disagreement in diagnosis | PER | Vertical flipping | InceptionV3 | Accuracy<br>Sensitivity<br>Specificity | 82.0<br>81.0<br>83.0 |
| Lee et al.[43] | BW | 304 (254, -, 50) | 2 | 2 postgradu-ate dental students | N | Image distortion, overlapping of proximal surfaces | PER | Flipping, rotation, scaling | U-Net | Precision<br>Recall<br>F1-score | 63.3<br>65.0<br>64.1 |
| Li et al.[44] | PA | 4129 (3829, -, 300) | 3 | 2 dentists | N | Primary teeth, restorations, pulp therapy, low quality images, | PER | Resizing, cropping | ResNet-18 | Sensitivity<br>Specificity<br>F1-score | 80.0<br>79.0<br>79.6 |
| Lian et al.[45] | OPG | 1160 (1071, -, 89) | 6 | 3 dentists | N | Primary teeth, blurred images | PER | Cropping | nnU-net | Accuracy<br>Sensitivity<br>Specificity<br>F1-score | 98.6<br>82.1<br>100<br>90.2 |
| Majanga and Viriri[46] | N/A | 120 | 2 | Hessian analysis (blob detection) | N | N/A | N/A | Rotation, scaling, resizing | U-net | Accuracy | 98.0 |
| Mao et al.[47] | BW | 278 (195, -, 83) | 2 | 3 dentists | N | N/A | N/A | Cropping, horizontal and vertical translations | AlexNet, GoogleNet, Vgg19 ResNet50 | Accuracy | 80.3 -90.3 |
| Moran et al.[48] | BW | 112 | 3 | 1 dental specialist | N | Dental implants, malocclusion | N/A | Rotation, flipping and cropping | ResNet and Inception at different learning rates | Accuracy | 0-100 |
| Oztekin et al.[49] | PA | 13870 (13270, -, 600) | 2 | 1 dental specialist | N | N/A | PER | Cropping, rotation, zoom, flipping, shifting | EfficientNet-B0 DenseNet-121 ResNet-50 | Accuracy<br>Sensitivity<br>Specificity<br>F1-score | 91.3<br>85.9<br>96.7<br>90.8 |

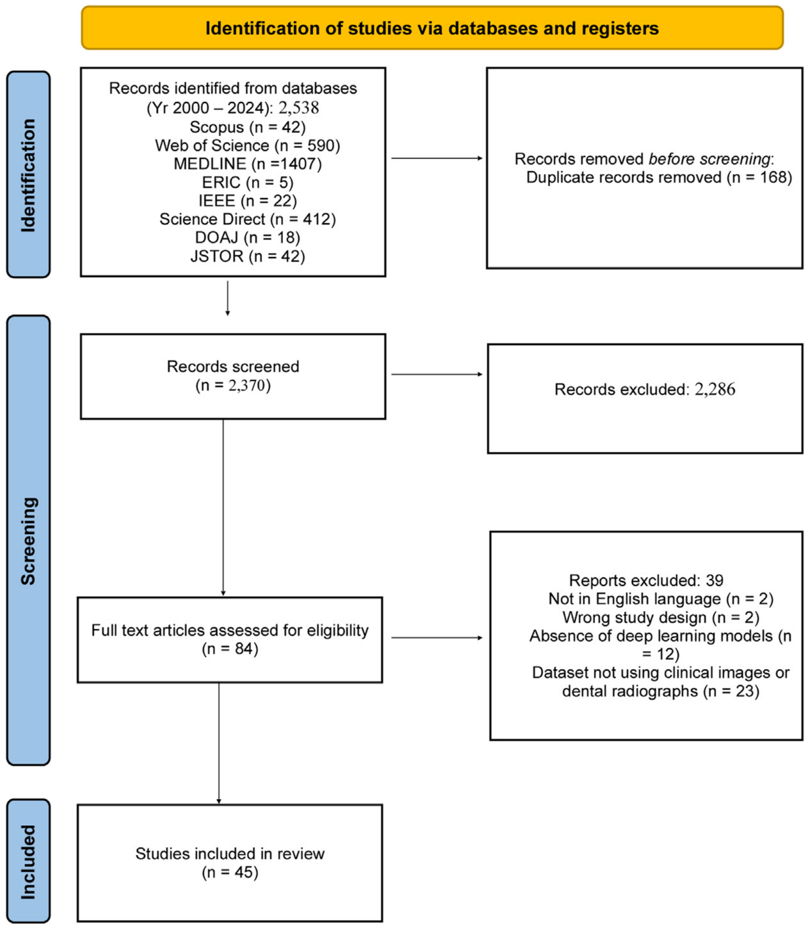
(*continued on next page*)

Table 2 (*continued*)

| Author year | Data modality | Dataset (train/ validation/ test) | Dental caries classifi-cation scale | Labelling procedure | Annot-ation tool | Exclusion criteria | Dentition | Preproc essing aug-mentation | AI platform | Results (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Panyarak et al.[50] | BW | 994 | 7 | 3 oral and maxillofacial radiologists | Y | severe interproximal overlap | PER | Horizontal flip, random transforma-tion, transla-tions, scaling | YOLOv3 | Precision<br>F1-score | 32-77<br>27-75 |
| Panyarak et al.[51] | BW | 2758 | 7 | 3 oral and maxillofacial radiologists | Y | Caries free, attrition, poor image quality, severe interproximal overlap | PER | Random movements in vertical and horizontal directions, rotation | ResNet-18<br>ResNet-50<br>ResNet-101<br>ResNet-152 | Accuracy<br>Sensitivity<br>Specificity | 46-71<br>68-80<br>16-61 |
| Pun[52] | BW | 190 | 5 | 1 experienced dentist | Y | Caries free | N/A | Image en-hancement and cropping | EfficientDet-Lite1 | Precision<br>Sensitivity<br>F1-score | 70.6<br>62.5<br>66.1 |
| Qayyum et al.[53] | PA | 229<br>(90%, -, 10%) | 2 | 1 dental specialist | Y | N/A | N/A | Horizontal and flip, rotation, cropping | ResNet-50<br>ResNet-101<br>MobileNet-V3 | Accuracy | 97.0 |
| Valizadeh et al.[54] | N/A | 183 | 2 | Histological analysis | N | Fractures, dental anomalies, gross caries | N/A | Cropping, rotations | FCM (fuzzy c-means) | Accuracy | 61-97 |
| Ying et al.[55] | N/A | 40<br>(90%, -, 10%) | 2 | 2 dental specialists | N | Developmental defects, malocclusion, irregular dentition, prosthesis | PER | Flipping, rotation, translation, | U-net | Precision<br>Sensitivity<br>Specificity | 74.0<br>94.0<br>92.0 |
| Zhou et al.[56] | OPG | 210 | 2 | 2 dentists | Y | N/A | N/A | Cropping | CNN | Accuracy<br>F1-score | 82.7<br>86.5 |

Abbreviations: **BW**, bitewing radiograph; **PA**, periapical radiograph; **OPG**, panoramic radiograph; **CBCT**, cone beam computed tomography; **PER**, permanent dentition; **PRI**, primary dentition; **N/A,** not applicable.

