

Diagnosis of Oral and maxillofacial cysts using artificial intelligence: a Literature Review

Manuscript Info

Manuscript History

Received: xxxxxxxxxxxxxxxxx

Final Accepted: xxxxxxxxxxxxx

Published: xxxxxxxxxxxxxxxxx

Key words:-

Oral cysts, Maxillofacial cysts, Cavities,
Asymptomatic

Abstract

Oral and maxillofacial cysts are cavities that can pose significant risks if not detected and treated promptly. Many of these cysts are asymptomatic, often going unnoticed until complications arise. The introduction of artificial intelligence (AI) presents a promising opportunity for early detection and management of these cysts. To explore current studies on the use of artificial intelligence in diagnosing oral and maxillofacial cysts. To examine the existing literature in this field, assess the accuracy, effectiveness, and limitations of AI models, and identify challenges in implementing AI in clinical practice. This literature review followed a systematic approach, identifying 223 studies from PUBMED and SCOPUS databases between 1975 and 2024. After applying inclusion and exclusion criteria, 26 retrospective cohort studies were included in the final analysis. A risk of bias assessment was conducted using the ROBINS I tool. The investigation revealed that AI models consistently demonstrate high accuracy in detecting oral cysts in both radiographs and digital histopathology. The ROBINS I tool indicated a moderate risk of bias in most of the included studies. Notable limitations include limited datasets, variable data quality, and a lack of explainability in AI models results. AI models have shown considerable effectiveness and speed in detecting both simple and complex cysts. However, to fully leverage AI's potential in clinical settings, further rigorous studies are needed to evaluate its risks, benefits, and feasibility, ensuring compliance with governmental health regulations on AI.

Copy Right, IJAR, 2019,. All rights reserved.

Introduction:-

Maxillofacial cysts can be defined as pathologic cavities lined by epithelium and generally containing fluid or semi-solid material. One exceptional subgroup, known as pseudocysts, lacks an epithelial lining and possesses various diagnostic considerations

¹. The lesions possess great heterogeneity in origin, biologic behavior, and clinical presentation, and therefore they are one of the significant challenges in diagnosis and also treatment. Epidemiological data of a United Kingdom-based retrospective study, spanning over 55,000 specimens, showed that odontogenic cysts accounted for approximately 12.8% of the cases, with a higher occurrence in males and in the age group of 30 to 60 years². The most prevalent types are radicular cysts, dentigerous cysts, and odontogenic keratocysts (OKCs). Diagnosis is paramount to effective clinical management and is typically founded upon a combination of patient history, clinical inspection, radiographic imaging, and histopathological examination³.

Maxillofacial cyst classification has changed considerably over time. Initial efforts by Bland-Sutton in 1888 laid the foundation for systematic classification, building on which World Health Organization (WHO) classifications, starting in 1971, 1992, 2005, and more recently in 2017⁸, took place. The present WHO system reclassified odontogenic cysts and grouped jaw lesions based on contemporary understanding of their pathogenesis. Odontogenic cysts are grouped into inflammatory and developmental types. Inflammatory cysts typically arise from pulpal infections, whereas developmental cysts form during odontogenesis³.

Radicular cysts, the most common inflammatory type, are formed due to necrotic pulp tissue that irritates epithelial cell rests^{4,5}. Root canal treatment or extraction with following curettage is commonly performed clinically. Residual cysts may persist when radicular cysts are incompletely removed³, while paradental cysts, commonly associated with third molars, share the same histological features as radicular cysts⁴. Among developmental cysts, dentigerous cysts are most frequent, most commonly associated with unerupted mandibular third molars. The lesions are radiographically unilocular but may become multilocular and displace neighboring structures. Management involves enucleation and histopathological analysis to eliminate the possibility of neoplastic transformation^{4,6,7}.

Odontogenic keratocysts (OKCs) are highly aggressive with a strong tendency for recurrence. Having been previously reclassified as tumors due to mutations in the PTCH gene, they were reclassified as cysts in the 2017 WHO update^{8,9}. OKCs predominantly appear in Caucasian males aged 20-40 years and are most commonly found in the posterior mandible¹⁰. Histologically, they are characterized by a thin epithelial lining, palisaded basal cells, and a keratin-filled lumen, some of the reasons why they have a strong tendency to recur¹¹. OKCs are also associated with Gorlin-Goetz syndrome, a genetic condition with multiple cysts, basal cell carcinomas, and skeletal defects^{12,13}.

Nonodontogenic cysts, which arise from non-tooth-forming epithelium, include nasopalatine duct cysts and neonatal palatal cysts¹⁴. Nasopalatine duct cysts are typically heart-shaped radiolucencies in the anterior maxilla and are histologically confirmed by the existence of epithelial-lined fibrous tissue. They are generally curative with surgical removal and have minimal chance of recurrence. Neonatal palatal cysts, Epstein pearls, and Bohn nodules are small, asymptomatic, and resolve spontaneously. These are histologically keratin-filled epithelial cysts¹⁵.

Despite histopathology as the gold standard for diagnosis, there are also several challenges. Radiographic overlap is common; OKCs and ameloblastomas may be indistinguishable, for instance, while nasopalatine duct cysts may be mistaken for periapical pathology^{16,17}. Dentigerous cysts may also present as OKCs on imaging examinations¹⁸. Histopathological ambiguity also worsens the diagnosis, with features such as mucous cells and keratinisation occasionally overlapping among lesions, necessitating additional staining techniques for successful differentiation^{19,20}. Further, inaccurate diagnoses by less experienced practitioners underscore the need for technological assistance tools, including artificial intelligence, to assist in enhancing diagnostic precision and preventing misclassification^{17,21}.

Artificial Intelligence (AI) refers to the capacity of computers and machines to perform operations that traditionally require human intelligence, such as reasoning, learning, decision-making, and pattern recognition²². The theoretical foundation of AI dates back to Alan Turing's seminal work in 1950 and was formally put forward by John McCarthy in 1956²³. AI is now a multidisciplinary field of research with applications in medicine, engineering, and social sciences. At the heart of AI is the usage of algorithms—computational methods enabling machines to accept inputs and generate outputs. Some of the most innovative AI developments are Machine Learning (ML) and Deep Learning (DL), which enable systems to learn from data and improve with experience.

Machine Learning encompasses supervised learning (learning from labeled examples), unsupervised learning (identifying patterns in unlabeled examples), and reinforcement learning (optimization via trial-and-error). Deep Learning, an ML branch, involves artificial neural networks with multiple layers based on the anatomy of the human brain. DL models are particularly good at image classification and recognition tasks and, therefore, are highly appropriate for analyzing radiographs and histological slides in the clinical setting²⁴. The second key component is Natural Language Processing (NLP) which enables machines to understand and process human language with applications in clinical documentation, chatbots, and automated diagnostic reporting. AI systems range from Artificial Narrow Intelligence (ANI) which are task-specific to the theoretical constructs of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI) which attempt to equalize or exceed human intellectual capabilities.

AI in healthcare is a developing decision support system that has the potential to deliver faster diagnosis, reduced human error, and personalized treatment planning. Its use in maxillofacial cyst diagnosis, in particular, holds great promise because of the diagnostic dilemma and overlapping features of these lesions. AI models were employed with panoramic radiographs, cone-beam computed tomography (CBCT), and magnetic resonance imaging (MRI) for enhancing the detection and classification of lesions. Algorithms such as YOLOv2 (You Only Look Once) have proven effective in the detection of cystic lesions such as dentigerous cysts, OKCs, and ameloblastomas by detecting shape, size, and anatomical location features²⁵. Similarly, DetectNet, a CNN-based framework, uses bounding boxes to delineate lesion borders and predict classification²⁶. These models have demonstrated high sensitivity and specificity, with performance better than that of general practitioners and even some specialists, with real-time feedback that is compatible with both clinical workflow and tele-dentistry platforms.

AI also helps minimize inter-observer variability, a well-known problem in radiographic interpretation, especially between less experienced clinicians. In histopathology, AI helps ease the diagnostic bottleneck by automating image classification. Deep learning models such as VGG16, pre-trained on ImageNet, have been fine-tuned to analyze histological oral lesion slides with high accuracy²⁷. These systems are able to pick up subtle histopathological differences between cystic lesions that are similar in appearance, such as distinguishing keratin-filled OKCs from dentigerous cysts or identifying proliferative epithelial features of neoplastic transformation. Rather than replacing pathologists, AI software is an adjunct that brings areas of suspicion to the attention of the pathologist, triages slides for review, and assists in decision-making in equivocal cases.

Clinical deployment of AI integration has several advantages. Lesion detection, even at asymptomatic stages, can be facilitated by automated screening. AI algorithms can also assist treatment planning by measuring lesional volume and determining proximity to vital anatomical structures to inform surgery. AI can function as a triage system in resource-scarce settings, guiding referrals and coordinating specialist intervention. Moreover, AI-driven platforms are didactic and can be incorporated into training modules to enhance pattern recognition in novice clinicians.

