Health Science Reports

WILEY

Health Science Reports
Open Access

SYSTEMATIC REVIEW OPEN ACCESS

Artificial Intelligence Models Accuracy for Odontogenic Keratocyst Detection From Panoramic View Radiographs: A Systematic Review and Meta-Analysis

Reyhaneh Shoorgashti[1] | Mohadeseh Alimohammadi[2] | Sana Baghizadeh[2] | Bahareh Radmard[3] | Hooman Ebrahimi[1] | Simin Lesan[1]

[1]Department of Oral and Maxillofacial Medicine, School of Dentistry, Islamic Azad University of Medical Sciences, Tehran, Iran | [2]Faculty of Dentistry, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran | [3]School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Correspondence: Reyhaneh Shoorgashti (reyhanehshoorgashti@gmail.com)

Received: 9 November 2024 | **Revised:** 4 February 2025 | **Accepted:** 8 March 2025

Funding: The authors received no specific funding for this work.

Keywords: artificial intelligence | deep learning | odontogenic cysts | odontogenic keratocysts | oral diagnosis | oral health | panoramic radiography

ABSTRACT

Background and Aims: Odontogenic keratocyst (OKC) is a radiolucent jaw lesion often mistaken for similar conditions like ameloblastomas on panoramic radiographs. Accurate diagnosis is vital for effective management, but manual image interpretation can be inconsistent. While deep learning algorithms in AI have shown promise in improving diagnostic accuracy for OKCs, their performance across studies is still unclear. This systematic review and meta-analysis aimed to evaluate the diagnostic accuracy of AI models in detecting OKC from panoramic radiographs.

Methods: A systematic search was performed across 5 databases. Studies were included if they examined the PICO question of whether AI models (I) could improve the diagnostic accuracy (O) of OKC in panoramic radiographs (P) compared to reference standards (C). Key performance metrics including sensitivity, specificity, accuracy, and area under the curve (AUC) were extracted and pooled using random-effects models. Meta-regression and subgroup analyses were conducted to identify sources of heterogeneity. Publication bias was evaluated through funnel plots and Egger's test.

Results: Eight studies were included in the meta-analysis. The pooled sensitivity across all studies was 83.66% (95% CI:73.75%–93.57%) and specificity was 82.89% (95% CI:70.31%–95.47%). YOLO-based models demonstrated superior diagnostic performance with a sensitivity of 96.4% and specificity of 96.0%, compared to other architectures. Meta-regression analysis indicated that model architecture was a significant predictor of diagnostic performance, accounting for a significant portion of the observed heterogeneity. However, the analysis also revealed publication bias and high variability across studies (Egger's test, $p = 0.042$).

Conclusion: AI models, particularly YOLO-based architectures, can improve the diagnostic accuracy of OKCs in panoramic radiographs. While AI shows strong capabilities in simple cases, it should complement, not replace, human expertise, especially in complex situations.

1 | Introduction

Odontogenic keratocysts (OKCs) are cystic lesions of odontogenic origin that present unique challenges in diagnosis and management. Although benign, OKCs are known for their aggressive nature, high recurrence rates, and potential to cause significant bone destruction [1–5]. Early and accurate detection of OKCs is critical for

improving treatment outcomes and minimizing surgical morbidity [6–10].

The initial diagnosis of OKCs traditionally relies on clinical examinations and radiographic imaging, particularly panoramic radiographs [11–14]. These imaging methods are widely available and cost-effective. Panoramic radiographs provide a comprehensive view of the teeth, jaws, and surrounding structures in a single image. However, interpreting these radiographs can be challenging, especially when distinguishing between radiolucent lesions such as OKCs and ameloblastoma (AM). Subtle differences in shape, size, and internal architecture of these lesions may be overlooked, even by experienced clinicians, leading to uncertainty in diagnosis [14–19].

In recent years, artificial intelligence (AI), particularly deep learning algorithms, has gained significant traction in medical image analysis, including dental and maxillofacial radiology [20–22]. These algorithms have demonstrated the capacity to enhance diagnostic accuracy by automating the detection and classification of pathologies on radiographic images [20–25]. AI-driven models, such as convolutional neural networks (CNNs), have been used to detect and differentiate various cystic lesions, including odontogenic tumors and cysts, with promising results [18, 26, 27]. The application of such technologies could potentially alleviate the diagnostic burden on clinicians and improve the consistency of radiographic interpretations [26, 28].

This systematic review and meta-analysis aimed to evaluate the diagnostic accuracy of AI models in detecting and differentiating OKCs from other jaw radiolucent lesions on panoramic radiographs. By synthesizing data from multiple studies, this review seeks to provide a comprehensive assessment of the current performance of AI technologies in this domain, highlight the potential benefits of their integration into clinical practice, and identify areas where further research is needed.

2 | Methods

2.1 | Study Design and Protocol

This systematic review and meta-analysis were conducted according to the Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines [29]. The study protocol has been registered to PROSPERO with the registration number CRD42024606360.

2.2 | Search Strategy

A comprehensive search was conducted in multiple electronic databases, including PubMed, EMBASE, Scopus, ScienceDirect, and Google Scholar up to October 25, 2024. The search terms included combinations of the following keywords: “artificial intelligence,” “deep learning,” “convolutional neural networks,” “CNN,” “automation,” “odontogenic cysts,” “odontogenic keratocyst,” and “OKC.” We also reviewed the reference lists of included studies for additional relevant articles.

2.3 | Eligibility Criteria

Studies were included if they examined the PICO question of whether AI models (I) could improve the diagnostic accuracy (O) of odontogenic keratocysts (OKC) in panoramic radiographs (P) compared to reference standards (C):

Population (P): Studies involving panoramic radiographs of human subjects with radiographically and/or histopathologically confirmed OKCs diagnosed;

Intervention (I): The application of AI models for the automatic detection and classification of OKC in panoramic radiographs;

Comparison (C): Studies comparing the performance of AI models with ground truth diagnoses (based on histopathology) or other diagnostic tools, including manual radiologist readings or alternative machine learning models;

Outcomes (O): The outcome measures include parameters such as diagnostic accuracy, sensitivity, specificity, F1-score for OKC detection, area under the curve (AUC), precision, and recall.

Articles such as case reports, reviews, and commentaries were not included. Additionally, studies lacking diagnostic performance data or those relying on histopathological analysis or Cone beam computed tomography (CBCT) imaging for diagnosis were omitted from the analysis.

2.4 | Study Selection

Two independent reviewers (M.A. and B.R.) screened the titles and abstracts of retrieved studies based on the inclusion and exclusion criteria. Full-text articles were obtained for further assessment, and any discrepancies between the reviewers were resolved through discussion and consultation with a third reviewer (R.S.).

2.5 | Data Extraction

Data extraction was carried out independently by two reviewers (M.A. and B.R.). All extracted data were double-checked by S.B. and any discrepancies between the reviewers were resolved through discussion and consultation with a third reviewer (R.S.). The extracted data included study characteristics (authors, publication year, sample size), lesion prevalence in the data set, patients' demographic features, inclusion and exclusion criteria, AI model architecture, data set details (training, validation, and testing sets), and key performance metrics (sensitivity, specificity, accuracy, and AUC for OKC detection). Where available, confusion matrices (true positives, false positives, true negatives, and false negatives) were also extracted.

2.6 | Risk of Bias Assessment

Two independent reviewers (R.S. and S.L.) assessed the risk of bias (ROB) using the QUADAS-2 tool (Quality Assessment of