Figure 1. PRISMA flow diagram illustrating the search strategy.




3-point scale 76.9% [95% CI (66.8% - 87.1%)], 4-point scale 82.6% [95% CI (43.4% - 100%)], 5-point scale 79.0% [95% CI 56.0% - 100%)], 6-point scale 79.4% [95% CI (30.9% -100%)], 7-point scale 60.3% [95% CI (30.1% - 90.4%)].

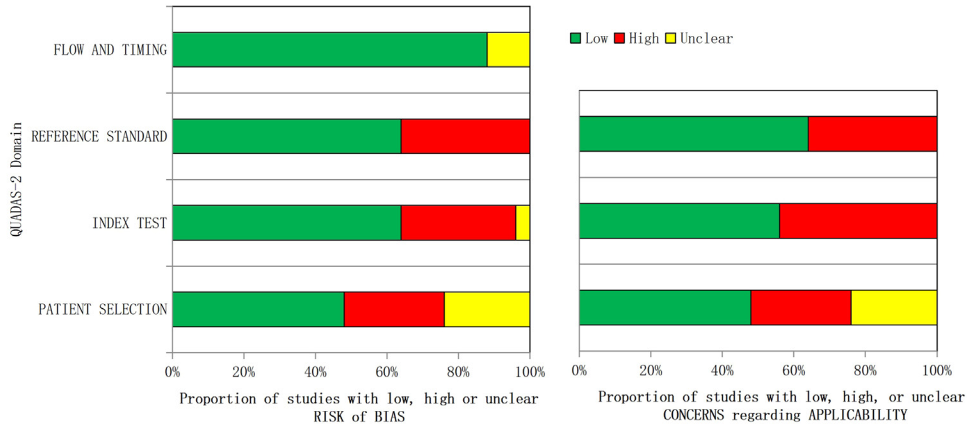
## Variation in the Inclusion and Exclusion Criteria across included studies

Three studies (7%) included primary and permanent dentitions, 15 studies (33%) did not disclose the assessed dentition, and 27 studies (60%) only included the permanent dentition. In addition, 20 studies (44%) excluded images of poor diagnostic quality, 12 studies (27%) excluded teeth with restorations, 6 studies (13%) excluded teeth undergoing fixed orthodontics, 4 studies (9%) excluded teeth with developmental defects of enamel, 2 studies (4%) excluded teeth with noncarious tooth structure loss and 1study excluded teeth that were caries-free.

## AI Platforms

The accuracy of the platforms was established by assessing it against the test dataset. Eighteen of the studies (40%) did not report on the size of the test dataset, with the remaining 27 studies using an average of 14.8%, [95% CI (10.8% - 18.9%)] of the total available data. Accuracy, sensitivity, specificity, and F1-scores had the following respective ranges: 41.5% to 98.6%, 18.5% to 94%, 38.5% to 100%, and 45.1% to 92.2%. The mean accuracy of correctly diagnosing dental caries was 78.2% [95% CI (72.0% - 84.4%)] for clinical image studies and 81.5% [95% CI (72.7% - 90.3%)] for dental radiograph studies. Of the included studies, 21 different AI platforms were used, including ResNet-50 in 11 studies, U-Net in 8 studies, YOLO in 7 studies, Inception-V3 and R-CNN in 6 studies and ResNet101 in 5 studies. The accuracy ranges of these platforms were observed at 41% to 97% for ResNet-50, 74% to 99% for U-Net, 42% to 89% for YOLO, 43% to 87% for Inception-V3 and 60% to 87% for R-CNN. The various AI platforms used for dental caries detection using dental ra-

Figure 2. Risk of bias assessment using quality assessment tool for diagnostic accuracy studies (QUADAS-2).




diographs and clinical images are tabulated in Table 2 and 3 .

## RoB assessment

ROB demonstrated that the index test domain showed the highest degree of bias, with 16 studies (36%) categorised with a high ROB Figure 2. Nine studies (20%) were found to have a low ROB of all 4 domains Table 4 and 5 , the performance of these 9 studies with low ROB demonstrated a mean accuracy of 86.6% [95% CI (75.8% - 97.5%)], sensitivity of 70.4 [95% CI (52.4% - 88.5%)], and specificity of 89.2% [95% CI (83.4% - 95.0%)].

## CLAIM checklist

The findings from applying the CLAIM checklist to the 45 studies include 84% of the studies not calculating sample size, 76% of the studies not measuring the inter- and intrarater variability, 27% of studies not commenting on their studies' limitations and none of the studies validating their model on external data Supplementary File 1.

## Quantitative Analysis

Due to limited data reporting, the meta-analysis included only 7 studies: 5 using dental radiographs and 2 using clinical images. The overall summary estimates for AI in dental caries detection showed a sensitivity of 76% [95% CI (65% - 85%)] and a specificity of 91% [95% CI (86% - 95%)], as seen in Figure 3. The HSROC curve generated using the bivariate hierarchical model indicated an AUC of 92% [95% CI (89% - 94%)], shown in Figure 4. Heterogeneity, assessed by Higgins' $I^2$, was 99.1% for sensitivity and 98% for specificity, demonstrating high heterogeneity across the studies.

## Subgroup Analysis

For the 5 studies that used dental radiographs, the summary estimates showed a sensitivity of 73% [95% CI (59% - 84%)] and specificity of 91% [95% CI (82% - 96%)] for AI in detecting dental caries Supplementary Figure 5. The AUC for these studies was 90% [95% CI (87% - 92%)] Supplementary Figure 6. Heterogeneity was 99.5% for sensitivity and 99% for specificity, indicating significant heterogeneity.

For the 2 studies that used clinical images, the summary estimates for AI in detecting dental caries showed a sensitivity of 83% [95% CI (74% - 90%)] and a specificity of 92% [95% CI (88% - 95%)] Supplementary Figure 7. The AUC for these studies was 95% [95% CI (92% - 96%)] Supplementary Figure 8. Heterogeneity was 89.3% for sensitivity and 65% for specificity, indicating high heterogeneity for sensitivity but lower heterogeneity for specificity.