Looking ahead, the future of AI in maxillofacial pathology lies in building larger annotated datasets and multi-center research that enhances model accuracy and generalizability. Future models may integrate radiographic, histologic, and genomic data to provide multidimensional diagnostic possibilities.

Despite the potential benefits of artificial intelligence (AI) in medicine, AI in oral and maxillofacial pathology must transcend several limitations. One of the key challenges is the absence of big annotated datasets required to build robust diagnostic models. In oral pathology, such datasets continue to be limited, hindering AI algorithm development and generalizability²⁸. Privacy and ethical concerns are additional significant challenges. AI software must satisfy stringent data protection regulations, particularly when dealing with sensitive patient information. In the United Kingdom, regulators such as the Medicines and Healthcare products Regulatory Agency (MHRA) and the NHS AI Lab ensure that AI applications are consistent with ethical and legal requirements^{29,30}. Another essential concern is the risk of overdependence on AI systems. While AI can enhance diagnostic accuracy and speed, it cannot replace clinical judgment. Instead, it must be constructed as a supportive tool that complements the intelligence of medical practitioners²⁸.

Based on these findings, it is apt and essential to undertake an exhaustive literature review of AI application in oral and maxillofacial cyst diagnosis. To date, no review has systematically worked towards the utilization of both machine learning (ML) and deep learning (DL) approaches in the detection of such lesions in radiographic and histopathological modalities. The current review seeks to fill a significant knowledge gap in the literature by summarizing available knowledge, evaluating technological progress, and establishing directions for future studies.

The purpose of this literature review is to discuss the current state and recent advances in the application of AI—more precisely ML and DL—for diagnosing oral and maxillofacial cysts. The main goals are to review current studies that apply AI-driven methods and tools in this field, and to evaluate the accuracy, efficacy, and limitations of these algorithms in identifying different cystic lesions. A secondary objective is to identify persisting gaps and implementation challenges in AI integration, and to provide recommendations that can guide future research and clinical practice.

Methods

A systematic approach was employed to ensure methodological rigour and accountability in selection, analysis, and identification of included studies within this review. The process adhered to standards of literature synthesis, from which there were well-defined eligibility criteria and systematic data extraction protocols.

Included studies were randomized controlled trials and cohort studies that investigated the application of artificial intelligence (AI) i.e. machine learning (ML) or deep learning (DL) in the detection or diagnosis of oral and maxillofacial cysts. To be included, studies had to have used AI-based methods in a diagnostic role and been English-language publications from January 1975 through to 15 June 2024. Studies were excluded if they were review articles, case reports, letters, editorials, conference abstracts, book chapters, or other non-research-based formats. Participants for eligible studies were patients who presented with oral cysts or cyst-like lesions found through radiographic evaluation or histopathological diagnosis. Interventions of interest were AI-driven diagnostic algorithms applied on radiographic (for example, panoramic radiographs, CBCT) or histopathological samples. Reference groups consisted of reference standards like previously established confirmed diagnosis or expert clinician opinion. Primary outcomes of interest measured included diagnostic accuracy metrics like sensitivity, specificity, precision, and F1-score.

Systematic search approach was conducted in two leading databases: PubMed and Scopus. Scopus search query was ('artificial intelligence' OR 'machine learning' OR 'deep learning') AND ('diagnosis' OR 'diagnostic imaging' OR 'detection') AND ('oral cysts' OR 'jaw cysts' OR 'odontogenic cysts'). The search query for PubMed was: (('artificial intelligence' OR 'machine learning' OR 'deep learning' OR 'neural networks') AND ('maxillofacial cysts' OR 'jaw lesion' OR 'oral cysts' OR 'odontogenic cysts')) AND ('diagnosis' OR 'detection'). The search was narrowed to studies published within the specified timeframe and language needs.

Independent study selection was performed by two reviewers. Titles and abstracts were screened for potentially eligible studies, and full texts were retrieved in studies with inclusion criteria or where it was not possible to determine from the abstract alone. As no randomized controlled trials were identified, the final decision included retrospective cohort studies. Data were independently withdrawn in a pre-standardized form with the intent to record main study characteristics. Extracted variables included author(s), year of publication, study design, type of applied AI, cyst classification, imaging modality, methods of validation, and diagnostic results reported.

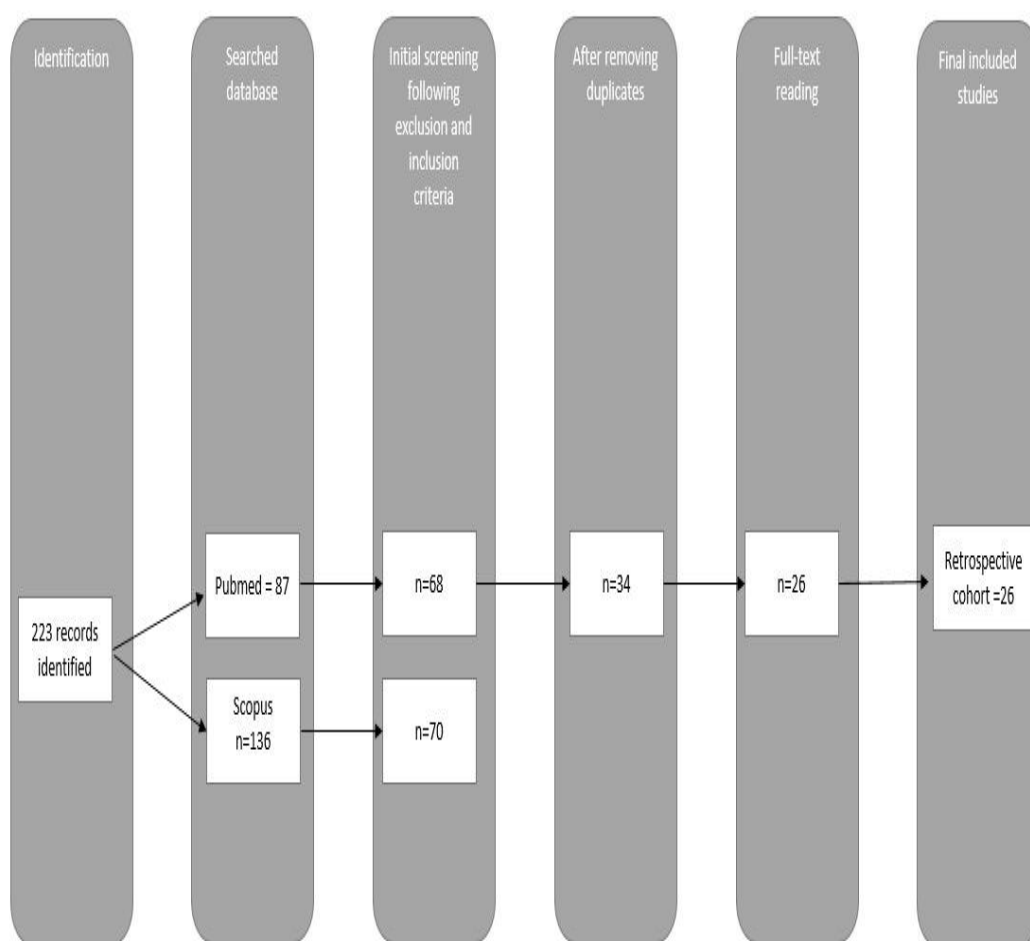
Methodological quality and potential bias within each study that was included were assessed using the Risk Of Bias In Non-randomized Studies - of Interventions (ROBINS-I) tool³¹. The instrument evaluates seven domains: confounding, participant selection, intervention classification, deviations from allocated interventions, missing data, outcome measurement, and reporting selection. Each domain was assigned a low, moderate, or critical risk of bias score. Tabulation of judgments regarding risk in all the included studies is shown in Appendix 2A. This systematic evaluation yielded methodologically sound data on which the review results were derived, allowing for critical evaluation of the validity and reliability of AI instruments for diagnosing oral cysts.

223 articles were initially identified through a systematic search of two leading databases—PubMed and Scopus. Following the preselected inclusion and exclusion criteria, full-text screening, and the removal of duplicate data, the final list comprised 26 articles. All included studies were retrospective observational studies that investigated the application of artificial intelligence (AI) in diagnosing or detecting oral and maxillofacial cysts.

The Scopus search yielded 136 articles. These were filtered out at first by selecting English-language research articles and excluding non-research article types such as conference abstracts, book chapters, mini reviews, case

reports, and editorials. This left the pool narrowed to 111 articles. Further filtering based on subject categories of interest to medicine and dentistry left the selection at 70 articles. This was followed by a critical assessment of titles and abstracts, and subsequent inclusion of 12 articles found to be eligible for the study.