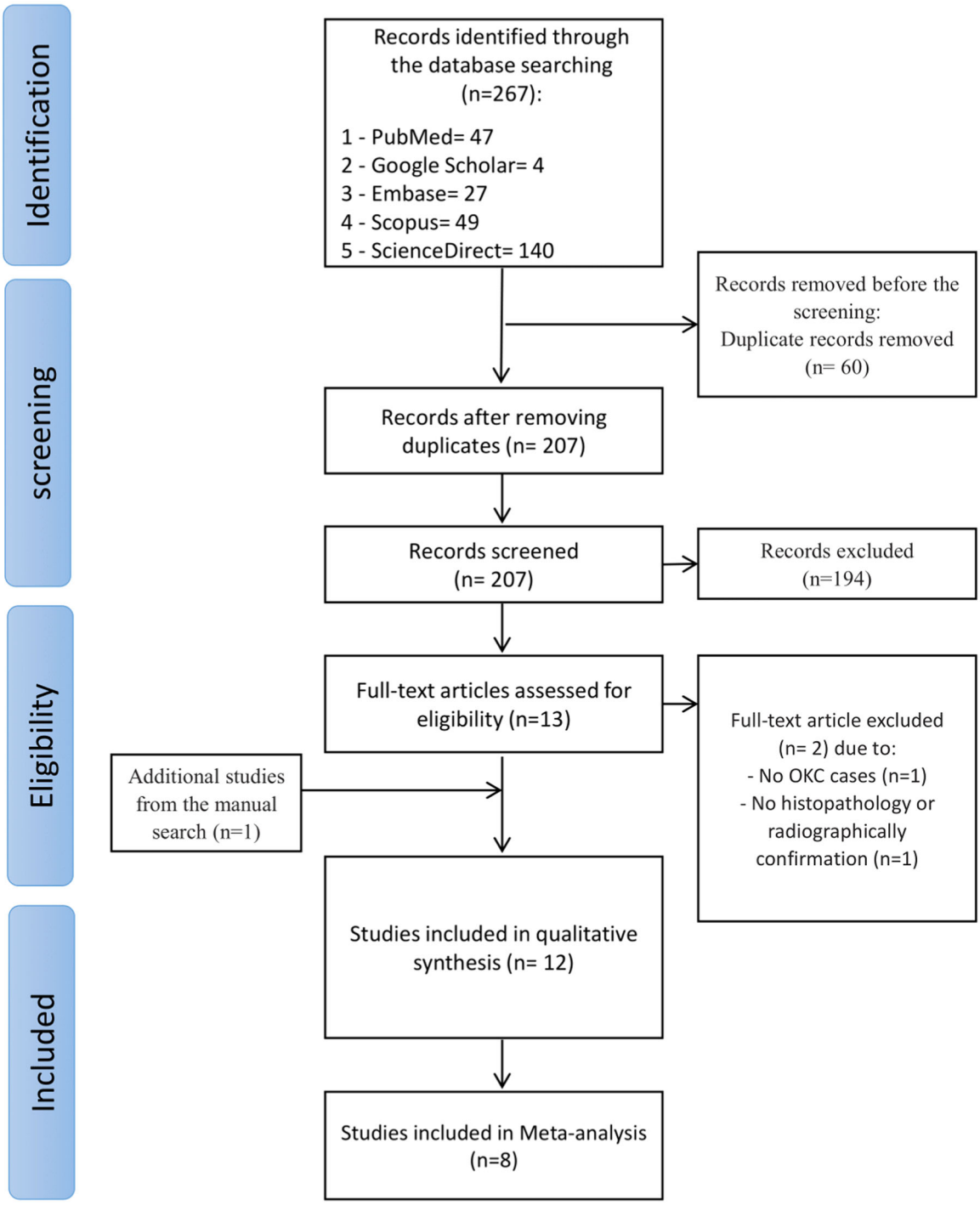

FIGURE 1 | The PRISMA flowchart used for this systematic review resulting in the selection of 12 studies, 8 of which were included in the meta-analysis.

Diagnostic Accuracy Studies). The domains evaluated included patient selection, index test, reference standard, and flow/timing. Based on the assessment, each study was classified as having a low, high, or unclear risk of bias. Disagreements between reviewers were resolved through consensus.

2.7 | Data Synthesis and Statistical Analysis

A meta-analysis was conducted to pool the diagnostic performance metrics of AI models across included studies, following standard meta-analytical guidelines (*Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*) [30]. Pooled estimates of sensitivity and specificity were computed using a random-effects model with inverse variance weighting to account for both within- and between-study variability. Study-specific weights were assigned based on sample size and variance [31].

The diagnostic odds ratio (DOR) and 95% confidence intervals (CIs) were estimated using the DerSimonian and Laird method [30]. To evaluate heterogeneity, we calculated

TABLE 1 | The main finding summary of studies evaluated AI performance in the detection and classification of OKC.

Authors and publication year	Country	Research objective	Data set	Data set breakdown (training, validation, testing)	Participant demographics	Inclusion and exclusion criteria	Annotation and labeling method	Machine learning task description	Data pre-processing techniques	Model type and algorithm used	The best OKC-related performance results
Sim et al. (2024) [34]	Republic of Korea	Distinguish OKC from simple bone cyst	188 panoramic radiographs (63 OKC; 125 simple bone cysts)	Training: 80% validation: 20% fivefold cross-validation to ensure model performance	NA	Inclusion: Radiographs of histopathologically confirmed diagnoses Exclusion: Pantographs without clear depictions of well-defined, circumscribed lesions	Manual regions of interest announcement based on histopathological diagnosis	Detection	299 × 299 pixel-image cropping and resizing	Inception-ResNet-V2	Accuracy: 82.9% precision: 80.0% recall: 61.5% F1-score: 69.5%
Kise et al. (2023) [35]	Japan	Investigate the classification performance of a pre-trained model for radiolucent lesions	310 patients (38 AMs; 41 OKCs; 82 DCs; 84 RCs; 65 Stafne's bone cavities)	Training: 200 images validation: 50 images testing: 60 images (training: 26/ validation: 7/ testing: 8 for OKCs specifically)	Average age: 47.9 ± 16.1years gender distribution (male: 26, female: 15)	Inclusion: cyst-like radiolucent lesions in the mandible Exclusion: history of previous surgery or malignant lesions	Manual regions of interest announcement by two radiologists with over 10 years of experience	Detection classification	900 × 900 pixel-image cropping	DetectNet FCN three target models (T10, T21, and T42)	Detection sensitivity: 88% classification: accuracy: 92% sensitivity: 57% specificity: 98%
Rašić et al. (2023) [36]	Croatia	Evaluate a deep learning model (YOLOv8) for detecting and segmenting radiolucent lesions	226 radiolucent lesions (13 Ams; 33 OKCs; 29 DCs; 138 RCs; 13 residual cysts)	200 panoramic radiographs (training: 60%/ validation: 20%/ testing: 20%) fivefold cross-validation to improve generalization	NA	Inclusion: The presence of a radiolucent lesion in the lower jaw Histopathologic verification of the diagnosis Exclusion: NA	Manual regions of interest announcement by a radiologist and an oral and maxillofacial surgeon	Detection segmentation	Translation, scaling, rotation, horizontal flipping, and mosaic effects to each batch	CNN (YOLOv8)	Detection: mean average precision: 97.5% recall: 94.4% segmentation: precision: 76% recall: 75.5%

(Continues)

TABLE 1 | (Continued)

Authors and publication year	Country	Research objective	Data set	Data set breakdown (training, validation, testing)	Participant demographics	Inclusion and exclusion criteria	Annotation and labeling method	Machine learning task description	Data pre-processing techniques	Model type and algorithm used	The best OKC-related performance results
Feher et al. (2022) [37]	Austria	Emulate clinical diagnostic reasoning for the detection and classification of jaw cysts	855 OPGs (positive cases) 384 OPGs (additional evaluation set: 240 negative cases)	1239 (80%/10%/10%) training: 855 positive OPGs testing: 384 OPGs	Average age: 53.5 years (41–63) gender distribution: (male: 733, female: 506)	Inclusion: Cystic lesions OPGs with confirmed histopathological diagnoses and random negative OPGs Exclusion: NA	Manual bounding boxes around lesions	Detection classification	Adding Gaussian noise and intensity shifts	RetinaNet U-Net++ architectures Random Forest classifier	Classification: precision: 89% sensitivity: 84% specificity: 59% F1-score: 86%
Tajima et al. (2022) [38]	Japan	Detect cyst-like radiolucent lesions of the jaws	7160 panoramic radiographs	Training: 7160 radiographs testing: 100 radiographs	NA	Inclusion: Panoramic radiographs with cyst-like radiolucent lesions Exclusion: maxillary molars area lesions	Labeling by two oral surgeons with more than 15 years of experience using LabelMe software	Detection	Rotation, inverting, moving, and trimming	CNN (YOLO v3)	Accuracy:98.3% precision: 99% recall: 94.4% Specificity 99.7% F1-score: 96.6%
Lee et al. (2021) [39]	Republic of Korea	Differentiate SBC from other mandibular pathologies	458 panoramic images (93 AM; 91 OKC; 98 DC; 176 SBC)	Training: 70% testing: 30%	NA	Inclusion: Panoramic radiographs of histopathologically confirmed lesions Exclusion: NA	Manual regions of interest announcement using free-drawn lines outlining	Classification	Horizontal flipping and reseizing	DenseNet121	Accuracy: 99.25% sensitivity: 98.08% specificity: 100%
Liu et al. (2021) [40]	China	Differentiate AM and OKC	420 panoramic images (209 AM; 211 OKC)	Training: 295 images validation: 42 images testing: 83 images	NA	Inclusion: Samples confirmed by histopathological	Radiologists cropped the region of interest	Classification	Random rotation and flip (horizontally or vertically) transform (used only in the training step)	CNN	Accuracy: 90.36% sensitivity: 92.88% specificity: 87.80%