# DISCUSSION

This systematic review and meta-analysis aimed to examine the different AI platforms and deep learning methods used for dental caries detection, using both clinical images and radiographs, and identify areas for future research. Following the appraisal and synthesis of the included studies, this systematic review and meta-analysis established the sensitivity and specificity of AI detecting dental caries was 76% and 91% respectively. However, these figures fail to identify the significant limitations of current research using AI to detect dental caries, which requires further discussion.

First, standardisation for dental caries' identification, classification and annotation is essential. This systematic review found that the included studies identified dental caries using up to 20 human experts and classified dental caries using a 2 to 7-point scale. As a result, each of the 45 platforms is

Table 3. Details of the studies using artificial intelligence to detect dental caries from clinical images.

| Author year | Data modality | Dataset (train/ validation /test) | Dental caries classification scale | Labelling procedure | Annotation tool | Exclusion criteria | Dentition | Pre-processing augmentation | AI platform | Results (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Askar et al.[57] | DSLR | 434 | 3 | 2 dental specialists | Y | Fixed orthodontics | N/A | Cropping and resizing of images | SqueezeNet | Accuracy<br>Sensitivity<br>Specificity<br>F1-score | 88.0<br>58.0<br>85.0<br>68.0 |
| Ding et al.[58] | Mobile phone | 640 (570, -, 70) | 2 | N/A | Y | Fixed orthodontics, restricted mouth opening | PER | Rotations, colour changes | YOLOv3 | Precision<br>F1-score | 77-100<br>50-80 |
| Felsch et al.[59] | DSLR | 18179 | 5 | 1 experienced dentist | Y | Restorations, orthodontic appliances, rare dental diseases | PRI and PER | Cropped, rotate, resized, randomly distorted | SegFormer-B5 | Accuracy<br>Sensitivity<br>Specificity | 90-99<br>37-89<br>91-99 |
| Kim et al.[60] | Intraoral camera | 610 (410, 90, 90) | 3 | 2 dentists | N | Low resolution photos | N/A | Resized, contrast enhancement | ResNext | Precision<br>F1-score | 83.5<br>84.3 |
| Kuhnisch et al.[61] | DSLR | 2417 (1891, 47, 479) | 3 | 1 dentist | N | Non-carious hard tissue defects, restorations | PER | Resizing, rotations, cropping | MobileNetV2 | Accuracy<br>Sensitivity<br>Specificity | 92.5<br>89.6<br>94.3 |
| Mehdizadeh et al.[62] | N/A | 1020 | 7 | 1 dentist | Y | N/A | PRI and PER | Rotation, height and width shifts, scaling, horizontal and vertical flipping | VGG-16<br>ResNet-50<br>Inception-v3 | Accuracy<br>Specificity<br>F1-score | 47-79<br>81-100<br>66-83 |

(*continued on next page*)

Table 3 (*continued*)

| Author year | Data modality | Dataset (train/validation/test) | Dental caries classification scale | Labelling procedure | Annotation tool | Exclusion criteria | Dentition | Pre-processing augmentation | AI platform | Results (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Moutselos et al.[63] | Intraoral Camera | 88 (79, -, 9) | 6 | 2 dentists | Y | Restorations, teeth with hypoplastic or hypomineralised areas | PER | Horizontal and vertical flipping, rotations, shear, scaling | Mask R-CNN | Accuracy | 59.6 |
| Park et al.[64] | Intraoral camera | 2348 (1638, 410, 300) | 4 | 1 dentist | Y | Blurred, duplicate or unintended images | PER | Shifting, blurring, image symmetry | U-Net<br>ResNet-18<br>Faster R-CNN | Precision<br>Sensitivity | 87.4<br>89.0 |
| Tareq et al.[65] | Mobile phone | 233 (107, 28, 33) | 3 | 3 dentists | Y | Molar teeth | PER | Blurring, focus changes, light filters, angulation changes | ResNet50<br>ResNet101<br>VGG16<br>AlexNet<br>DenseNet<br>YOLO | Accuracy | 78.9 |
| Thanh et al.[66] | Mobile Phone | 2652 (1902, -, 750) | 4 | 1 dentist | N/A | Restorations, enamel defects | N/A | Flipping and rotation | YOLOv3<br>Faster R-CNN<br>RetinaNet<br>SSD | Accuracy<br>Sensitivity<br>Specificity | 61-69<br>0-37<br>71-100 |
| Yoon et al.[67] | DSLR | 24578 | 4 | 20 data labellers | Y | Mirror scratches, obstructed images, primary teeth, duplicate images, out-of-focus images | PER | Resizing, flipping, random shifting, photometric distortions | R-CNN | Accuracy<br>Sensitivity<br>Specificity | 93-98<br>71-78<br>96-99 |
| Zhang et al.[68] | Mobile Phone | 3932 (2507, 300, 1125) | 2 | 3 dentists | N | N/A | PER | Shifting, rotation, hue/chroma/exposure changes | ConvNet | Sensitivity<br>Specificity | 68.7<br>81.9 |

Abbreviations: **DSLR,** Digital Single Lens Reflex Camera; **PER**, permanent dentition; **PRI**, primary dentition; **N/A,** not applicable.

Table 4. QUADAS 2 assessment of included studies using artificial intelligence to detect dental caries from dental radiographs.

| Study | risk of bias | | | | applicability concerns | | |
|---|---|---|---|---|---|---|---|
| | Patient selection | Index test | Reference standard | flow and timing | patient selection | index test | reference standard |
| Ahmed et al.[24] | Unknown | Low | Low | Low | Unknown | Low | Low |
| Amasya et al.[25] | Low | Low | Low | Low | Low | Low | Low |
| Ayhan et al.[26] | Low | Low | High | Low | Low | Low | High |
| Azhari et al.[27] | Low | Low | Unknown | Low | Low | Low | Low |
| Bayrakdar et al.[28] | Low | Low | High | Low | Unknown | High | High |
| Bayraktar and Ayan[29] | Low | Low | High | Unknown | Low | High | High |
| Cantu et al.[30] | Low | Low | Low | Low | Low | Low | Low |
| Chaves et al.[31] | Unknown | Low | Unknown | Low | Unknown | Low | High |
| Chen et al.[32] | Unknown | High | Low | Low | Unknown | High | Low |
| Devito et al.[33] | High | High | High | Low | High | High | High |
| Devlin et al.[34] | Low | Low | Low | Low | Low | Low | Low |
| Estai et al.[35] | Unknown | Low | Low | Low | Unknown | Low | Low |
| Frutos et al.[36] | Unknown | High | Low | Low | Low | High | Low |
| Geetha et al.[37] | High | High | High | Low | High | High | High |
| Gunec et al.[38] | Unknown | High | High | Low | Unknown | Unknown | High |
| Imak et al.[39] | Unknown | High | High | Low | Unknown | High | High |