PubMed search generated 87 articles. With identical inclusion criteria—English-language research articles—and non-research articles excluded, the number of research articles eligible was reduced to 68. Relevance of articles to the study's purpose was assessed through title and abstract review, and 25 articles were selected accordingly. After cross-matching and duplicate elimination between the two databases, 26 unique articles remained for final analysis. Collectively, these studies provide a core data set from which the value of AI-based approaches to the diagnostic practice of oral and maxillofacial cysts can be evaluated.



The selection process for the studies is summarized in **Figure 1**.

Literature Review

The reviewed studies altogether demonstrate the growing application of artificial intelligence (AI) to diagnose jaw cysts and related lesions through panoramic radiographs (OPG). The five studies employed similar retrospective

study designs and various machine learning (ML) or deep learning (DL) models with varying validation techniques from hold-out testing to cross-validation.

Berne et al.³² developed a DL model with YOLOv3 to differentiate radicular cysts (RC) and periapical granulomas (GC). Using a 10-fold cross-validation approach to OPG images, the study had 80 RC cases, 72 GC cases, 197 normal images, and some other radiolucent controls. The model was good for RC detection with 1.00 sensitivity, 0.95 specificity, and an AUC of 0.97. In the case of GC, sensitivity was slightly lower at 0.77 but specificity remained flawless at 1.00 with an AUC of 0.88. These results validate the accuracy of the model in distinguishing between granulomatous and inflammatory lesions.

Building upon this, Yu et al.³³ employed a Unet-based deep learning model to distinguish odontogenic cysts and tumors, i.e., periapical cysts (PC), dentigerous cysts (DC), odontogenic keratocysts (OKC), and ameloblastomas (AB). The collection of samples comprised 87 lesion cases and 200 normal samples. Using a hold-out validation test, the model was 90.66% accurate, 85.23% precise, 84.27% sensitive, 93.50% specific, and 84.74% F1 score. These metrics represent how robust the model is in distinguishing pathological from normal images, particularly in mixed-lesion datasets.

In another approach, Kumar et al.³⁴ examined ML techniques—specifically support vector machines (SVM)—for dental cyst, tumor, and abscess identification. The study utilized 172 lesion images and hold-out validation. The three feature extraction methods were compared: GLCM, wavelet analysis, and GLRLM. GLCM had the highest accuracy of 98%, while GLRLM yielded 95% and wavelet analysis yielded 91%. AUC values testified to the superior diagnostic capability of GLCM, highlighting the importance of texture-based features in radiographic classification.

Feher et al.³⁵ broadened the scope by simulating clinical reasoning using RetinaNet, a ML model that was trained on 855 OPG images and validated with 384. The model was able to differentiate between odontogenic and non-odontogenic cysts. For odontogenic cysts, sensitivity was 0.84 and specificity was 0.59, while for non-odontogenic cysts, sensitivity dropped to 0.56 but specificity rose to 0.84. Compared to a human control group, the AI model had higher sensitivity for both groups, although specificity varied. The research illustrates the potential of AI in assisting diagnostic reasoning, especially in uncertain cases.

Also supporting these findings, Watanabe et al.³⁶ conducted a preliminary study using DetectNet for maxillary cyst-like lesion and benign tumor detection. The patient dataset was 412 cases with 323 RC, 37 DC, 23 OKC, 3 AB, 2 odontogenic myxomas, and 1 adenomatoid odontogenic tumor. The model was validated in two steps. In the first step, recall was 74.6%, precision 89.8%, and F1 score 81.5%. In the second step, recall was 77.1%, precision 90.0%, and F1 score 83.1%. These results indicate consistent performance on datasets and validate the applicability of DL for multi-class lesion detection.

Liu et al.³⁷ explored the use of machine learning (ML) algorithms to discriminate between ameloblastomas and odontogenic keratocysts (OKCs) on panoramic radiographs. Knowing that these lesions have overlapping diagnoses, the authors demonstrated that ML could learn to identify radiographic features successfully in an attempt to assist clinicians to accurately classify them. By utilizing model training on annotated datasets, promising diagnostic accuracy was achieved, which showed that AI can assist clinicians in differentiating between lesionally similar radiographic-appearing lesions.

Following this radiographic foundation, Cai et al.²¹ progressed to digital pathology by developing AI models to differentially diagnose and predict the prognosis of sporadic OKCs on haematoxylin and eosin-stained slides. The approach merged image-based deep learning with prognostic modeling to not just diagnose lesions but also to forecast recurrence risk. This dual-purpose utility renders AI an effective tool in both diagnostic as well as treatment planning procedures, especially for aggressive behavior lesions like OKCs.

In their corresponding effort to counteract diagnostic uncertainty, Lee et al.³⁸ resolved a common conundrum by deep learning neural networks to separate Stafne's bone cavity—a benign anatomical variation—from pathological radiolucent mandibular lesions. In their proof on heterogeneous panoramic radiographs, the authors demonstrated

DL models can significantly differentiate between normal anatomy and disease, precluding false positives and unwarranted procedures.

Scaling AI application in histopathology, Rao et al.²⁷ developed a DL-driven microscopic diagnostic system for OKCs and non-keratocysts on the basis of histological slides. Trained on haematoxylin and eosin-stained incisional biopsies, the model was extremely accurate in classifying cyst types. The study highlighted the potential of AI to streamline histopathological processes, highlight uncertain cases, and support pathologists in high-volume diagnostic settings.

Similarly, Yang et al.²⁵ employed deep learning in automating detection of jaw cysts and tumors from panoramic radiographs. Their model performed well for a variety of lesions, i.e., dentigerous cysts, OKCs, and ameloblastomas. Employing convolution neural networks, the model was able to delineate the borders of lesions and make suggestions for classifications and offer real-time aid in radiographic interpretation.

Adding additional importance to anatomical precision, Rašić et al.³⁹ focused on segmentation tasks by deep neural networks to locate and define radiolucent lesions in the lower jaw. From panoramic radiographs, their model was very accurate in detecting lesion margins, which is crucial in surgical planning and follow-up. This study highlights the prospect of AI in enhancing image-based diagnosis by precise anatomical localization.

For the early detection of disease, Endres et al.⁴⁰ developed a DL algorithm specific for the identification of periapical disease from dental radiographs. The model, which was trained to identify subtle radiolucencies of pathologic structures, was very accurate for diagnosis. All these results validate the use of AI in everyday dental practice, particularly for the early detection of apical lesions, which, otherwise, remain undiagnosed.

Following the differential diagnosis thread, Li et al.⁴¹ used a deep learning network to differentiate ameloblastomas from OKCs on panoramic radiographs. Tested against expert scores and learned on a big data set, the model proved that AI could match or exceed human ability at recognizing such clinically significant lesions.

Kwon et al.⁴² extended this by developing a DL model of YOLOv3 to classify odontogenic cysts and ameloblastomas of 1,282 panoramic radiographs automatically. In the database, there were 302 RC, 350 DC, 300 OKC, 230 AB, and 100 normal jaw views. Without augmentation, sensitivity of the model was 78.2%, specificity was 93.9%, and total accuracy was 91.3%, with an AUC of 0.86. Performance was greatly improved with augmentation, yielding sensitivity of 88.9%, specificity of 97.2%, accuracy of 95.6%, and AUC of 0.94—highlighting the way that training advances are advantageous to model precision.

Complementing this, Bispo et al.⁴³ sought to differentiate between OKCs and ameloblastomas using MDCT images and a Google Inception v3 convolutional neural network. The dataset consisted of 350 images of 18 OKC and 22 ameloblastoma cases. By using five-fold cross-validation, the model established stable accuracy in iterations of 90.16% to 92.48%, showcasing the reliability of DL in CT-based lesion classification.

By employing a multimodal approach, Committeri et al.⁴⁴ blended ML models—logistic regression (LR) and Fourier transform (FT)—with inflammatory biomarkers and CBCT imaging to classify dentigerous cysts, OKCs, and unicystic ameloblastomas. For 103 patients, biomarkers alone yielded less than 50% accuracy. But when integrated with imaging, LR was at 95% accuracy with an AUC of 0.96, and FT was at 94.3% accuracy with an AUC of 0.95, reinforcing the benefit of data fusion towards enhanced diagnostic performance.

Yilmaz et al.⁴⁵ also validated ML's variability by comparing various models like SVM, Naive Bayes, decision tree, random forest, and neural networks for diagnosing periapical cysts and OKCs from 50 CBCT images. The performance of SVM model was superior in three validation methods: ten-fold cross-validation yielded accuracy and F1 score of 100%; split-sample validation yielded accuracy and F1 score of 96%; and leave-one-out cross-validation yielded 94% accuracy and an F1 score of 93.88%, which testified to the strength of SVM for lesion detection based on CBCT.

In a comparative study, Lee et al.⁴⁶ employed Google Inception v3 for AI-based diagnosis of periapical cysts, DCs, and OKCs from OPGs and CBCT scans. The database comprised 2,126 images, 1,140 OPGs, and 986 CBCTs. The model performed higher accuracy on CBCT with an AUC of 0.914, sensitivity of 96.1%, and specificity of 77.1%. For OPG, the AUC was 0.87, sensitivity was 88.2%, and specificity was 77.0%, indicating that CBCT offers superior diagnostic precision for AI models.