(Continues)

TABLE 1 | (Continued)

Authors and publication year	Country	Research objective	Data set	Data set breakdown (training, validation, testing)	Participant demographics	Inclusion and exclusion criteria	Annotation and labeling method	Machine learning task description	Data pre-processing techniques	Model type and algorithm used	The best OKC-related performance results
						analysis, reviewed by the radiologists The images clearly showed the lesion area Exclusion: NA					F1-score: 90.70% AUC: 94.6%
Watanabe et al. (2021) [41]	Japan	Detect maxillary cyst-like lesions	436 maxillary cyst-like lesions (23 OKC; 37 DC; 44 nasopalatine duct cysts; 9 other lesions)	Training: 330 images testing: 106 images (testing set 1: 71 images; testing set 2: 35 images)	Average age: 44.8 ± 16.6 years gender distribution: (male: 221, female: 191)	Inclusion: Cyst-like lesions or benign tumors of the maxilla, with bone resorption of at least 10 mm in diameter, confirmed by histopathology Exclusion: NA	Manual labeling by radiologists	Detection classification	900 × 900 pixels and 24-bit depth per pixel standardization	DetectNet neural network	Detection: precision: 100% recall: 70% F1-score: 82.4% classification: sensitivity: 71.4%
Kwon et al. (2020) [42]	Republic of Korea	Automatically diagnose odontogenic cysts and tumors	1282 panoramic radiographs (230 AM; 300 OKC; 350 DC; 302 PC; 100 normal jaws)	Training: 80% testing: 20%	NA	Inclusion: radiographs of histopathological and radiological confirmed cysts and tumors Exclusion: NA	Manual bounding boxes around lesions using ImageJ software	Detection classification	Image contrast enhancement and 608 × 608 pixel-standardization horizontal flipping, ±1° rotation, grayscale intensity adjustments	CNN (YOLOv3)	detection: precision: 91% recall: 83% F1-score: 76% classification: accuracy: 94% sensitivity: 98.4% specificity: 92.3% AUC: 97%
Lee et al. (2020) [43]	Republic of Korea	Compare the diagnostic	1140 panoramic radiographs	Training: 80% testing: 20%	Average age: 40–59 years	Inclusion: Radiographs of	Manual regions of	Detection	Horizontal and vertical flipping, rotation, width	CNN (GoogLeNet	Accuracy: 81.8%

(Continues)

TABLE 1 | (Continued)

Authors and publication year	Country	Research objective	Data set	Data set breakdown (training, validation, testing)	Participant demographics	Inclusion and exclusion criteria	Annotation and labeling method	Machine learning task description	Data pre-processing techniques	Model type and algorithm used	The best OKC-related performance results
		accuracy of OKCs, DCs, and PCs	(260 OKC; 436 DC; 417 PC)		gender distribution (male:167, female: 80)	histopathologically confirmed diagnoses Exclusion: Severe image distortion, artificial noise, blur, or poor-quality radiographs	interest announcement	classification	and height shifting, shearing, and zooming	Inception v3 model)	sensitivity: 88.2% specificity: 77% AUC: 84.7%
Yang et al. (2020) [44]	Republic of Korea	Detect and classify odontogenic cysts and tumors	1603 panoramic radiographs (160 AM; 316 OKC; 1094 DC, 33 normal jaw)	Training: 90% testing: 10%	Average age: training set: 42 years (31–53) (68% male, 32% female) testing set: 37 years (25–48) (65.7% male, 34.3% female)	Inclusion: Radiographs of radiographically confirmed diagnoses Exclusion: Postoperative radiographs and ambiguous cases	Radiograph assessment by two general dentists and 3 oral and maxillofacial specialists labeling by radiologists using the YOLO mark tool	Detection classification	Horizontal and vertical flipping, translation, and scaling, expanding the original 1422 images to 16,224 images	CNN (YOLO v2)	Accuracy: 66.3% precision: 63.5% recall: 70.2% F1-score: 69.3%
Ariji et al. (2019) [45]	Japan	Automatically detect and classify radiolucent lesions in mandibles	235 (34 AM; 39 OKC; 74 DC; 75 RC; 13 Simple bone cysts)	Training: 210 radiographs testing: 50 radiographs from the training data set and 25 new radiographs	Average age: 45 years (15–85) gender distribution (male: 128, female: 82)	Inclusion: Histopathologically confirmed lesions with at least 10 mm diameters Exclusion: NA	Manual labeling by one radiologist	Detection classification	900 × 900 pixel-image cropping	DetectNet	Detection sensitivity: 100% classification sensitivity: 13%

Abbreviations: AM, Ameloblastoma; CLAHE, contrast-limited adaptive histogram equalization; CNN, convolutional neural network; DC, dentigerous cyst; FCN, fully conventional network; NA, not available; OKC, odontogenic keratocyst; OPG, orthopantomogram; PC, periapical cyst; RC, radicular cyst; SBC, Stafne's bone cavity.

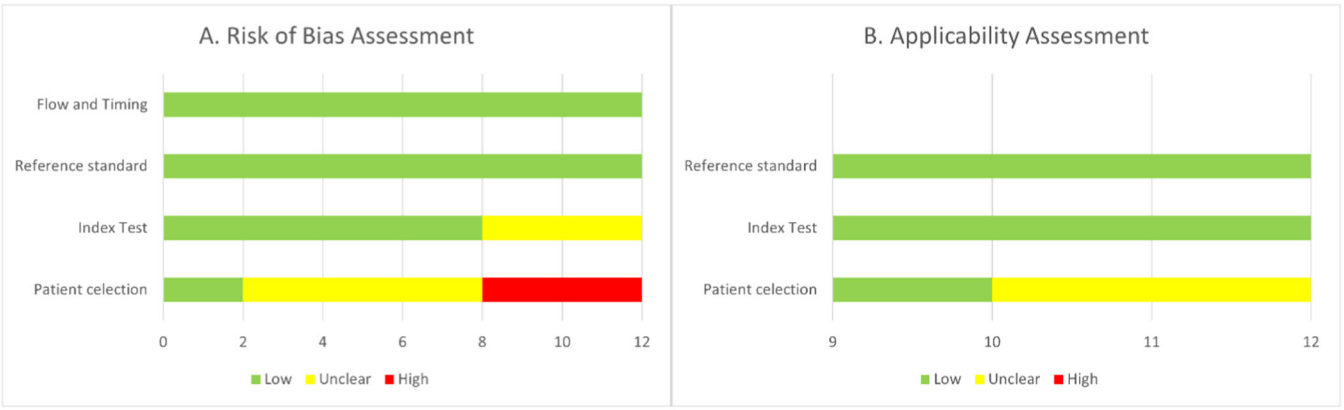

FIGURE 2 | The ROB assessment (A) and applicability analysis (B) results of 12 included studies using QUADAS-2.

TABLE 2 | The details of ROB and applicability assessed articles using QUADAS-2.

	Risk of bias assessment				Applicability assessment		
Study	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Sim et al. [34]	High	Unclear	Low	Low	low	low	low
Kise et al. [35]	High	Unclear	Low	Low	low	low	low
Rašić et al. [36]	Unclear	Low	Low	Low	low	low	low
Feher et al. [37]	High	Low	Low	Low	Unclear	low	low
Tajima et al. [38]	Unclear	Low	Low	Low	low	low	low
Lee et al.[39]	High	Unclear	Low	Low	low	low	low
Liu et al. [40]	Unclear	Low	Low	Low	low	low	low
Watanabe et al. [41]	Unclear	Low	Low	Low	Unclear	low	low
Kwon et al. [42]	Low	Low	Low	Low	low	low	low
Lee et al. [43]	Unclear	Low	Low	Low	low	low	low
Yang et al. [44]	Unclear	Unclear	Low	Low	low	low	low
Ariji et al. [45]	Low	Low	Low	Low	low	low	low

the I^2 statistic, with $I^2 > 50\%$ indicating substantial heterogeneity [30, 32]. The Cochran Q test was performed, with a significance level of $p < 0.05$ suggesting statistical heterogeneity [30].

Pre-specified analyses included subgroup analyses based on AI model architecture, publication year, and sample size. Exploratory analyses involved meta-regression to assess the impact of study characteristics (model type, publication year) on diagnostic performance. The proportion of variance explained by the meta-regression model was quantified using the coefficient of determination (R^2), which indicates the percentage of heterogeneity accounted for by the included covariates.