(*continued on next page*)

Table 4 (*continued*)

| Study | risk of bias | | | | applicability concerns | | |
|---|---|---|---|---|---|---|---|
| | Patient selection | Index test | Reference standard | flow and timing | patient selection | index test | reference standard |
| Kawazu et al.[40] | High | Low | High | Low | High | Low | High |
| Kim et al.[41] | Low | High | Unknown | Low | Low | High | Unknown |
| Lee et al.[42] | Low | Low | Low | Low | Low | Low | Low |
| Lee et al.[43] | High | Low | Low | Low | High | Low | Low |
| Li et al.[44] | Low | Low | Low | Low | Low | Low | Low |
| Lian et al.[45] | Low | Low | Low | Low | Low | Low | Low |
| Majanga and Viriri[46] | High | High | High | Unknown | High | High | High |
| Mao et al.[47] | Low | Low | High | Low | Low | High | High |
| Moran et al.[48] | Unknown | High | Low | Low | Unknown | High | Low |
| Oztekin et al.[49] | Unknown | High | High | Low | Unknown | High | High |
| Panyarak et al.[50] | Low | Low | Low | Low | Low | Unknown | Low |
| Panyarak et al.[51] | Low | Low | Low | Low | Low | Unknown | Low |
| Pun[52] | High | High | Low | Low | High | High | Unknown |
| Qayyum et al.[53] | High | High | Unknown | Unknown | High | High | Unknown |
| Valizadeh et al.[54] | High | Low | Low | Low | High | Low | Low |
| Ying et al.[55] | High | Unknown | Low | Low | High | High | Low |
| Zhou et al.[56] | Unknown | High | High | Unknown | High | High | High |

Table 5. QUADAS 2 assessment of included studies using artificial intelligence to detect dental caries from clinical images.

| Study | risk of bias | | | | applicability concerns | | |
|---|---|---|---|---|---|---|---|
| | Patient selection | Index test | Reference standard | flow and timing | patient selection | index test | reference standard |
| Askar et al.[57] | Unknown | Low | Low | Low | Low | Low | Low |
| Ding et al.[58] | Low | High | High | Low | Low | High | High |
| Felsch et al.[59] | High | Unknown | Low | Low | High | unknown | Low |
| Kim et al.[60] | Unknown | Low | Low | Low | Unknown | Low | Low |
| Kuhnisch et al.[61] | Low | Low | Low | Low | Low | Low | Low |
| Mehdizadeh et al.[62] | High | High | Low | Low | High | High | Low |
| Moutselos et al.[63] | High | Low | Low | Unknown | High | Low | Low |
| Park et al.[64] | Low | Unknown | Low | Low | Unknown | Unknown | Low |
| Tareq et al.[65] | Low | Low | Low | Low | High | Low | Low |
| Thanh et al.[66] | Low | Low | Low | Low | Low | Low | Low |
| Yoon et al.[67] | Low | Low | Low | Low | Low | Low | Low |
| Zhang et al.[68] | Low | High | High | Low | Low | High | High |

Figure 3. Forest plot for sensitivity and specificity.

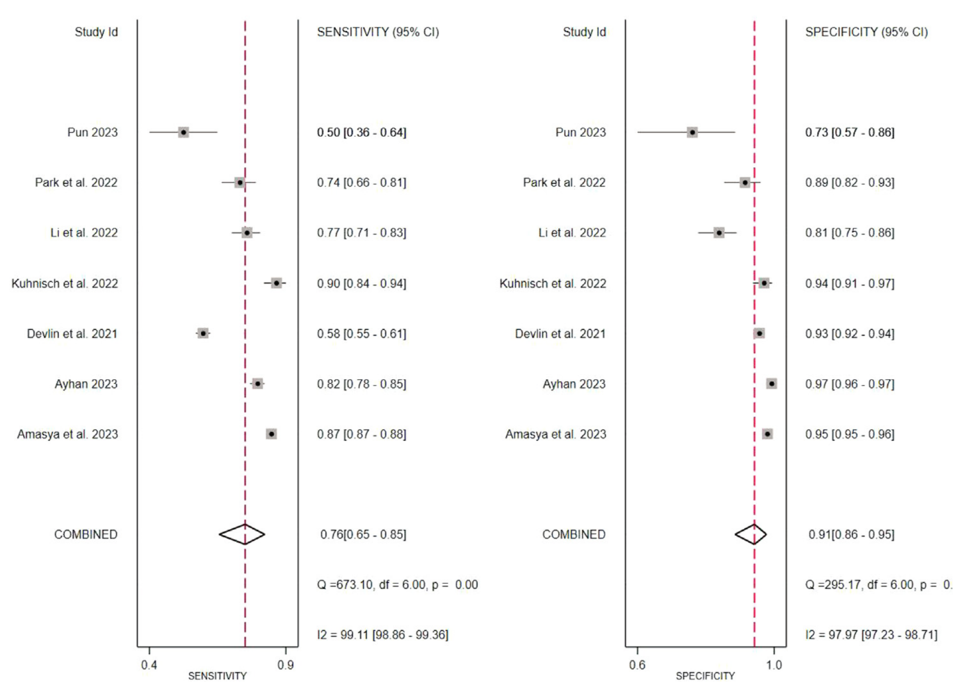
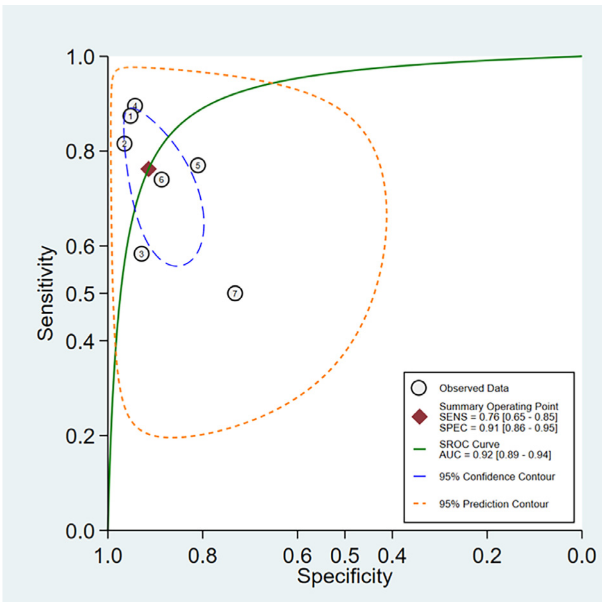