For the case of anterior maxillary lesions, Kise et al.¹⁷ developed a DL system using DetectNet and Inception v3 to discriminate nasopalatine duct cysts (NDC) from radicular cysts (RC) with 300 panoramic radiographs. The data set included 100 NDC, 100 RC, and 100 normal ones. The DL system achieved a total accuracy of 0.88, higher than dental residents with 0.77. For NDC, the model achieved recall of 0.83, precision of 0.92, and F1 score of 0.87. For RC, recall was 0.85, precision 0.94, and F1 score 0.89. Normal images achieved recall of 0.95, precision of 0.79, and F1 score of 0.86, commenting on the good performance of the model by categories.

Lee et al.⁴⁷ conducted a comparative study of five DL models, namely EfficientDet-D3, Faster R-CNN, YOLO v5, RetinaNet, and SSD, to identify and classify nasopalatine duct cysts and periapical cysts from 1,269 panoramic radiographs. The data set comprised 603 RC, 606 NDC, and 60 normal images. EfficientDet-D3 outperformed all the other models, with mean average precision (mAP) of 93.8%, accuracy of 94.4%, sensitivity of 94.4%, specificity of 97.2%, and predictive values greater than 94%. The results reassert EfficientDet-D3's superiority in undertaking complex radiographic differentiation tasks.

Ariji et al.²⁶ suggested the use of a deep learning model with DetectNet for the automatic detection and classification of radiolucent lesions in the mandible. The collection had 210 panoramic radiographs and comprised 68 radicular cysts (RC), 66 dentigerous cysts (DC), 33 odontogenic keratocysts (OKC), 31 ameloblastomas (AB), and 12 simple bone cysts (SBC). Two independent test sets were used to test and validate the model. In both, sensitivity was always 0.88, while the false positive rate was remarkably low—0.00 per image in the first test and 0.04 in the second—demonstrating the accuracy and reliability of the model for multi-class lesion detection.

Taking AI to histopathology, Mohanty et al.⁴⁸ developed an automation pipeline for the discrimination of sporadic OKCs and non-keratocysts from WSI. The study utilized 113 slides, 48 OKCs, 20 DCs, and 37 RCs. With P-C-ReliefF algorithm and hold-out validation, the model had an AUC of 0.97 and overall accuracy of 97.4%, managing over 128,000 parameters. Such results reflect the scalability of AI in high-resolution slide interpretation and its capability to assist digital pathology workflows.

Frydenlund et al.⁴⁹ utilized machine learning algorithms—support vector machines (SVM) and Bayesian logistic regression (BLR)—for the identification of four developmental odontogenic cysts based on micrographic digital images. The dataset involved 73 slides: 20 DCs, 20 OKCs, 20 lateral periodontal cysts (LPCs), and 13 glandular odontogenic cysts (GOCs). BLR outperformed SVM in all three experiments. In the first experiment, it was 95.4% compared to 92.3% for SVM. In cross-validation, BLR was 90% and SVM was 87.9%. Without DCs, SVM gave 100% accuracy and BLR 96.78%, indicating the models' adaptability and robustness with various lesion types.

Tajima et al.⁵⁰ overcome the issue of limited training data by developing a deep learning model based on YOLOv3 for the detection of cyst-like radiolucent lesions on panoramic radiographs. 7,160 images were employed to train the model, 100 for validating it, and 100 for testing it. Even with the small test set, the model worked incredibly well: 98.3% accuracy, 94.4% sensitivity, 99.7% specificity, 99.0% precision, and F-score of 0.966. All these prove the strength of high-quality AI models even in conditions where there is limited data.

These radiographic studies have been supplemented by Setzer et al.⁵¹, who designed a DL model based on the Unet architecture to detect periapical lesions in CBCT images. Using a training set of 20 CBCT images and five-fold cross-validation, the model achieved 0.93 accuracy, 0.88 specificity, positive predictive value of 0.87, and negative predictive value of 0.93. These results establish the excellence of the model for the detection of fine apical pathoses in three-dimensional imaging.

Ekert et al.⁵² conducted a high-throughput analysis of 2,001 panoramic radiographs for apical lesion detection using deep learning. Model validation was conducted with 10-time repeated group shuffling with 0.85 (± 0.04) AUC.

Sensitivity was moderate at 0.65 (± 0.12), but specificity was up to 0.87 (± 0.04). The positive predictive value was 0.49 (± 0.10), but the negative predictive value was extremely high at 0.93 (± 0.03), indicating good performance in excluding disease and enabling early detection.

Results

Demographics and Population

As seen in Figure 2. 26 retrospective studies on the use of AI in diagnosing oral and maxillofacial cysts were conducted across 16 countries. China and South Korea led the research efforts with five studies each. India followed with three studies. Japan and Germany both contributed with two studies each. Other countries, including Croatia, Belgium, Austria, the Netherlands, Brazil, Italy, Turkey, Canada, and the USA, each had one published study.