Publication bias was assessed using funnel plots and Egger's test, with a p-value < 0.05 suggesting the presence of bias. To further address potential publication bias, the Trim-and-Fill method was applied to estimate and adjust for the number of missing studies [33].

All statistical analyses were performed using Python (version 3.8). The following libraries were utilized: scipy.stats for statistical tests, numpy for numerical computations, pandas for data manipulation, matplotlib and seaborn for data visualization, and scikit-learn for meta-regression analyses. Two-sided p-values were reported throughout, with statistical significance defined as $\alpha = 0.05$.

3 | Results

Figure 1 shows the PRISMA fellow chart utilized in this systematic review.

3.1 | Study Characteristics

Table 1 displays the main summary of the findings. A total of 12 studies published between 2019 and 2024 met the inclusion criteria. The studies originated from various countries,

including Republic of Korea ($n = 5$) [34, 39, 42–44], Japan ($n = 4$) [35, 38, 41, 45], China ($n = 1$) [40], Croatia ($n = 1$) [36], and Austria ($n = 1$) [37].

Regarding study design, all included studies were retrospective analyses of AI model performance on panoramic radiographs. The total number of included patients varied across studies, with sample sizes ranging from 188 to 7160 radiographs. The data sets used for training, validation, and testing varied in composition, with some studies employing fivefold cross-validation for improved generalization.

The investigated AI models included YOLO variants ($n = 4$) [36, 38, 42, 44], DetectNet implementations ($n = 3$) [35, 41, 45], Inception-based models ($n = 2$) [34, 43], and other architectures ($n = 3$) [37, 39, 40]. The investigated models employed different approaches to medical image analysis, ranging from pure classification to combined detection and segmentation tasks (Table 1).

3.2 | Risk of Bias Assessment

ROB assessment across 12 studies showed that patient selection was a key area of concern, with 33.3% of studies rated high risk and 50% rated unclear due to inadequate reporting samples. The index test domain was generally well-handled, with 66.7% of studies rated low risk, though 33.3% were unclear due to insufficient details. Both the reference standard and flow/timing domains were consistently rated low risk in all studies. Regarding applicability concerns, most studies showed low concern across all domains, with 83.3% rated low for patient selection and 100% rated low for the index test and reference standard domains (Figure 2 and Table 2).

3.3 | Diagnostic Performance

Among the studies, the performance of individual AI models varied notably. The study by Lee et al. [39] reported the highest diagnostic performance, with a sensitivity of 98.08% and specificity of 100%. Conversely, Sim et al. [34] reported the lowest sensitivity and specificity, both at 61.5%.

The overall accuracy of the AI models across the 12 studies ranged from 66.3% to 99.25%, with a mean accuracy of 88.11% ± 10.86%. Five of the studies reported accuracies above 90% [35, 38–40, 42], indicating that AI models can demonstrate excellent diagnostic performance in detecting OKC.

3.4 | Meta-Analysis

Eight studies that provided complete data on both sensitivity and specificity were included in the diagnostic accuracy meta-analysis [34, 35, 37–40, 42, 43]. The pooled sensitivity across all studies was 84.31% (95% CI: 71.24%–97.38%), while the pooled specificity was 84.41% (95% CI: 70.52%–98.30%), indicating a strong ability to correctly identify both positive and negative cases (Figure 3). The pooled diagnostic odds ratio (DOR) was 27.84 (95% CI: 12.46%–62.19%), reflecting a high overall diagnostic accuracy.

There was substantial heterogeneity among the included studies. Sensitivity exhibited significant heterogeneity

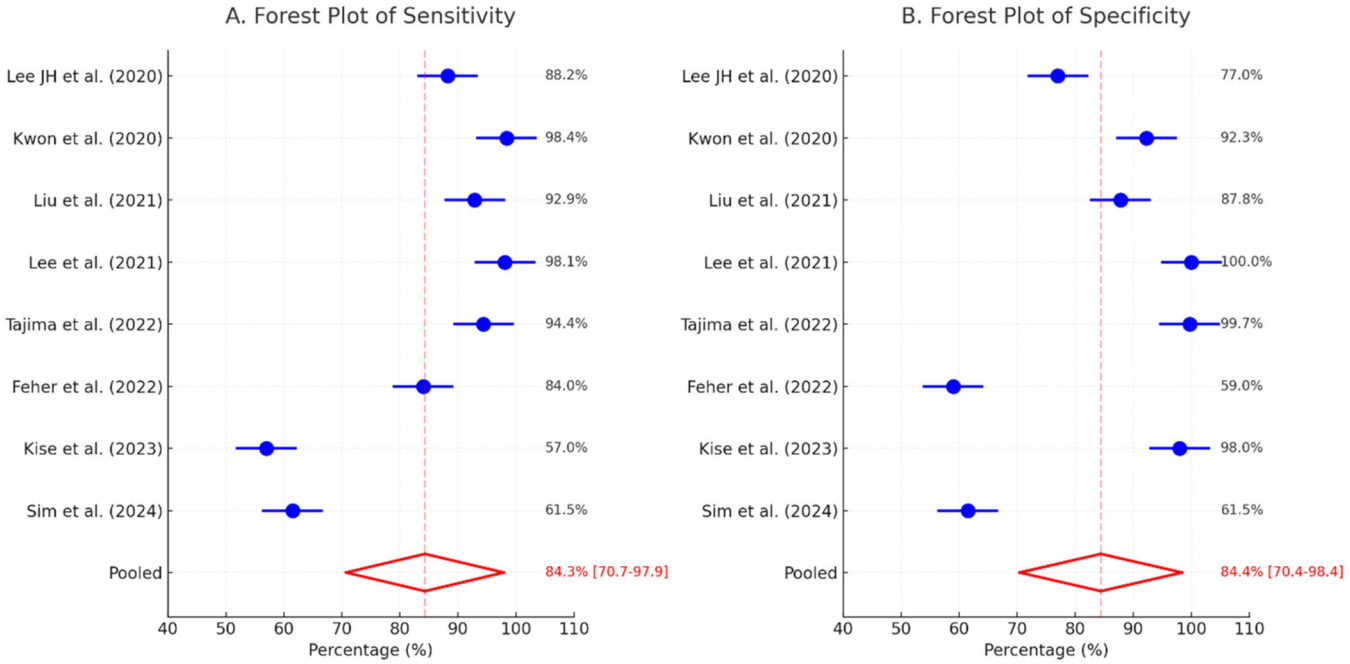

FIGURE 3 | Forest plots for pooled sensitivity and specificity: This figure shows the forest plots for sensitivity (A) and specificity (B) from the meta-analysis. Each dot represents the estimate for a study, with error bars indicating uncertainty. The red dashed lines and diamonds represent the pooled sensitivity (84.3%) and specificity (84.4%) across all studies.

($I^2 = 89.5\%$), as did specificity ($I^2 = 92.3\%$) (Figure 4). The overall area under the summary ROC curve (AUC) was 0.912, suggesting robust diagnostic accuracy despite inter-study variability.

3.5 | Subgroup and Meta-Regression Analysis

Subgroup analysis revealed notable differences in performance based on model architecture. YOLO-based architectures demonstrated superior accuracy, with a pooled sensitivity of 96.4% and specificity of 96.0%, outperforming Inception-based models (sensitivity: 74.9%, specificity: 69.3%) and DetectNet (sensitivity: 57.0%, specificity: 98.0%).

Meta-regression analysis explored sources of heterogeneity, evaluating the effects of publication year and model architecture (Figure 4). Model architecture was a significant predictor of diagnostic performance: YOLO models demonstrated significantly higher accuracy (coefficient: 2.147, $p < 0.001$), while Inception-based models showed significantly lower performance (coefficient: −0.892, $p = 0.043$). Publication year was also associated with a decline in diagnostic accuracy, with more recent studies reporting lower performance (coefficient: −0.285, 95% CI: −0.472 to −0.098, $p = 0.003$).