Figure 4. Summary receiver operating characteristic curve (sROC) for predictive studies.




trained, validated, and tested to assess dental caries at different locations, from an initial "enamel" carious lesion to a cavitated "dentine" carious lesion encroaching the pulp. In addition, each platform's development and assessment reflect the quality and standardisation of the initial annotations. In assessing this, the QUADAS-2 and CLAIM checklists found that of the included studies, 31% exhibited a high RoB in their reference standard domain, and 76% did not report on the calibration of their annotations. Therefore, the accuracy of these annotations remains to be determined. This uncertainty is further highlighted through Kawazu and colleagues using imaging processing software to artificially create dental caries onto sound teeth,[40] and multiple authors simplifying standardised dental caries classification systems such as the International Caries Detection and Assessment System (ICDAS) from a 7-point scale to a 3-point scale.[65,66] Through changing dental caries classification systems and artificially creating dental caries, AI platforms restrict their ability to be replicated and compared against 1 another, as well as the generalisability of an AI platform on external datasets.

Second, the 45 included studies highlight significant variations in their overall dataset sizes ranging from 40 to 24,578 images. This variation impacts the reliability and validity of each platform in detecting dental caries and limits its generalisability. Coupled with this is the size of the test dataset, which, on average, represented 14.8% of the overall dataset, with 40% of the studies failing to report this. As a result, the variation in dataset size between the included articles introduces further heterogeneity and makes further analysis less meaningful.

Third, this systematic review found variations in inclusion and exclusion criteria, with 27% of the included studies excluding teeth with restorations and 9% excluding teeth with developmental defects of enamel (DDEs). In having too broad exclusion criteria, AI platforms risk making themselves not applicable to a wide range of patients, demonstrated through the prevalence of DDEs estimated at 33%,[69] and 77% of Australian adults having at least 1 dental restoration.[70] However, as dental caries is associated with DDEs with an odds ratio of 2.21,[71] and an estimated 3.6% of restorations having secondary caries,[72] there are clear benefits for future AI platforms attempting to overcome these limitations.

Several AI platforms are currently used to create the algorithms underpinning each deep learning framework. This systematic review found 21 different AI platforms from the 45 included studies with similar accuracy ranges observed in the studies using ResNet-50, U-Net, YOLO Inception-V3 and R-CNN. Such similarities in the different platforms are also displayed within studies, as Tareq and colleagues used the same dataset on 5 different AI platforms, including VGG16, ResNet-50, ResNet-101, AlexNet, and DenseNet121 to achieve accuracies between 78.3% and 87.0%.[65] This comparison between AI platforms is even more complex after considering the learning rate used. This learning rate is a numerical size of the updated hyperparameters.[1] Authors need to balance using enough data for the network to produce meaningful results while keeping this manageable to ensure the process can be completed quickly and without excessive memory requirements.[73] Moran and colleagues assessed the performance of 0.1, 0.001 and 0.001 learning rates to conclude that 0.001 was the most accurate.[48] However due to the lack of standardization in the learning rates of each of the 45 included studies, further complexity is created that makes comparisons between studies more difficult.

1 growing observation from the included studies is using the AI platform using tooth surface segmentation to improve its accuracy. Park and colleagues investigated this process to demonstrate an improvement in precision from 75% to 87% following tooth surface segmentation.[64] It is believed that by removing the background noise and blur, AI platforms can focus on the tooth surface and, therefore, arrive at a more accurate diagnosis. However, this has yet to be assessed comprehensively.

While an in-depth analysis of the architecture of each of these platforms is beyond the scope of this systematic review, future research should examine these variations to better understand how algorithms can enhance the resolution and, ultimately, the accuracy of their outputs. This research is needed now more than ever, as He and colleagues report that increasing layers in a convolutional neural network (CNN) resulted in over-saturation and the degradation of the platform's performance.[74] In addition, due to the black box problem, whereby AI cannot explain how it arrived at its conclusion, little is known about the benefits or potential harms of each AI platform's architecture.

Pooling the performance of studies using a low risk of bias or different dental caries classification scales did not demonstrate a statistical difference from the overall pooled accuracy ranges. An explanation for this could include the limited number of included studies using 4, 6 and 7-point dental caries classification scales, thereby increasing the level of uncertainty in the 95% confidence interval, the additional complexity for the AI platform to identify subtle changes in enamel carious lesions, and the heterogeneity of the datasets whereby multiple variables are contributing to the outcomes. These variables include the type of AI architecture, dataset size, use and application of augmented data, and the robustness of the dental caries annotations. Consequently, the current published data using AI to detect dental caries cannot isolate individual factors for comparison between studies, and the only method for comparison is to test each AI platform against a standardised and universally accepted external test dataset.

With high levels of heterogeneity only 7 studies,[25,26,34,44,52,61,64] were found suitable for meta-analysis. Overall, good sensitivity and specificity were noted across 5 studies.[25,26,34,61,64] This could be due to many of these studies using a large sample size. However, analysis by Amasya and coworkers demonstrated high sensitivity and specificity irrespective of the small sample size used (n=500) and absence of an annotation tool.[25] This study did however use a dichotomous caries classification and a rich dataset consisting of CBCTs. Conversely, Li and colleagues exhibited poor sROC scores despite its larger sample size (n = 4129). This may be attributed to using periapical radiographs as a data modality, not using an annotation tool and ResNet-18 as an AI platform.[44] Another study by Pun and colleagues showed a wide discrepancy in sensitivity (77%) and specificity (33%) values in the sROC plot. Potential reasons include a small sample size (n = 190), caries classification using 5 segmentation categories, labelling procedure done by 1 experienced dentist, poor image quality and Efficient DetLite 1 as an AI platform.[52]

AI for dental caries detection brings several other potential implications. One key advantage is its ability to process and analysis large volumes of data rapidly, which can enhance diagnostic efficiency in busy clinical settings. AI tools can also standardize diagnostic practices by reducing human error and variability among practitioners, potentially leading to more consistent outcomes.[2] However, there are notable disadvantages as well. AI systems are often trained on specific datasets that may not fully represent the diverse population seen in real-world dental practices, potentially leading to biased or less accurate outcomes in certain patient groups.[15] Additionally, the integration of AI into dental practices may require significant investments in modern technologies and ongoing staff training, which could present financial and logistical challenges for clinics.[3] As AI continues to evolve, it is essential for dental professionals to balance the potential benefits with an understanding of AI's current limitations and the need for human oversight in the diagnostic process.