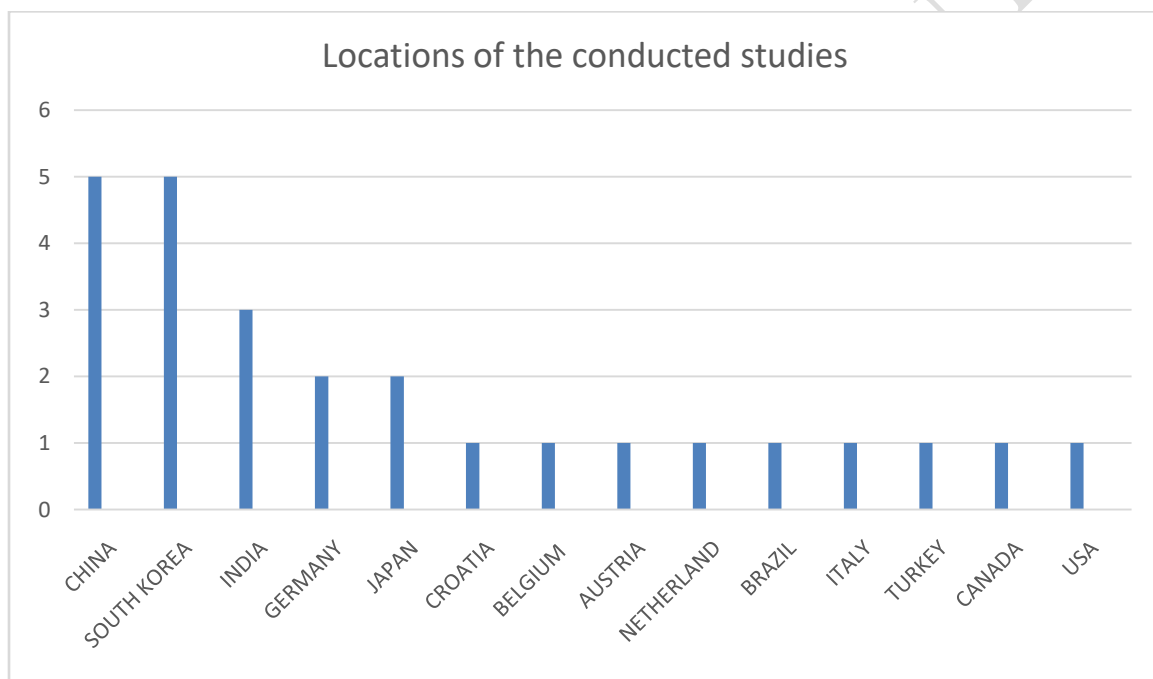


Figure 2. location of the conducted studies

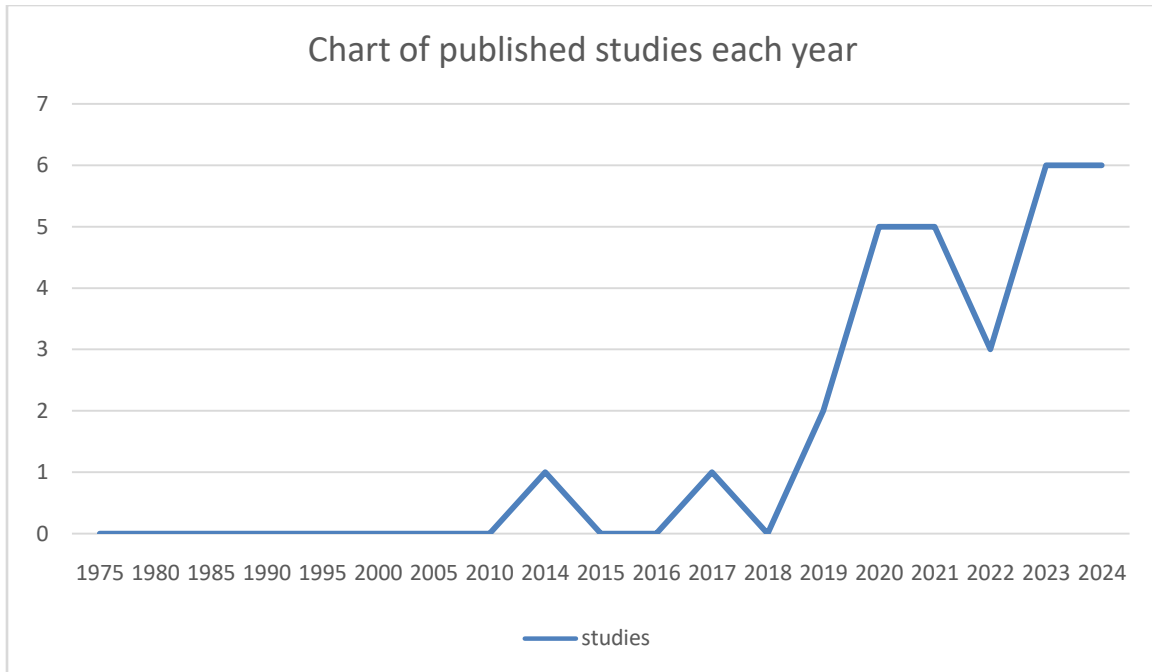


Figure 3. Years of published studies

We began our search from 1975. However, in 2014, we have only found a single study from that year. There was a gap until 2017, which yielded another study. In 2019, two studies were found. A noticeable increase in research occurred between 2020 and 2024, with five studies published in both 2020 and 2021. In 2022, we identified three studies. The highest number of articles, six, were found in 2024. Explained on Figure 3.

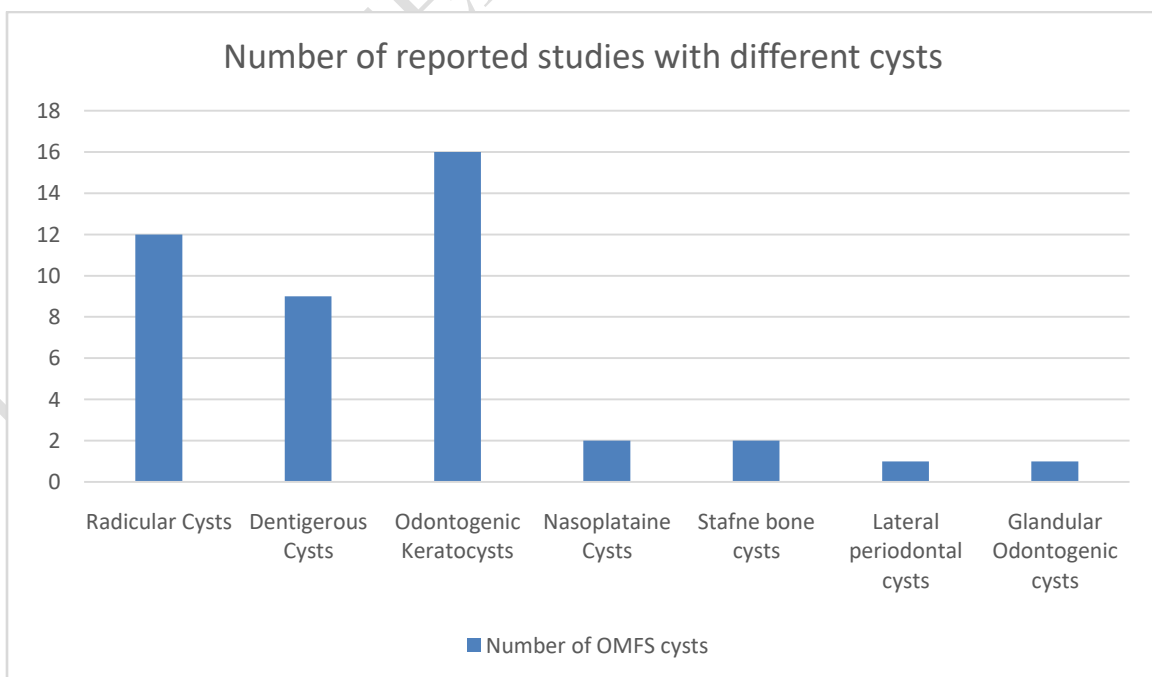


Figure 4. Number of reported cysts in included studies

In our research, the most frequently reported type of cyst was OKC, with 16 studies. Radicular cysts followed with 12 studies, and dentigerous cysts were reported in nine studies. Nasopalatine duct cysts and Stafne bone cysts each had two studies. Lateral periodontal cysts and glandular odontogenic cysts were the subjects of only one study each. Figure 4 illustrates this breakdown.

Table 1. Accuracy rate of Deep Learning and Machine Learning models.

<i>Deep Learning Models</i>	<i>Author(s)</i>	<i>Accuracy</i>	<i>Machine Learning Models</i>	<i>Author(s)</i>	<i>Accuracy</i>
<i>YOLO</i> (v2,v3,v5,v8)	Kown et al. ⁴³ , Yang et al. ²⁵ , Lee et al., ³⁸ Rasic et al. ³⁹ , Berne et al. ³²	V2= 66% v3=91.3% 98.3% v5= NA v8= NA	SVM	Frydenlund A et al. ⁴⁹ , Cai et al. ²¹ , Yilmaz et al. ⁴⁵ Kumar et al. ³⁴ ,	87.9% NA 98% 98%
<i>VGG (16,19)</i>	Li et al. ⁴¹ , Liu et al. ³⁷ , Rao et al. ²⁷ ,	80.72% 85.5% , 93%	BLR	Frydenlund A et al. ⁴⁹ ,	90.0 ± 0.92%
<i>Google Inception3</i>	Bispo et al. ⁴³ , Lee et al. ⁴⁶ , Cai et al. ²¹ , Kise et al. ¹⁷ ,	90%- 92% NA NA	Naïve Bayse	Yilmaz et al. ⁴⁵ ,	98%
<i>ResNet-50</i>	Li m et al. ⁴¹ , Lee et al. ⁴⁷ ,	82.50%, 78.31%	Decision tree	Yilmaz et al. ⁴⁵ ,	89%
<i>DetectNet</i>	Kise et al. ¹⁷ , Ariji et al. ²⁶ , Watanabe et al. ³⁶	88% NA NA	Random forest	Yilmaz et al. ⁴⁵ ,	92%
<i>DenseNet</i>	Lee et al., Rao et al. ²⁷ ,	93%, 99.25%	Neural Network	Yilmaz et al. ⁴⁵ ,	92%
<i>EfficientNet</i>	Li et al. ⁴¹ ,	87.50%	MLP	Cai et al. ²¹ ,	NA
<i>Faster R-CNN</i>	Lee et al.,	NA	Extra Trees	Cai et al. ²¹ ,	NA
<i>RetinaNet</i>	Lee et al.,	NA	XGBoost	Cai et al. ²¹ ,	NA
<i>SSD</i>	Lee et al. ⁴⁷ ,	NA	Linear Regression	Committeri U et al. ⁴⁴ ,	95%
<i>EfficientDet-D3</i>	Lee et al.,	94.4%	Fine Tree	Committeri U et al. ⁴⁴ ,	94.3%
<i>Unet</i>	Setzer et al. ⁵¹ , Yu et al. ³³ ,	93%,90.6%			
<i>P-C-ReliefF</i>	Mohanty et al. ⁴⁸ ,	97.4%			

A. Outcomes

The primary outcomes of this review involved examining the existing literature on AI-based techniques and technologies utilised in the diagnosis of oral and maxillofacial cysts and assessing their accuracy and limitations.

1) Deep Learning (DL) Models

The most widely used DL model in this review was YOLO, with five published studies using various versions of the model. Yang et al.²⁵ reported that their YOLOv2 model achieved an accuracy of 66%, while Kwon et al.⁴²

demonstrated an improved accuracy of 91.3% with YOLOv3. Lee et al.⁴⁷, Berne et al.³² and Rasic et al.³⁹ used YOLOv5 and YOLOv8 models, respectively, and found promising results in terms of precision, although specific accuracy values were not provided.

The second most frequently used model was Google InceptionV3. Four studies utilised this model. Bispo et al.⁴³ reported accuracy values ranging from 90% to 92% across five iterations. Kise et al.¹⁷ achieved an accuracy of 88% with InceptionV3. However, the studies by Lee et al.⁴⁶ and Cai et al.²¹ did not report accuracy, focusing instead on area under the curve (AUC) and sensitivity metrics.