The combined model explained 79.3% of the observed heterogeneity ($R^2 = 0.793$), although significant residual heterogeneity remained ($p < 0.001$). Model architecture alone accounted for 68.1% of the variation ($R^2 = 0.681$), whereas publication year explained 43.7% ($R^2 = 0.437$). Sensitivity analysis reinforced these findings, with YOLO models showing a significantly higher effect (coefficient: 15.621, $p < 0.001$) and Inception models demonstrating significantly lower performance (coefficient: −12.334, $p < 0.001$). These findings emphasize the critical role of model selection in diagnostic imaging, highlighting the superior performance of YOLO-based architectures. However, persistent residual heterogeneity suggests that

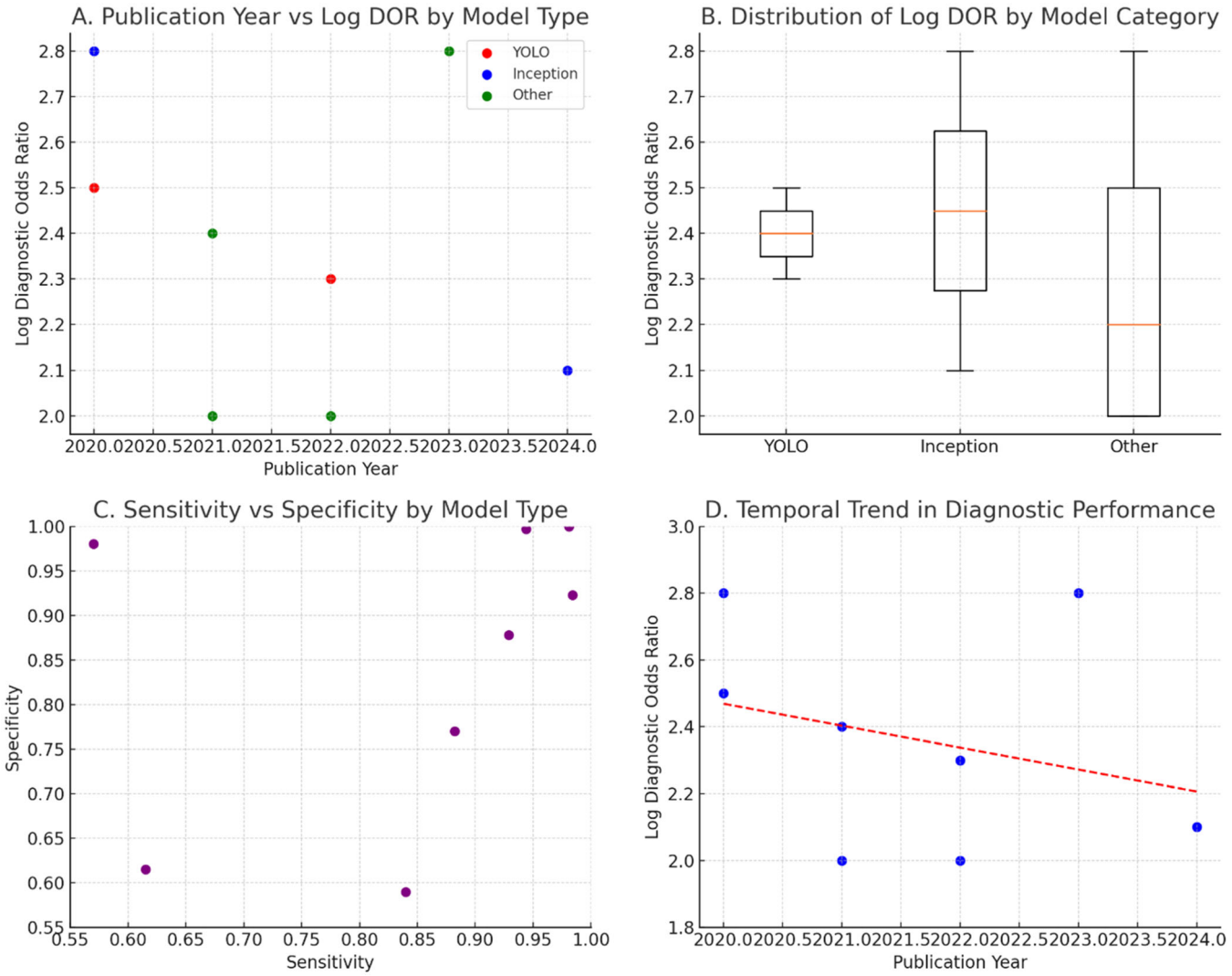

FIGURE 4 | This figure presents four subplots illustrating the diagnostic performance of different models across multiple studies. (A) The relationship between publication year and Log Diagnostic Odds Ratio (Log DOR) for different model types, with distinct colors representing each model type (Inception, YOLO, Other). (B) The box plot comparing the variability of Log DOR values across model types, indicating that YOLO models show less variability. (C) The scatter plot displaying the trade-off between sensitivity and specificity across different studies, highlighting the variability in model performance. (D) The temporal trend in Log DOR over the years, with a linear trendline.

additional unmeasured factors may influence diagnostic accuracy across studies.

3.6 | Publication Bias

Assessment of publication bias revealed significant asymmetry in the funnel plot (Egger's test, $p = 0.042$) (Figure 5), indicating a potential bias toward studies reporting higher diagnostic accuracy. Additionally, small-study effects were observed, as smaller studies tended to report inflated accuracy estimates.

4 | Discussion

The use of AI in diagnosing OKC and other jaw lesions has advanced significantly [36, 38, 42, 44, 46–49]. Deep learning models show great potential in enhancing clinical workflows by improving both the speed and accuracy of diagnoses [49–52]. Unlike the variability often seen in human diagnostic performance, AI models can provide a more consistent and reliable output. This consistency addresses a critical issue in clinical practice, where human performance can vary due to factors like experience, fatigue, or complex image interpretations [36–38, 42, 44, 53].

As demonstrated in multiple studies included in this review, AI models—especially those based on YOLO architectures—have achieved diagnostic accuracy that is comparable to, and in some cases exceeds, that of human radiologists [36–38, 42, 44]. Historically, human performance in this area has varied, with sensitivity and specificity generally ranging from 70% to 80% and 60% to 75%, respectively [37]. In contrast, models developed by Lee et al. [39] and Tajima et al. [38] exhibited sensitivity levels exceeding 94% and specificity levels exceeding 99%.

AI's ability to handle routine cases efficiently, particularly those with well-defined radiographic characteristics, enables it to reduce the burden on clinicians [45, 54–56]. Its capacity for rapid analysis is crucial in high-volume clinical settings where human resources are limited [57–59]. By offering accurate preliminary assessments, AI can help pinpoint cases that need immediate attention. This allows human experts to concentrate on more complex or uncertain cases, where their expertise is essential. Such cases often require the integration of clinical history, symptoms, and advanced imaging techniques like CBCT, which AI models cannot fully replicate [37, 40].

The limitations of AI should not be overlooked. Despite AI's advantages in terms of speed and consistency, its performance remains closely tied to the quality and diversity of the training data [60–62]. Models trained on limited or homogeneous datasets often face challenges in generalizing to diverse clinical scenarios, particularly when tasked with detecting lesions in complex anatomical regions like the maxilla [34, 38, 41, 42, 44]. The overlapping structures in these areas often obscure lesion boundaries, leading to performance gaps compared to human radiologists, who can incorporate contextual information and clinical judgment into their diagnostic decisions [42, 44].

The integration of more advanced imaging modalities, such as CBCT, has been shown to improve AI's diagnostic accuracy. Lee HJ et al. [43] has demonstrated that models trained on CBCT data are better equipped to handle complex cases than those relying solely on panoramic radiographs. This underscores the importance of multi-modal approaches in optimizing AI performance in clinical environments.

AI's role should thus be viewed as a complement to, rather than a replacement for, human expertise [37, 40, 63–65]. Its strength lies in its ability to enhance workflow efficiency by handling routine cases and flagging potential anomalies for further review. However, human oversight remains critical, particularly in cases where clinical context, patient history, and complex imaging are required for a definitive diagnosis [37, 40, 65].