This systematic review and meta-analysis demonstrates a robust and scientifically sound approach to evaluating AI's accuracy in detecting dental caries. The reported accuracy aligns with previous systematic review by Mohammad-Rahimi and colleagues (71% - 96%),[3] Revilla-Leon and colleagues (76% - 88%),[14] Talpur and colleagues (69% - 97%),[13] and Rokhshad and colleagues at (60% - 98%).[75] This review however, stands out by pooling data from 45 included studies published up to March 2024 and focuses on key methodological factors influencing AI performance in dental caries detection. A thorough search across 8 major databases ensured a comprehensive coverage of relevant literature, and the use of the CLAIM checklist ensured transparency and consistency in reporting. By incorporating studies utilizing both dental radiographs and clinical images, this review offers a deeper insight into the impact of data modality on AI accuracy. While this meta-analysis was restricted to eligible studies, the inclusion of subgroup analyses on pooled sensitivity and specificity further strengthens the findings. One acknowledged limitation is the exclusion of non-English studies, which may affect the generalizability of the conclusions. Nonetheless, this review contributes significantly to the growing evidence on the role of AI in dental diagnostics.

Future studies using AI to detect dental caries need to learn from the limitations of existing research to validate their performance, utilise funding most appropriately and ultimately improve patient outcomes. This includes ensuring future studies meet minimum reporting standards, such as the STARD-AI protocol,[76] CLAIM,[20] and the 2021 checklist created by Schwendicke and co-workers.[77] Also, it is essential to address data protection for each platform, ensuring that clinical images can be deidentified to meet legal standards like the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union. After this, a standardised test dataset should serve as a gold standard for evaluating future AI platforms. This dataset should ideally include high-quality, diverse dental radiographs and clinical images that have been calibrated and annotated according to standardised, reproducible dental caries classification scales, with input from multiple international centres.

## CONCLUSION

Significant variability exists in AI performance for detecting dental caries, with reported accuracy ranging from 41.5% to 98.6% across different AI platforms. Meta-analysis indicates a mean sensitivity of 76% and specificity of 91% for AI-based caries detection, with subgroup analysis showing marginally higher sensitivity for clinical images than dental radiographs. Although AI is promising for dental caries detection, further refinement is necessary to achieve consistent and reliable performance across varying imaging modalities.

### Clinical Significance

AI holds significant potential in dental caries detection, but its performance remains variable, with accuracy rates spanning from 41.5% to 98.6%. Due to this inconsistency, current AI tools are not yet reliable enough to replace traditional diagnostic methods. Instead, they should be considered as valuable supplementary aids in clinical practice. To fully harness the impact of AI in dental diagnostics, further refinement and rigorous validation are essential to ensure consistent and dependable applications in real-world clinical settings.

## ACKNOWLEDGMENTS

Not applicable.

## ETHICAL STATEMENT

This is a systematic review. Hence, ethical approval was not required.

## CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

**LYNDON P ABBOTT:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **ANKITA SAIKIA:** Writing – review & editing, Validation, Data curation. **ROBERT P ANTHONAPPA:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

## SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebdp.2024.102077.

## REFERENCES

1. Longley D, Shain M. *Data and Computer Security: a Dictionary of Terms and Concepts*. York: Springer; 1989.
2. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res*. 2020;99(7):769–774. doi:10.1177/0022034520915714.
3. Mohammad-Rahimi H, Motamedian SR, Rohban MH, et al. Deep learning for caries detection: a systematic review. *J Dent*. 2022;122:104115. doi:10.1016/j.jdent.2022.104115.
4. Bejnordi BE, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318(22):2199–2210. doi:10.1001/jama.2017.14585.
5. Haenssle HA, Fink C, Schneiderbauer R, 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;29(8):1836–1842. doi:10.1093/annonc/mdy166.
6. Rajpurkar P, Irvin J, Zhu K, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. https://arxiv.org/abs/1711.05225. Accessed on 15/02/2024.
7. Bilgir E, Bayrakdar İŞ, Çelik Ö, et al. An artifıcial ıntelligence approach to automatic tooth detection and numbering in panoramic radiographs. *BMC Med Imaging*. 2021;21(1):124. doi:10.1186/s12880-021-00656-7.
8. Kuwada C, Ariji Y, Fukuda M, et al. Deep learning systems for detecting and classifying the presence of impacted supernumerary teeth in the maxillary incisor region on panoramic radiographs. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2020;130(4):464–469. doi:10.1016/j.oooo.2020.04.813.
9. Ekert T, Krois J, Meinhold L, et al. Deep learning for the radiographic detection of apical lesions. *J Endod*. 2019;45(7):917–922.e5. doi:10.1016/j.joen.2019.03.016.
10. Wen PYF, Chen MX, Zhong YJ, Dong QQ, Wong HM. Global burden and inequality of dental caries, 1990 to 2019. *J Dent Res*. 2022;101(4):392–399. doi:10.1177/00220345211056247.
11. Do LG, Spencer J. Oral Health of Australian Children: the National Child Oral Health Study 2012–14. (Do LG, Spencer AJ, eds.). University of Adelaide Press; 2016. https://www.adelaide.edu.au/press/ua/media/631/ncohs-ebook.pdf. Accessed on: 03/01/2024.
12. Alabdullah JH, Daniel SJ. A systematic review on the validity of teledentistry. *Telemedicine and e-Health*. 2018;24(8):639–648. doi:10.1089/tmj.2017.0132.
13. Talpur S, Azim F, Rashid M, Syed SA, Talpur BA, Khan SJ. Uses of different machine learning algorithms for diagnosis of dental caries. *J Healthc Eng*. 2022;2022:1–13. doi:10.1155/2022/5032435.
14. Revilla-León M, Gómez-Polo M, Vyas S, et al. Artificial intelligence applications in restorative dentistry: a systematic review. *J Prosthet Dent*. 2022;128(5):867–875. doi:10.1016/j.prosdent.2021.02.010.
15. Khanagar SB, Alfouzan K, Awawdeh M, Alkadi L, Albalawi F, Alfadley A. Application and performance of artificial intelligence technology in detection, diagnosis and prediction of dental caries (DC)—a systematic review. *Diagnostics*. 2022;12(5):1083. doi:10.3390/diagnostics12051083.

16. Albano D, Galiano V, Basile M, et al. Artificial intelligence for radiographic imaging detection of caries lesions: a systematic review. *BMC Oral Health*. 2024;24(1):274. doi:10.1186/s12903-024-04046-7.