VGG16 and VGG19 models were used in three retrospective studies. Liu et al.³⁷ reported an accuracy of 80.72%, Rao et al.²⁷ achieved 93% accuracy, and Li et al.⁴¹ reported an accuracy of 85.50% with their VGG16/19 models. DetectNet was featured in three studies. Kise et al.¹⁷ reported an accuracy of 88%, while Arijji et al.²⁶ and Watanabe et al.³⁶ did not provide accuracy metrics but reported high scores in sensitivity and precision, respectively. Lee et al.⁴⁷ reported high mean average precision for models like Faster R-CNN, YOLOv5, RetinaNet and SSD but did not provide accuracy values. However, the EfficientDet-D3 model scored a notable 94.4% accuracy.

The highest accuracy was reported for DenseNet, with Lee et al.³⁸ achieving an accuracy of 99.25% and Rao et al.²⁷ reporting 93%. Unet models were used by Setzer et al.⁵¹ and Yu et al.³³, who reported accuracies of 93% and 90%, respectively. Lastly, Mohanty et al.⁴⁸ reported a high accuracy of 97.4% with their P-C-ReliefF model.

2) Machine Learning (ML) Models

Of the 26 studies reviewed, only five utilised ML models. Frydenlund et al.⁴⁹ employed an SVM and reported an accuracy of 87.9%. Additionally, the same study used bagging with logistic regression, achieving an accuracy of $90.0 \pm 0.92\%$. Yilmaz et al.⁴⁵ evaluated multiple ML models and reported accuracy scores of 98% for SVM, 98% for Naïve Bayes, 89% for decision tree, 92% for random forest and 92% for neural networks. Committeri et al.⁴⁴ reported 95% accuracy using a linear regression model and 94.3% accuracy using a fine tree model. Cai et al.²¹ utilised four ML models: SVM, extra trees, XGBoost and MLP. Although the authors did not report specific accuracy numbers, these models achieved high values in the AUC metric. Finally, Kumar et al.³⁴ reported an accuracy of 98% using an SVM model.

3) Digital Pathology

Interestingly, out of all 26 studies, only four applied AI models for digital pathology. Frydenlund et al.⁴⁹ pioneered the AI ML application for detecting odontogenic development cysts using H & E micrographic imaging with an accuracy between 87.9% and 92%. The other three studies – Rao et al.²⁷, Mohanty et al.⁴⁸ and Cai et al.²¹ – utilised DL models to detect and diagnose OKC lesions using whole-slide imaging. The accuracy results for Rao et al.²⁷ and Mohanty et al.⁴⁸ were 91% and 97.4%, respectively. Cai et al.²¹ went further and measured the diagnosis and prognosis of OKCs using AUC, with the diagnosis being 0.935 and the prognosis 0.840, indicating high performance in distinguishing between other cysts.

4) Odontogenic Cysts

The majority of the reviewed studies focused on reporting odontogenic cysts using OPG, and AI models reported high performance results with an average accuracy higher than 85%^{25,26, 32, 33, 34,35, 36, 37, 38, 41, 43,52}. The reported studies using CBCT included Committeri et al.⁴⁴, Bispo et al.⁴³, Lee et al.⁴⁶, Setzer et al.⁵¹ and Yilmaz et al.⁴⁵ and had higher performance than those using OPG.

5) Nonodontogenic Cysts

We found only two studies reporting on nasopalatine duct cysts in comparison to periapical cysts^{17,47}. The results of both achieved high accuracy in relation to dentists, with an accuracy of 88% and precision of 99.8%.

6) Validation Tests

Among the 26 retrospective studies reviewed, 17 applied a hold-out validation test, five implemented a five-fold cross-validation, two used a 10-fold cross-validation, and only one utilised a leave-one-out validation test. Only one study compared the effectiveness of these validation methods on the accuracy of AI models. Yilmaz et al.⁴⁵ reported that the highest accuracy was achieved with the 10-fold cross-validation at 100%, followed by the split-sample (or hold-out test) at 96% and, lastly, leave-one-out cross-validation at 90%.

7) Duration Performance

Ariji et al.²⁶ reported that their model, DetectNet, required 3 hours for training, 13 seconds for validation and 13 seconds for each testing session. Yang et al.²⁵ calculated the average time to evaluate 181 images and revealed that oral surgeons and general dentists took an average of 33.8 minutes, while their model, YOLOv2, provided real-time detection capabilities. Liu et al.³⁷ achieved an average processing time of 0.15 seconds per image using their models. Similarly, Lee et al.³⁸ reported that their DenseNet model required only 0.06 seconds per sample during testing.

B. Limitations of Artificial Intelligence (AI) Models

Across the reviewed studies, several limitations of using AI were identified. Computational complexity is a significant challenge, as noted by Rao et al.²⁷, with many models requiring advanced and faster hardware to function effectively. As highlighted by Cai et al.²¹, the reliance on single-centre samples, which means the collection of data from a single hospital or institute, limits the generalisability (the ability of the AI model to perform well on new and unseen data) of findings, thus emphasising the need for multicentre cohorts for more robust validation. Small sample sizes, as mentioned by Berne et al.³², often lead to concerns about overfitting and the reliability of a model's generalisability. Additionally, the interpretability of DL models, another concern raised by Berne et al.³², remains an issue, making it difficult to understand and trust their decisions in clinical settings. Variability in data quality, such as image clarity, significantly impacts model performance, as seen in Mohanty et al.⁴⁸, thus necessitating consistent data standards. Finally, diverse validation methods reveal differences in accuracy, as reported by Yilmaz et al.⁴⁵, underscoring the need for standardised validation approaches to ensure consistent and reliable results.

1) Application of Artificial Intelligence (AI) in Clinical Settings

Watanabe et al.³⁶ emphasised the need for extensive and more training datasets to reduce error effects. In addition, improved network architecture may pave the way for clinical applications. Endres et al.⁴⁰ highlighted the regulatory constraints and questions regarding whether AI predictions influence clinicians' responses and patients' trust and whether these effects will evolve over time with confidence in AI. Studies, such as Cai et al.²¹, that employed retrospective designs may not provide the strongest evidence compared to prospective studies, potentially hindering the implementation of AI in clinical practice. Bispo et al.⁴³ stated that, even though their model demonstrated high diagnostic value, their practical application remains limited. This is primarily due to the time-consuming nature of the segmentation and training processes, coupled with the high computational costs involved, which pose substantial limitations to their widespread use.

2) Risk of Bias in Included Studies

We used Cochrane's tool ROBINS-I to assess the risk of bias in each individual study (presented in Figure 3 and Figure 4), which are summarised in Table 1 in the Appendix

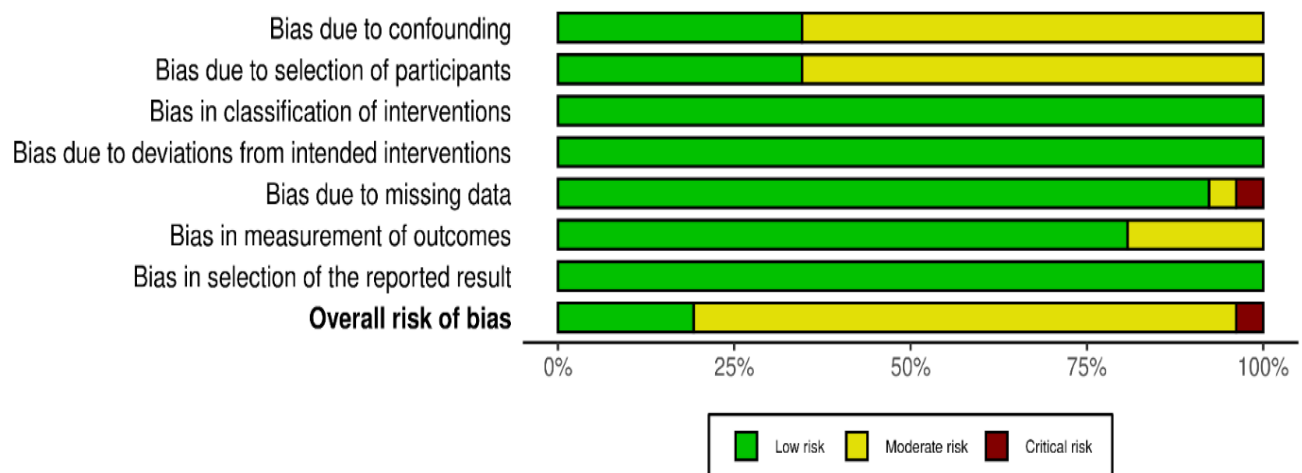


Figure 5. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

Most of the included studies had moderate bias due to confounding factors and the selection of participants. Only one study, published by Frydenlund et al.⁴⁹, had critical risk bias due to missing data. However, all studies had low risks of bias in the classification of interventions and deviations from intended interventions.

Discussion

Oral and maxillofacial cysts are epithelial-lined lesions that may cause infection, pain, or bone expansion, often affecting adjacent anatomical structures. However, most are asymptomatic and found incidentally on radiographs. Aside from periapical cysts, which are linked to dental caries, most cysts lack a well-defined etiology.