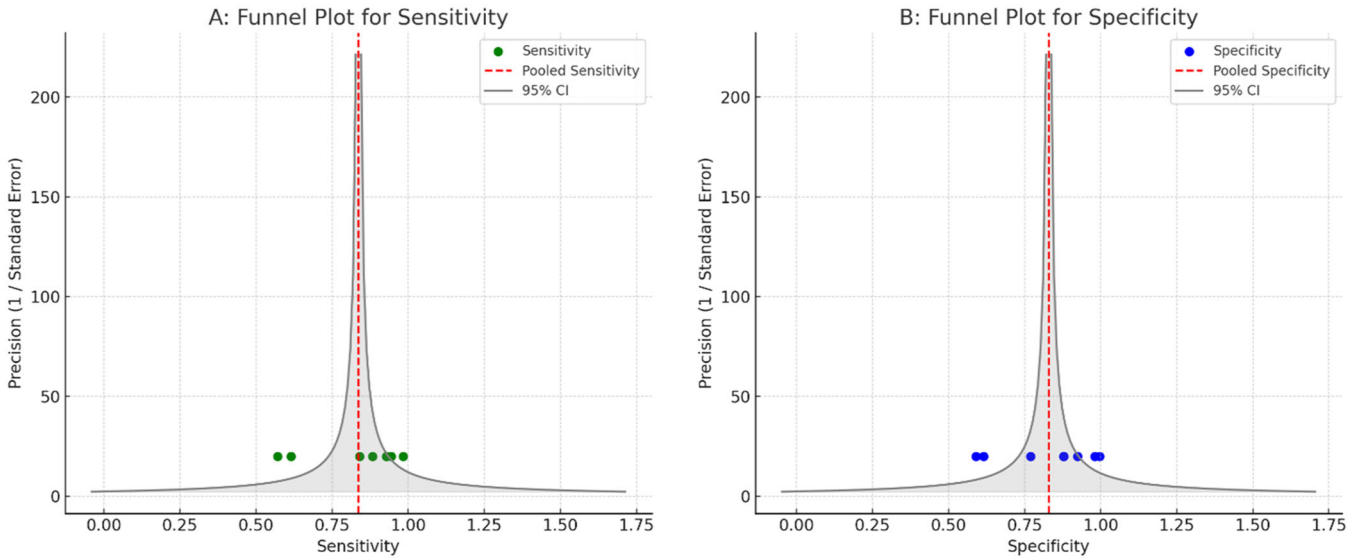

FIGURE 5 | Funnel plots for sensitivity (A) and specificity (B). The figure shows funnel plots to assess publication bias. The red dashed lines mark the pooled estimates, while the dots represent individual studies. The gray regions represent the 95% confidence intervals. The uneven distribution of studies suggests the presence of publication bias and high heterogeneity.

This review also highlights several limitations. (1) Four studies [35, 36, 41, 45] have a sample size of less than 50 samples in the OKC group, which restricts their generalizability to diverse clinical settings. (2) The significant variability among the included studies, highlighted by substantial heterogeneity metrics, indicates that considerable work is needed to standardize AI's application in various clinical settings. (3) The potential for publication bias further complicates the assessment of AI's true capabilities, as studies with positive outcomes may be more likely to be published. So, it is essential to address these limitations by promoting the use of larger, more diverse datasets and standardizing reporting methods to ensure AI's continued advancement and integration into clinical practice. (4) Our meta-analysis indicates that deep learning models, especially those based on YOLO architectures, have shown superior diagnostic accuracy in various studies. However, the results of this meta-analysis are limited by the small number of studies included ($n = 8$) and the variability in their sample sizes and inclusion criteria. The sample sizes varied significantly, ranging from fewer than 200 to over 7100 radiographs, which may contribute to inconsistencies in performance. (5) The choice of data pre-processing techniques varied significantly among the included studies, potentially influencing the reported diagnostic accuracy. Some studies applied contrast adjustment, histogram equalization, and noise reduction to enhance image quality, while others incorporated data augmentation techniques such as rotation, flipping, and scaling to improve model generalization. While pre-processing steps can help optimize model performance, their effects were not systematically analyzed across studies, making it difficult to draw definitive conclusions about their direct impact. Notably, YOLO-based models, which showed superior diagnostic performance, were often trained on datasets that incorporated diverse augmentation techniques. However, excessive augmentation may introduce artifacts that do not accurately represent real-world clinical images, potentially affecting the model's applicability. Establishing standardized pre-processing protocols could improve comparability between studies and provide clearer insights into their role in AI model performance.

5 | Conclusion

In conclusion, AI models, particularly those based on YOLOv3, show great promise in diagnosing OKCs from panoramic radiographs. These models not only match human diagnostic performance but also provide better specificity in some cases, which could enhance clinical outcomes by reducing false positives and unnecessary procedures. While AI demonstrates strong diagnostic capabilities, especially in routine and straightforward cases, its role in clinical settings should be viewed as complementary to the nuanced judgment of human experts.

Authors Contributions

Reyhaneh Shoorgashti: conceptualization, data curation, formal analysis, investigation, methodology, visualization, project administration, software, validation, writing – original draft, writing – review and editing. **Mohadeseh Alimohammadi:** data curation, investigation, methodology, writing – original draft. **Sana Baghizadeh:** data curation, investigation, methodology. **Bahareh Radmard:** data curation, investigation, methodology. **Hooman Ebrahimi:** supervision, validation. **Simin Lesan:** project administration, supervision, validation.

Acknowledgments

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Also, all authors have read and approved the final version of the manuscript. Dr. Reyhaneh Shoorgashti had full access to all of the data in this study and takes complete responsibility for the integrity of the data and the accuracy of the data analysis.

Transparency Statement

The lead author Reyhaneh Shoorgashti affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References

1. E. A. Al-Moraissi, A. Kaur, R. S. Gomez, and E. Ellis 3rd, "Effectiveness of Different Treatments for Odontogenic Keratocyst: A Network Meta-Analysis," *International Journal of Oral and Maxillofacial Surgery* 52, no. 1 (2023): 32–43, https://doi.org/10.1016/j.ijom.2022.09.004.

2. R. Winters, M. Garip, J. Meeus, R. Coropciuc, and C. Politis, "Safety and Efficacy of Adjunctive Therapy in the Treatment of Odontogenic Keratocyst: A Systematic Review," *British Journal of Oral and Maxillofacial Surgery* 61, no. 5 (2023): 331–336, https://doi.org/10.1016/j.bjoms.2023.04.006.

3. R. Shoorgashti, D. Sadri, and S. Farhadi, "Expression of Epidermal Growth Factor Receptor by Odontogenic Cysts: A Comparative Study of Dentigerous Cyst and Odontogenic Keratocyst," *Journal of Research in Dental Sciences* 17, no. 2 (2020): 127–136.

4. S. Mirzaee, R. Shoorgashti, D. Sadri, and S. Farhadi, "Comparison of EGFR Expression in Ameloblastoma and Odontogenic Keratocyst," *Journal of Research in Dental and Maxillofacial Sciences* 8, no. 4 (2023): 274–279, https://doi.org/10.61186/jrdms.8.4.274.

5. R. Shoorgashti, A. Moshiri, and S. Lesan, "Evaluation of Oral Mucosal Lesions in Iranian Smokers and Non-Smokers," *Nigerian Journal Of Clinical Practice* 27, no. 4 (2024): 467–474, https://doi.org/10.4103/njcp.njcp_702_23.

6. T. O. F. Gonçalves, R. M. R. Rangel, G. A. Marañón-Vásquez, et al., "Management and Recurrence of the Odontogenic Keratocyst: An Overview of Systematic Reviews," *Oral and Maxillofacial Surgery* 28, no. 4 (2024): 1457–1478, https://doi.org/10.1007/s10006-024-01277-4.

7. J. K. Brooks, A. S. Sultan, M. P. Rabkin, et al., "Recurrent Peripheral Odontogenic Keratocyst: Review of the Literature and Presentation of a Novel Case Initially Masquerading as an Atypical Infected Lateral Periodontal Cyst," *Journal of Stomatology, Oral and Maxillofacial Surgery* 125, no. S4 (2024): 101540, https://doi.org/10.1016/j.jormas.2023.101540.

8. S. Grover, S. Hegde, and R. Mascarenhas, "Management Regulations for Odontogenic Keratocyst: A Case Report and Review of the Literature," *Journal of Medical Case Reports* 18, no. 1 (2024): 152, https://doi.org/10.1186/s13256-024-04473-8.

9. E. A. Bilodeau and B. M. Collins, "Odontogenic Cysts and Neoplasms," *Surgical Pathology Clinics* 10, no. 1 (2017): 177–222, https://doi.org/10.1016/j.path.2016.10.006.

10. N. Mirhosseini, R. Shoorgashti, and S. Lesan, "The Evaluation of Clinical Factors Affecting Oral Health Impacts on the Quality of Life of Iranian Elderly Patients Visiting Dental Clinics: A Cross-Sectional Study," *Special Care in Dentistry* 44, no. 4 (2024): 1219–1227, https://doi.org/10.1111/scd.12980.

11. D. Rioux-Forker, A. C. Deziel, L. S. Williams, and A. R. Muzaffar, "Odontogenic Cysts and Tumors," *Annals of Plastic Surgery* 82, no. 4 (2019): 469–477, https://doi.org/10.1097/sap.0000000000001738.

12. J. Kitisubkanchana, N. H. Reduwan, S. Poomsawat, S. Pornprasertsuk-Damrongsri, and C. Wongchuensoontorn, "Odontogenic Keratocyst and Ameloblastoma: Radiographic Evaluation," *Oral Radiology* 37, no. 1 (2021): 55–65, https://doi.org/10.1007/s11282-020-00425-2.