17. McInnes MDF, Moher D, Thombs BD, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies. *JAMA*. 2018;319(4):388. doi:10.1001/jama.2017.19163.

18. Mejàre I, Källestål C, Stenlund H. Incidence and progression of approximal caries from 11 to 22 years of age in Sweden: a prospective radiographic study. *Caries Res*. 1999;33(2):93–100. doi:10.1159/000016502.

19. Pitts N, Ekstrand K. International Caries Detection and Assessment System (ICDAS) and its International Caries Classification and Management System (ICCMS) – methods for staging of the caries process and enabling dentists to manage caries. *Commun Dent Oral Epidemiol*. 2013;41(1):1–84. doi:10.1111/cdoe.12025.

20. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2(2):e200029. doi:10.1148/ryai.2020200029.

21. Whiting PF. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529. doi:10.7326/0003-4819-155-8-201110180-00009.

22. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20(19):2865–2884. doi:10.1002/sim.942.

23. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–560. doi:10.1136/bmj.327.7414.557.

24. Ahmed WM, Azhari AA, Fawaz KA, et al. Artificial intelligence in the detection and classification of dental caries. *J Prosthet Dent.*. 2023:1–7. doi:10.1016/j.prosdent.2023.07.013.

25. Amasya H, Alkhader M, Serindere G, et al. Evaluation of a decision support system developed with deep learning approach for detecting dental caries with cone-beam computed tomography imaging. *Diagnostics*. 2023;13(22):3471. doi:10.3390/diagnostics13223471.

26. 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. doi:10.1007/s00784-024-05566-w.

27. Azhari AA, Helal N, Sabri LM, Abduljawad A. Artificial intelligence (AI) in restorative dentistry: performance of AI models designed for detection of interproximal carious lesions on primary and permanent dentition. *Digit Health*. 2023;9. doi:10.1177/20552076231216681.

28. Bayrakdar IS, Orhan K, Akarsu S, et al. Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. *Oral Radiol*. 2021;38(4):468–479. doi:10.1007/s11282-021-00577-9.

29. 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. doi:10.1007/s00784-021-04040-1.

30. Cantu AG, Gehrung S, Krois J, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent*. 2020;100:103425. doi:10.1016/j.jdent.2020.103425.

31. Chaves ET, Vinayahalingam S, van Nistelrooij N, et al. Detection of caries around restorations on bitewings using deep learning. *J Dent*. 2024;143:104886. doi:10.1016/j.jdent.2024.104886.

32. Chen H, Li H, Zhao Y, Zhao J, Wang Y. Dental disease detection on periapical radiographs based on deep convolutional neural networks. *Int J Comput Assist Radiol Surg*. 2021;16(4):649–661. doi:10.1007/s11548-021-02319-y.

33. Devito KL, Barbosa FD, Felippe WN. An artificial multilayer perceptron neural network for diagnosis of proximal dental caries. *Oral Surg Oral Med Oral Pathol Oral Radiol Endodontol*. 2008;106(6):879–884. doi:10.1016/j.tripleo.2008.03.002.

34. Devlin H, Williams T, Graham J, Ashley M. The ADEPT study: a comparative study of dentists' ability to detect enamel-only proximal caries in bitewing radiographs with and without the use of AssistDent artificial intelligence software. *Br Dent J*. 2021;231(8):481–485. doi:10.1038/s41415-021-3526-6.

35. Estai M, Tennant M, Gebauer D, et al. Evaluation of a deep learning system for automatic detection of proximal surface dental caries on bitewing radiographs. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2022;134(2):262–270. doi:10.1016/j.oooo.2022.03.008.

36. Pérez de Frutos J, Holden Helland R, Desai S, et al. AI-Dentify: deep learning for proximal caries detection on bitewing x-ray - HUNT4 Oral Health Study. *BMC Oral Health*. 2024;24(1):344. doi:10.1186/s12903-024-04120-0.

37. Geetha V, Aprameya KS, Hinduja DM. Dental caries diagnosis in digital radiographs using back-propagation neural network. *Health Inf Sci Syst*. 2020;8(1):8. doi:10.1007/s13755-019-0096-y.

38. Güneç HG, Ürkmez EŞ, Danaci A, Dilmaç E, Onay HH, Aydin KC. Comparison of artificial intelligence vs. junior dentists' diagnostic performance based on caries and periapical infection detection on panoramic images. *Quant Imaging Med Surg*. 2023;13(11):7494–7503. doi:10.21037/qims-23-762.

39. Imak A, Celebi A, Siddique K, Turkoglu M, Sengur A, Salam I. Dental caries detection using score-based multi-input deep convolutional neural network. *Ieee Access*. 2022;10:18320–18329. doi:10.1109/ACCESS.2022.3150358.

40. Kawazu T, Takeshita Y, Fujikura M, Okada S, Hisatomi M, Asaumi J. Preliminary study of dental caries detection by deep neural network applying domain-specific transfer learning. *J Med Biol Eng*. 2024;44(1):43–48. doi:10.1007/s40846-024-00848-w.

41. Kim C, Jeong H, Park W, Kim D. Tooth-related disease detection system based on panoramic images and optimization

through automation: development study. *JMIR Med Inform*. 2022;10(10):e38640. doi:10.2196/38640.

42. 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;77:106–111. doi:10.1016/j.jdent.2018.07.015.

43. 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):8. doi:10.1038/s41598-021-96368-7.

44. Li S, Liu J, Zhou Z, et al. Artificial intelligence for caries and periapical periodontitis detection. *J Dent*. 2022;122:104107. doi:10.1016/j.jdent.2022.104107.

45. Lian LY, Zhu TE, Zhu FD, Zhu HH. Deep learning for caries detection and classification. *Diagnostics*. 2021;11(9):11. doi:10.3390/diagnostics11091672.

46. Majanga V, Viriri S. Automatic blob detection for dental caries. *Appl Sci-Basel*. 2021;11(19):12. doi:10.3390/app11199232.

47. Mao YC, Chen TY, Chou HS, et al. Caries and restoration detection using bitewing film based on transfer learning with CNNs. *Sensors*. 2021;21(13):19. doi:10.3390/s21134613.

48. Moran M, Faria M, Giraldi G, Bastos L, Oliveira L, Conci A. Classification of approximal caries in bitewing radiographs using convolutional neural networks. *Sensors (Basel)*. 2021;21(15):1–12. doi:10.3390/s21155192.

49. Oztekin F, Katar O, Sadak F, et al. An explainable deep learning model to prediction dental caries using panoramic radiograph images. *Diagnostics*. 2023;13(2):226. doi:10.3390/diagnostics13020226.