Diagnosing these lesions is complex due to overlapping clinical and radiographic features. Accurate diagnosis typically requires clinical evaluation, imaging, and histopathological confirmation. The emergence of artificial intelligence (AI) offers a promising tool that may enhance diagnostic accuracy and reduce diagnostic time.

Interest in AI has surged globally. Our review included studies from 1975 to 2024, though no AI-based diagnostic research on oral cysts was found before 2014. Notably, publications have increased recently, with six studies released in the first half of 2024 alone. This trend may reflect growing awareness of AI and increased productivity during the COVID-19 pandemic. Most studies were from East Asia, particularly China and South Korea, suggesting regional leadership in dental AI research.

All included studies used standard AI performance metrics:

- **Accuracy:** Overall correctness of model predictions.
- **Sensitivity (Recall):** Ability to correctly identify positive cases.
- **AUC:** Model's ability to distinguish between classes.
- **F1-score:** Harmonic mean of precision and recall—especially useful for imbalanced datasets.

No single AI model demonstrated flawless performance across all metrics. For example, Yang et al.²⁵ reported that YOLOv3 achieved a 0.7 precision using a relatively simple neural network architecture. More recently, Rasic et al.³⁹ introduced YOLOv8, which significantly improved performance using 401 layers and over 45 million parameters.

Most studies applied AI to orthopantomogram (OPG) images, with only five using CBCT. Although both modalities yielded good results, Lee et al.⁴⁶ found CBCT superior for detecting cysts, though OPG remains more accessible and involves lower radiation.

Studies by Kise et al.¹⁷ and Lee et al.⁴⁷ evaluated AI detection of nasopalatine duct cysts, comparing AI models with general practitioners and residents. Results were promising, but further research involving experienced oral surgeons is needed to validate AI's clinical utility across expertise levels.

A common limitation was reliance on single-centre datasets, which may reduce generalisability. Multi-centre studies, such as those by Feher et al.³⁵ and Mohanty et al.⁴⁸, provide stronger evidence and should be encouraged.

Another key challenge is dataset size. To address this, many studies used techniques like transfer learning and fine-tuning, which adapt pretrained models for specific tasks^{21,25, 47}. These methods save time and reduce overfitting, making them valuable when data are limited.

The so-called "black box" nature of AI remains a concern. Kise et al.¹⁷ noted the lack of transparency in how models reach decisions. To improve safety and explainability, regulatory bodies like the MHRA, FDA, and Health Canada have issued joint guidelines for best practices in machine learning²⁹.

We excluded two recent systematic reviews by Tobias et al.⁵³ and Shrivastava et al.⁵⁴, as they did not address histopathology or nonodontogenic cysts. Our review is more comprehensive, covering both radiographic and digital

histological studies. However, both reviews reported high heterogeneity among studies, which aligns with our observations.

Currently, performing a robust meta-analysis in this field is challenging due to wide variability in models, datasets, and evaluation metrics. As Cai et al.²¹ suggest, future prospective studies and, eventually, randomized controlled trials are needed to assess AI's clinical value in diagnosing oral and maxillofacial cysts.

Our review has several limitations. Only two databases were searched, which may have excluded relevant studies. Expanding the search scope would increase comprehensiveness. Additionally, the Risk of Bias assessment was conducted by a single reviewer, which may introduce subjectivity. Collaborative or blinded assessments could strengthen reliability.

Conclusion

This literature review examined 26 retrospective studies on the diagnosis of oral and maxillofacial cysts using artificial intelligence (AI). The findings indicate a growing awareness and potential for AI in diagnosing oral lesions. AI models demonstrated high accuracy in both machine learning and deep learning across radiographic and digital histopathological settings.

However, several limitations were identified, including limited datasets, lack of explainability, variations in data quality, and the absence of standard validation tests. Oral and maxillofacial cysts pose significant health risks if not detected early. AI models have shown promising effectiveness and speed in detecting and diagnosing both simple and complex cysts. To enhance the reliability of AI in diagnosing oral and maxillofacial cysts, future research should focus on several key areas. Larger and more diverse datasets are needed to improve generalizability, while longitudinal and multi-centre prospective studies will provide comprehensive validation across different settings. Involvement of experienced clinicians as benchmarks, standardized validation protocols, and improved explainability are crucial for trust and accuracy. Additionally, randomized clinical trials should be conducted where applicable to assess the real-world impact of AI on clinical outcomes, aligning with governmental health guidelines. A systematic review with meta-analysis on YOLO models in oral cyst diagnosis could reduce heterogeneity and elevate study quality, given the multiple studies that have utilized YOLO models in this context. Interdisciplinary collaboration will also ensure the development of clinically relevant AI tools.

Data availability

All data in this literature review were accessed through publicly available databases, including Google Scholar and PubMed. All the sources have been included within the reference list and can be accessed through these publicly accessible databases.

Conflict of interest

Authors have not conflict of interest

Author contributions statement

Author was responsible of the conception, designing the review, conducting the primary literature search, carrying out the data extraction, the analysis, and writing the manuscript.

Acknowledgements

Ethics declarations

Ethical approval was not sought for this study as it involved a literature review and analysis of existing publicly available data. The information and data used in this study were obtained from previously published academic literature, which is freely accessible and does not involve direct interaction with human subjects or animals.

References

-
1. Nayyer NV, Macluskey M, Keys W. Odontogenic Cysts – An Overview. core.ac.uk [Preprint] 2015. https://core.ac.uk/reader/30664545?utm_source=linkout (Accessed: 2 November 2023).

2. Jones AM, Craig GT, Franklin CD. Range and demographics of odontogenic cysts diagnosed in a UK population over a 30-year period. *J Oral Pathol Med* 2006;35(8):500–507. <https://doi.org/10.1111/j.1600-0714.2006.00455.x>.
3. Martin L, Speight PM. Odontogenic cysts. *Diagn Histopathol* 2015;21(9):359–369. <https://doi.org/10.1016/j.mpdhp.2015.07.005>.
4. Wang LL, Olmo H. *Odontogenic Cysts*. Treasure Island (FL): StatPearls Publishing, 2022. <https://www.ncbi.nlm.nih.gov/books/NBK574529/> (Accessed: 13 November 2023).
5. Woo SB. Odontogenic cysts. In: *Oral and Maxillofacial Pathology*. New Delhi: Elsevier; 2023. <https://www.sciencedirect.com/science/article/pii/B9780323829182000238> (Accessed: 13 November 2023).
6. Austin RP, Nelson BL. Sine Qua Non: Dentigerous Cyst. *Head Neck Pathol* 2021;15(4):1261–1264. <https://doi.org/10.1007/s12105-021-01327-3>.
7. Dave M, Patel S, Shah A, et al. Adapting treatment approaches for dentigerous cysts in paediatric and adult patients: A case series. *Oral Surg* 2020;14(3):277–284. <https://doi.org/10.1111/ors.12575>.
8. Biswas P, Sivapathasundharam B, Preethi S. The World Health Organization classification of odontogenic and maxillofacial bone tumors: An appraisal. *J Oral Maxillofac Pathol* 2019;23(2):178. https://doi.org/10.4103/jomfp.jomfp_211_19.
9. Ghazy S, El-Mofty M, El-Masry M, et al. Odontogenic Keratocyst: A Review of Histogenesis, Classification, Clinical Presentation, Genetic Aspect, Radiographic Picture, Histopathology and Treatment. *Egypt J Histol* 2021;0(0). <https://doi.org/10.21608/ejh.2021.58363.1419>.
10. Greer Jr R, Parshar P. Odontogenic Keratocyst Pathology: Definition, Epidemiology, Etiology. *eMedicine [Preprint]* 2021. <https://emedicine.medscape.com/article/1731868-overview?form=fpf> (Accessed: 19 November 2023).
11. Idar Paveli B, Vuković A, Vuković R, et al. The Incidence of Satellite Cysts in Keratocystic Odontogenic Tumors. *Hrčak* 2014. <https://hrcak.srce.hr/file/178864> (Accessed: 20 November 2023).
12. Spiker AM, Troxell T, Ramsey ML. *Gorlin Syndrome*. Treasure Island (FL): StatPearls Publishing, 2023. <https://pubmed.ncbi.nlm.nih.gov/28613671/> (Accessed: 20 November 2023).
13. Zaher B, El Bouhairi M, Ben Yahya I. Gorlin-Goltz syndrome from diagnosis to treatment: Role of the dentist. *Adv Oral Maxillofac Surg* 2023;9:100370. <https://doi.org/10.1016/j.adoms.2022.100370>.
14. McKinney R, Olmo H. *Non-Odontogenic Cysts*. Treasure Island (FL): StatPearls Publishing, 2022. <https://www.ncbi.nlm.nih.gov/books/NBK572126/> (Accessed: 26 November 2023).
15. Diaz de Ortiz LE, Mendez MD. *Epstein Pearls*. Treasure Island (FL): StatPearls Publishing, 2020. <https://www.ncbi.nlm.nih.gov/books/NBK493177/> (Accessed: 28 November 2023).
16. Apajalahti S, Holappa H, Kontio R, et al. Imaging characteristics of ameloblastomas and diagnostic value of computed tomography and magnetic resonance imaging in a series of 26 patients. *Oral Radiol* 2015. <https://www.sciencedirect.com/science/article/pii/S2212440315008974?via%3Dihub> (Accessed: 24 April 2024).
17. Kise Y, Arijji Y, Fukuda M, et al. Deep learning system for distinguishing between nasopalatine duct cysts and radicular cysts arising in the midline region of the anterior maxilla on panoramic radiographs. *Imaging Sci Dent* 2023;53(1). <https://doi.org/10.5624/isd.20230169>.
18. Perez A, Lenoir V, Lombardi T. Dentigerous Cysts with Diverse Radiological Presentation Highlighting Diagnostic Challenges. *Diagnostics* 2022;12(8):2006. <https://doi.org/10.3390/diagnostics12082006>.
19. Hunter KD, Niklander S. Pitfalls in odontogenic lesions and tumours: a practical guide. *Oral Surg* 2021. <https://www.sciencedirect.com/science/article/pii/S1756231720300141#sec4> (Accessed: 25 April 2024).
20. Krishna A, Ramesh V, Krishnan R, et al. Role of artificial intelligence in diagnostic oral pathology—A modern approach. *J Oral Maxillofac Pathol* 2020;24(1):152. https://doi.org/10.4103/jomfp.jomfp_215_19.
21. Cai X, Zhang Y, Liu Y, et al. Digital pathology-based artificial intelligence models for differential diagnosis and prognosis of sporadic odontogenic keratocysts. *Int J Oral Sci* 2024;16(1):1–10. <https://doi.org/10.1038/s41368-024-00287-y>.
22. Sheikh H, Prins C, Schrijvers E. Artificial Intelligence: Definition and Background. In: *Research for Policy*. Springer; 2023:15–41. https://doi.org/10.1007/978-3-031-21448-6_2.
23. Anyoha R. *The History of Artificial Intelligence*. Science in the News, Harvard University 2017. <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/> (Accessed: 14 January 2024).