13. M. S. Bispo, M. Pierre Júnior, and A. L. ApolinárioJr., et al., "Computer Tomographic Differential Diagnosis of Ameloblastoma and Odontogenic Keratocyst: Classification Using a Convolutional Neural Network," *Dentomaxillofacial Radiology* 50, no. 7 (2021): 20210002, https://doi.org/10.1259/dmfr.20210002.

14. D. B. M. Alves, F. M. Tuji, F. A. Alves, et al., "Evaluation of Mandibular Odontogenic Keratocyst and Ameloblastoma by Panoramic Radiograph and Computed Tomography," *Dentomaxillofacial Radiology* 47, no. 7 (2018): 20170288, https://doi.org/10.1259/dmfr.20170288.

15. F. Cavarra, P. Boffano, M. Brucoli, et al., "Imaging of Odontogenic Keratocysts: A Pictorial Review," *Minerva Dental and Oral Science* 71, no. 1 (2022): 48–52, https://doi.org/10.23736/s2724-6329.21.04582-4.

16. Y. Meng, Y. N. Zhao, Y. Q. Zhang, D. G. Liu, and Y. Gao, "Three-Dimensional Radiographic Features of Ameloblastoma and Cystic Lesions in the Maxilla," *Dentomaxillofacial Radiology* 48, no. 6 (2019): 20190066, https://doi.org/10.1259/dmfr.20190066.

17. R. Zhang, J. Yang, J. Zhang, Y. Hong, X. Xie, and T. Li, "Should the Solid Variant of Odontogenic Keratocyst and Keratoameloblastoma be Classified as the Same Entity? A Clinicopathological Analysis of Nine Cases and a Review of the Literature," *Pathology* 53, no. 4 (2021): 478–486, https://doi.org/10.1016/j.pathol.2020.09.028.

18. R. S. Fedato Tobias, A. B. Teodoro, K. Evangelista, et al., "Diagnostic Capability of Artificial Intelligence Tools for Detecting and Classifying Odontogenic Cysts and Tumors: A Systematic Review and Meta-Analysis," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology* 138, no. 3 (2024): 414–426, https://doi.org/10.1016/j.oooo.2024.03.004.

19. J. P. P. Gomes, C. M. Ogawa, R. V. Silveira, et al., "Magnetic Resonance Imaging Texture Analysis to Differentiate Ameloblastoma From Odontogenic Keratocyst," *Scientific Reports* 12, no. 1 (2022): 20047, https://doi.org/10.1038/s41598-022-20802-7.

20. A. G. Iruvuri, G. Miryala, Y. Khan, et al., "Revolutionizing Dental Imaging: A Comprehensive Study on the Integration of Artificial Intelligence in Dental and Maxillofacial Radiology," *Cureus* 15, no. 12 (2023): 50292, https://doi.org/10.7759/cureus.50292.

21. K. Hung, C. Montalvao, R. Tanaka, T. Kawai, and M. M. Bornstein, "The Use and Performance of Artificial Intelligence Applications in Dental and Maxillofacial Radiology: A Systematic Review," *Dentomaxillofacial Radiology* 49, no. 1 (2020): 20190107, https://doi.org/10.1259/dmfr.20190107.

22. T. Bonny, W. Al Nassan, K. Obaideen, M. N. Al Mallahi, Y. Mohammad, and H. M. El-Damanhoury, "Contemporary Role and Applications of Artificial Intelligence in Dentistry," *F1000Research* 12 (2023): 1179, https://doi.org/10.12688/f1000research.140204.1.

23. R. H. Putra, C. Doi, N. Yoda, E. R. Astuti, and K. Sasaki, "Current Applications and Development of Artificial Intelligence for Digital Dental Radiography," *Dentomaxillofacial Radiology* 51, no. 1 (2022): 20210197, https://doi.org/10.1259/dmfr.20210197.

24. A. Thurzo, W. Urbanová, B. Novák, et al., "Where Is the Artificial Intelligence Applied in Dentistry? Systematic Review and Literature Analysis," *Healthcare (Basel, Switzerland)* 10, no. 7 (2022): 1269, https://doi.org/10.3390/healthcare10071269.

25. R. Rokhshad, F. Nasiri, N. Saberi, et al., "Deep Learning for Age Estimation From Panoramic Radiographs: A Systematic Review and Meta-Analysis," *Journal of Dentistry* 154 (2025): 105560, https://doi.org/10.1016/j.jdent.2025.105560.

26. R. Rokhshad, H. Mohammad-Rahimi, J. B. Price, et al., "Artificial Intelligence for Classification and Detection of Oral Mucosa Lesions on Photographs: A Systematic Review and Meta-Analysis," *Clinical Oral Investigations* 28, no. 1 (2024): 88, https://doi.org/10.1007/s00784-023-05475-4.

27. H. S. Lee, S. Yang, J. Y. Han, et al., "Automatic Detection and Classification of Nasopalatine Duct Cyst and Periapical Cyst on Panoramic Radiographs Using Deep Convolutional Neural Networks," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology* 138, no. 1 (2024): 184–195, https://doi.org/10.1016/j.oooo.2023.09.012.

28. K. F. Hung, Q. Y. H. Ai, Y. Y. Leung, and A. W. K. Yeung, "Potential and Impact of Artificial Intelligence Algorithms in Dento-Maxillofacial Radiology," *Clinical Oral Investigations* 26, no. 9 (2022): 5535–5555, https://doi.org/10.1007/s00784-022-04477-y.

29. M. D. F. McInnes, D. Moher, B. D. Thombs, et al., "Preferred Reporting Items for a Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement," *Journal of the American Medical Association* 319, no. 4 (2018): 388–396, https://doi.org/10.1001/jama.2017.19163.

30. J. P. T. Higgins, J. Thomas, and J. Chandler, "Analysing Data and Undertaking Meta-Analyses," in *Cochrane Handbook for Systematic Reviews of Interventions*. 6.4, eds. J. J. Deeks, J. P. T. Higgins, and D. G. Altman (Cochrane, 2023).

31. J. R. Dettori, D. C. Norvell, and J. R. Chapman, "Fixed-Effect vs Random-Effects Models for Meta-Analysis: 3 Points to Consider," *Global Spine Journal* 12, no. 7 (2022): 1624–1626, https://doi.org/10.1177/21925682221110527.

32. Y. Wang, N. DelRocco, and L. Lin, "Comparisons of Various Estimates of the I2 Statistic for Quantifying Between-Study Heterogeneity in Meta-Analysis," *Statistical Methods in Medical Research* 33, no. 5 (2024): 745–764, https://doi.org/10.1177/09622802241231496.

33. M. Egger, G. D. Smith, M. Schneider, and C. Minder, "Bias in Meta-Analysis Detected by a Simple, Graphical Test," *British Medical Journal* 315, no. 7109 (1997): 629–634, https://doi.org/10.1136/bmj.315.7109.629.

34. S.-Y. Sim, J. Hwang, J. Ryu, H. Kim, E.-J. Kim, and J.-Y. Lee, "Differential Diagnosis of OKC and SBC on Panoramic Radiographs: Leveraging Deep Learning Algorithms," *Diagnostics (Basel, Switzerland)* 14, no. 11 (2024): 1144, https://doi.org/10.3390/diagnostics14111144.

35. Y. Kise, Y. Ariji, C. Kuwada, M. Fukuda, and E. Ariji, "Effect of Deep Transfer Learning With a Different Kind of Lesion on Classification Performance of Pre-Trained Model: Verification With Radiolucent Lesions on Panoramic Radiographs," *Imaging Science in Dentistry* 53, no. 1 (2023): 27–34, https://doi.org/10.5624/isd.20220133.

36. M. Rašić, M. Tropčić, P. Karlović, D. Gabrić, M. Subašić, and P. Knežević, "Detection and Segmentation of Radiolucent Lesions in the Lower Jaw on Panoramic Radiographs Using Deep Neural Networks," *Medicina (Kaunas, Lithuania)* 59, no. 12 (2023): 2138, https://doi.org/10.3390/medicina59122138.

37. B. Feher, U. Kuchler, F. Schwendicke, et al., "Emulating Clinical Diagnostic Reasoning for Jaw Cysts With Machine Learning," *Diagnostics (Basel, Switzerland)* 12, no. 8 (2022): 1968, https://doi.org/10.3390/diagnostics12081968.

38. S. Tajima, Y. Okamoto, T. Kobayashi, et al., "Development of an Automatic Detection Model Using Artificial Intelligence for the

Detection of Cyst-Like Radiolucent Lesions of the Jaws on Panoramic Radiographs With Small Training Datasets," *Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology* 34, no. 5 (2022): 553–560, https://doi.org/10.1016/j.ajoms.2022.02.004.