50. 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. *Clin Oral Investig*. 2022;27(4):1731–1742. doi:10.1007/s00784-022-04801-6.

51. Panyarak W, Wantanajittikul K, Suttapak W, Charuakkra A, Prapayasatok S. Feasibility of deep learning for dental caries classification in bitewing radiographs based on the ICCMS™ radiographic scoring system. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2023;135(2):272–281. doi:10.1016/j.oooo.2022.06.012.

52. Pun MHJ. Real-time caries detection of bitewing radiographs using a mobile phone and an artificial neural network: a pilot study. *Oral*. 2023;3(3):437–449. doi:10.3390/oral3030035.

53. Qayyum A, Tahir A, Butt MA, et al. Dental caries detection using a semi-supervised learning approach. *Sci Rep*. 2023;13(1):749. doi:10.1038/s41598-023-27808-9.

54. Valizadeh S, Goodini M, Ehsani S, Mohseni H, Azimi F, Bakhshandeh H. Designing of a computer software for detection of approximal caries in posterior teeth. *Iran J Radiol*. 2015;12(4):8. doi:10.5812/iranjradiol.12(2)2015.16242.

55. Ying S, Wang BW, Zhu HH, Liu W, Huang F. Caries segmentation on tooth X-ray images with a deep network. *J Dent*. 2022;119:8. doi:10.1016/j.jdent.2022.104076.

56. Zhou X, Yu G, Yin Q, Liu Y, Zhang Z, Sun J. Context aware convolutional neural network for children caries diagnosis on dental panoramic radiographs. *Comput Math Methods Med*. 2022;2022:1–8. doi:10.1155/2022/6029245.

57. Askar H, Krois J, Rohrer C, et al. Detecting white spot lesions on dental photography using deep learning: a pilot study. *J Dent*. 2021;107:6. doi:10.1016/j.jdent.2021.103615.

58. Ding BC, Zhang Z, Liang YR, et al. Detection of dental caries in oral photographs taken by mobile phones based on the YOLOv3 algorithm. *Ann Transl Med*. 2021;9(21):11. doi:10.21037/atm-21-4805.

59. Felsch M, Meyer O, Schlickenrieder A, et al. Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model. *NPJ Digit Med*. 2023;6(1):198. doi:10.1038/s41746-023-00944-2.

60. Kim D, Choi J, Ahn S, Park E. A smart home dental care system: integration of deep learning, image sensors, and mobile controller. *J Ambient Intell Humaniz Comput. Published online*. 2021;14:1–9. doi:10.1007/s12652-021-03366-8.

61. Kuhnisch J, Meyer O, Hesenius M, Hickel R, Gruhn V. Caries detection on intraoral images using artificial intelligence. *J Dent Res*. 2022;101(2):158–165. doi:10.1177/00220345211032524.

62. Mehdizadeh M, Estai M, Vignarajan J, et al. A deep learning-based system for the assessment of dental caries using colour dental photographs. In. 2024. https://ebooks.iospress.nl/doi/10.3233/SHTI231097. Accessed on: 13/02/2024.

63. Moutselos K, Berdouses E, Oulis C, et al. Recognizing occlusal caries in dental intraoral images using deep learning. *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* Ieee; 2019:1617–1620. doi:10.1109/EMBC.2019.8856553.

64. Park EY, Cho H, Kang S, Jeong S, Kim EK. Caries detection with tooth surface segmentation on intraoral photographic images using deep learning. *BMC Oral Health*. 2022;22(1):573. doi:10.1186/s12903-022-02589-1.

65. Tareq A, Faisal MI, MdS Islam, et al. Visual diagnostics of dental caries through deep learning of non-standardised photographs using a hybrid YOLO ensemble and transfer learning model. *Int J Environ Res Public Health*. 2023;20(7):5351. doi:10.3390/ijerph20075351.

66. Thanh MTG, Van Toan N, Ngoc VTN, Tra NT, Giap CN, Nguyen DM. Deep learning application in dental caries detection using intraoral photos taken by smartphones. *Appl Sci*. 2022;12(11):5504. doi:10.3390/app12115504.

67. Yoon K, Jeong HM, Kim JW, Park JH, Choi J. AI-based dental caries and tooth number detection in intraoral photos: model development and performance evaluation. *J Dent*. 2024;141:104821. doi:10.1016/j.jdent.2023.104821.

68. Zhang X, Liang Y, Li W, et al. Development and evaluation of deep learning for screening dental caries from oral photographs. *Oral Dis*. 2022;28(1):173–181. doi:10.1111/odi.13735.

69. Jälevik B, Szigyarto-Matei A, Robertson A. The prevalence of developmental defects of enamel, a prospective cohort study of adolescents in Western Sweden: a Barn I TAnadvarden (BITA, children in dental care) study. *Eur Arch Paediat Dentist*. 2018;19(3):187–195. doi:10.1007/s40368-018-0347-7.

70. Peres KG, Ha DH. Christofis S. Trend and distribution of coronal dental caries in Australians adults. *Aust Dent J*. 2020;65(S1):S32–39. doi:10.1111/adj.12762.

71. Vargas-Ferreira F, Salas MMS, Nascimento GG, et al. Association between developmental defects of enamel and dental caries: a systematic review and meta-analysis. *J Dent*. 2015;43(6):619–628. doi:10.1016/j.jdent.2015.03.011.

72. Nedeljkovic I, De Munck J, Vanloy A, et al. Secondary caries: prevalence, characteristics, and approach. *Clin Oral Investig*. 2020;24(2):683–691. doi:10.1007/s00784-019-02894-0.

73. Carrillo-Perez F, Pecho OE, Morales JC, et al. Applications of artificial intelligence in dentistry: a comprehensive review. *J Esthet Restorat Dentist*. 2022;34(1):259–280. doi:10.1111/jerd.12844.

74. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. https://arxiv.org/pdf/1512.03385. Accessed on 04/03/2024.

75. Rokhshad R, Zhang P, Mohammad-Rahimi H, Shobeiri P, Schwendicke F. Current applications of artificial intelligence for pediatric dentistry: a systematic review and meta-analysis. *Pediatr Dent*. 2024;46(1):27–35.

76. Sounderajah V, Ashrafian H, Golub RM, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open*. 2021;11(6):e047709. doi:10.1136/bmjopen-2020-047709.

77. Schwendicke F, Singh T, Lee JH, et al. Artificial intelligence in dental research: checklist for authors, reviewers, readers. *J Dent*. 2021;107:103610. doi:10.1016/j.jdent.2021.103610.