39. A. Lee, M. S. Kim, S. S. Han, P. G. Park, C. Lee, and J. P. Yun, "Deep Learning Neural Networks to Differentiate Stafne's Bone Cavity From Pathological Radiolucent Lesions of the Mandible in Heterogeneous Panoramic Radiography," *PLoS One* 16, no. 7 July (2021): e0254997, https://doi.org/10.1371/journal.pone.0254997.

40. Z. Liu, J. Liu, Z. Zhou, et al., "Differential Diagnosis of Ameloblastoma and Odontogenic Keratocyst by Machine Learning of Panoramic Radiographs," *International Journal of Computer Assisted Radiology and Surgery* 16 (2021): 415–422, https://doi.org/10.1007/s11548-021-02309-0.

41. H. Watanabe, Y. Ariji, M. Fukuda, et al., "Deep Learning Object Detection of Maxillary Cyst-Like Lesions on Panoramic Radiographs: Preliminary Study," *Oral Radiology* 37, no. 3 (2021): 487–493, https://doi.org/10.1007/s11282-020-00485-4.

42. O. Kwon, T. H. Yong, S. R. Kang, et al., "Automatic Diagnosis for Cysts and Tumors of Both Jaws on Panoramic Radiographs Using a Deep Convolution Neural Network," *Dentomaxillofacial Radiology* 49, no. 8 (2020): 20200185, https://doi.org/10.1259/dmfr.20200185.

43. J. H. Lee, D. H. Kim, and S. N. Jeong, "Diagnosis of Cystic Lesions Using Panoramic and Cone Beam Computed Tomographic Images Based on Deep Learning Neural Network," *Oral Diseases* 26, no. 1 (2020): 152–158, https://doi.org/10.1111/odi.13223.

44. H. Yang, E. Jo, H. J. Kim, et al., "Deep Learning for Automated Detection of Cyst and Tumors of the Jaw in Panoramic Radiographs," *Journal of Clinical Medicine* 9, no. 6 (2020): 1839, https://doi.org/10.3390/jcm9061839.

45. Y. Ariji, Y. Yanashita, S. Kutsuna, et al., "Automatic Detection and Classification of Radiolucent Lesions in the Mandible on Panoramic Radiographs Using a Deep Learning Object Detection Technique," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology* 128, no. 4 (2019): 424–430, https://doi.org/10.1016/j.oooo.2019.05.014.

46. U. Committeri, S. Barone, A. Arena, et al., "New Perspectives in the Differential Diagnosis of Jaw Lesions: Machine Learning and Inflammatory Biomarkers," *Journal of Stomatology, Oral and Maxillofacial Surgery* 125, no. S4 (2024): 101912, https://doi.org/10.1016/j.jormas.2024.101912.

47. Y. Song, S. Ma, B. Mao, et al., "Application of Machine Learning in the Preoperative Radiomic Diagnosis of Ameloblastoma and Odontogenic Keratocyst Based on Cone-Beam CT," *Dentomaxillofacial Radiology* 53, no. 5 (2024): 316–324, https://doi.org/10.1093/dmfr/twae016.

48. W. Liu, X. Li, C. Liu, et al., "Automatic Classification and Segmentation of Multiclass Jaw Lesions in Cone-Beam CT Using Deep Learning," *Dentomaxillofacial Radiology* 53, no. 7 (2024): 439–446, https://doi.org/10.1093/dmfr/twae028.

49. P. K. Shrivastava, S. Hasan, L. Abid, R. Injety, A. K. Shrivastav, and D. Sybil, "Accuracy of Machine Learning in the Diagnosis of Odontogenic Cysts and Tumors: A Systematic Review and Meta-Analysis," *Oral Radiology* 40, no. 3 (2024): 342–356, https://doi.org/10.1007/s11282-024-00745-7.

50. G. Krishnan, S. Singh, M. Pathania, et al., "Artificial Intelligence in Clinical Medicine: Catalyzing a Sustainable Global Healthcare Paradigm," *Frontiers in Artificial Intelligence* 6 (2023): 1227091, https://doi.org/10.3389/frai.2023.1227091.

51. T. D. Pham, M. T. Teh, D. Chatzopoulou, S. Holmes, and P. Coulthard, "Artificial Intelligence in Head and Neck Cancer: Innovations, Applications, and Future Directions," *Current Oncology* 31, no. 9 (2024): 5255–5290, https://doi.org/10.3390/curroncol31090389.

52. N. N. Zhong, H. Q. Wang, X. Y. Huang, et al., "Enhancing Head and Neck Tumor Management With Artificial Intelligence: Integration and Perspectives," *Seminars in Cancer Biology* 95 (2023): 52–74, https://doi.org/10.1016/j.semcancer.2023.07.002.

53. S. A. Alowais, S. S. Alghamdi, N. Alsuhebany, et al., "Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice," *BMC Medical Education* 23, no. 1 (2023): 689, https://doi.org/10.1186/s12909-023-04698-z.

54. M. M. Chen, A. Terzic, A. S. Becker, et al., "Artificial Intelligence in Oncologic Imaging," *European Journal of Radiology Open* 9 (2022): 100441, https://doi.org/10.1016/j.ejro.2022.100441.

55. M. M. Chen, L. P. Golding, and G. N. Nicola, "Who Will Pay for AI?," *Radiology: Artificial Intelligence* 3, no. 3 (2021): e210030, https://doi.org/10.1148/ryai.2021210030.

56. F. Yu, A. Moehring, O. Banerjee, T. Salz, N. Agarwal, and P. Rajpurkar, "Heterogeneity and Predictors of the Effects of AI Assistance on Radiologists," *Nature Medicine* 30, no. 3 (2024): 837–849, https://doi.org/10.1038/s41591-024-02850-w.

57. C. Chakraborty, M. Bhattacharya, S. Pal, and S.-S. Lee, "From Machine Learning to Deep Learning: Advances of the Recent Data-Driven Paradigm Shift in Medicine and Healthcare," *Current Research in Biotechnology* 7 (2024): 100164, https://doi.org/10.1016/j.crbiot.2023.100164.

58. N. Nawaz, H. Arunachalam, B. K. Pathi, and V. Gajenderan, "The Adoption of Artificial Intelligence in Human Resources Management Practices," *International Journal of Information Management Data Insights* 4, no. 1 (2024): 100208, https://doi.org/10.1016/j.jjimei.2023.100208.

59. A. Haakenstad, C. M. S. Irvine, M. Knight, et al., "Measuring the Availability of Human Resources for Health and Its Relationship to Universal Health Coverage for 204 Countries and Territories From 1990 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019," *Lancet* 399, no. 10341 (2022): 2129–2154, https://doi.org/10.1016/S0140-6736(22)00532-3.

60. P. Mikalef and M. Gupta, "Artificial Intelligence Capability: Conceptualization, Measurement Calibration, and Empirical Study on Its Impact on Organizational Creativity and Firm Performance," *Information & Management* 58, no. 3 (2021): 103434, https://doi.org/10.1016/j.im.2021.103434.

61. N. Haefner, J. Wincent, V. Parida, and O. Gassmann, "Artificial Intelligence and Innovation Management: A Review, Framework, and Research Agenda☆," *Technological Forecasting and Social Change* 162 (2021): 120392, https://doi.org/10.1016/j.techfore.2020.120392.

62. D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, and R. K. Tekade, "Artificial Intelligence in Drug Discovery and Development," *Drug Discovery Today* 26, no. 1 (2021): 80–93, https://doi.org/10.1016/j.drudis.2020.10.010.

63. E. Ullah, A. Parwani, M. M. Baig, and R. Singh, "Challenges and Barriers of Using Large Language Models (LLM) Such as ChatGPT for Diagnostic Medicine With a Focus on Digital Pathology—A Recent Scoping Review," *Diagnostic Pathology* 19, no. 1 (2024): 43, https://doi.org/10.1186/s13000-024-01464-7.

64. X. Meng, X. Yan, K. Zhang, et al., "The Application of Large Language Models in Medicine: A Scoping Review," *iScience* 27, no. 5 (2024): 109713, https://doi.org/10.1016/j.isci.2024.109713.

65. D. B. Olawade, A. C. David-Olawade, O. Z. Wada, A. J. Asaolu, T. Adereni, and J. Ling, "Artificial Intelligence in Healthcare Delivery: Prospects and Pitfalls," *Journal of Medicine, Surgery, and Public Health* 3 (2024): 100108, https://doi.org/10.1016/j.glmedi.2024.100108.