24. Athanasopoulou K, Kourou K, Papageorgiou E, et al. Artificial Intelligence: The Milestone in Modern Biomedical Research. *BioMedInformatics* 2022;2(4):727–744. <https://doi.org/10.3390/biomedinformatics2040049>.
25. Yang H, Zhang Y, Liu Z, et al. Deep Learning for Automated Detection of Cyst and Tumors of the Jaw in Panoramic Radiographs. *J Clin Med* 2020;9(6):1839. <https://doi.org/10.3390/jcm9061839>.
26. Arijji Y, Fukuda M, Kise Y, et al. Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2019;128(4):424–430. <https://doi.org/10.1016/j.oooo.2019.05.014>
27. Rao RS, Patil S, Ganavi BS, et al. Deep Learning-Based Microscopic Diagnosis of Odontogenic Keratocysts and Non-Keratocysts in Haematoxylin and Eosin-Stained Incisional Biopsies. *Diagnostics* 2021;11(12):2184. <https://doi.org/10.3390/diagnostics11122184>.
28. Khan B, Ali S, Raza M, et al. Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. *Biomed Mater Devices* 2023;1(36785697):1–8. <https://doi.org/10.1007/s44174-023-00063-2>.
29. MHRA. Transparency for machine learning-enabled medical devices: guiding principles. GOV.UK 2024. <https://www.gov.uk/government/publications/machine-learning-medical-devices-transparency-principles/transparency-for-machine-learning-enabled-medical-devices-guiding-principles> (Accessed: 8 August 2024).
30. NHS Transformation Directorate. Understanding healthcare workers' confidence in Artificial Intelligence (AI). Health Education England | Digital Transformation 2022. <https://digital-transformation.hee.nhs.uk/building-a-digital-workforce/dart-ed/horizon-scanning/understanding-healthcare-workers-confidence-in-ai/foreword> (Accessed: 7 May 2024).
31. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919. <https://doi.org/10.1136/bmj.i4919>.
32. Ver Berne J, de Jongh M, van der Stelt P, et al. A deep learning approach for radiological detection and classification of radicular cysts and periapical granulomas. *J Dent* 2023;135:104581. <https://doi.org/10.1016/j.jdent.2023.104581>.
33. Yu D, Zhang Y, Li X, et al. Deep learning based diagnosis for cysts and tumors of jaw with massive healthy samples. *Sci Rep* 2022;12(1). <https://doi.org/10.1038/s41598-022-05913-5>.
34. Nishath Sayed Abdul, Kumar A, Sharma R, et al. Applications of artificial intelligence in the field of oral and maxillofacial pathology: a systematic review and meta-analysis. *BMC Oral Health* 2024;24(1). <https://doi.org/10.1186/s12903-023-03533-7>.
35. Feher B, Dombi A, Varga E, et al. Emulating Clinical Diagnostic Reasoning for Jaw Cysts with Machine Learning. *Diagnostics* 2022;12(8):1968. <https://doi.org/10.3390/diagnostics12081968>.
36. Watanabe H, Arijji Y, Kise Y, et al. Deep learning object detection of maxillary cyst-like lesions on panoramic radiographs: preliminary study. *Oral Radiol* 2021;37(3):487–493. <https://doi.org/10.1007/s11282-020-00485-4>.
37. Liu Z, Zhang Y, Li X, et al. Differential diagnosis of ameloblastoma and odontogenic keratocyst by machine learning of panoramic radiographs. *Int J Comput Assist Radiol Surg* 2021;16(3):415–422. <https://doi.org/10.1007/s11548-021-02309-0>.
38. Lee A, Kim JE, Park YH, et al. Deep learning neural networks to differentiate Stafne's bone cavity from pathological radiolucent lesions of the mandible in heterogeneous panoramic radiography. *PLoS One* 2021;16(7):e0254997. <https://doi.org/10.1371/journal.pone.0254997>.
39. Rašić M, Vuković A, Pavelić B, et al. Detection and Segmentation of Radiolucent Lesions in the Lower Jaw on Panoramic Radiographs Using Deep Neural Networks. *Medicina* 2023;59(12):2138. <https://doi.org/10.3390/medicina59122138>.
40. Endres MG, Hillen F, Grabe N, et al. Development of a Deep Learning Algorithm for Periapical Disease Detection in Dental Radiographs. *Diagnostics* 2020;10(6):430. <https://doi.org/10.3390/diagnostics10060430>.
41. Li M, Zhang Y, Liu H, et al. [Application of Deep Learning in Differential Diagnosis of Ameloblastoma and Odontogenic Keratocyst Based on Panoramic Radiographs]. *PubMed* 2023;45(2):273–279. <https://doi.org/10.3881/j.issn.1000-503x.15139>.

-
42. Kwon O, Lee JH, Kim Y, et al. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep convolution neural network. *Dentomaxillofac Radiol* 2020;49(8):20200185. <https://doi.org/10.1259/dmfr.20200185>.
43. Bispo MS, de Souza RM, de Oliveira AE, et al. Computer tomographic differential diagnosis of ameloblastoma and odontogenic keratocyst: classification using a convolutional neural network. *Dentomaxillofac Radiol* 2021;50(7):20210002. <https://doi.org/10.1259/dmfr.20210002>.
44. Committeri U, Rossi M, De Angelis F, et al. New perspectives in the differential diagnosis of jaw lesions: machine learning and inflammatory biomarkers. *J Stomatol Oral Maxillofac Surg* 2024;101912. <https://doi.org/10.1016/j.jormas.2024.101912>.
45. Yilmaz E, Kayikcioglu T, Kayipmaz S. Computer-aided diagnosis of periapical cyst and keratocystic odontogenic tumor on cone beam computed tomography. *Comput Methods Programs Biomed* 2017;146:91–100. <https://doi.org/10.1016/j.cmpb.2017.05.012>.
46. Lee J, Kim D, Jeong S. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. *Oral Dis* 2019;26(1):152–158. <https://doi.org/10.1111/odi.13223>.
47. Lee HS, Kim JE, Park YH, et al. Automatic detection and classification of nasopalatine duct cyst and periapical cyst on panoramic radiographs using deep convolutional neural networks. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2024;138(1):184–195. <https://doi.org/10.1016/j.oooo.2023.09.012>.
48. Mohanty S, Rao S, Patel A, et al. Building Automation Pipeline for Diagnostic Classification of Sporadic Odontogenic Keratocysts and Non-Keratocysts Using Whole-Slide Images. *Diagnostics* 2023;13(21):3384. <https://doi.org/10.3390/diagnostics13213384>.
49. Frydenlund A, Eramian M, Daley T. Automated classification of four types of developmental odontogenic cysts. *Comput Med Imaging Graph* 2014;38(3):151–162. <https://doi.org/10.1016/j.compmedimag.2013.12.002>.
50. Tajima S, Arijji Y, Kise Y, 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. *J Oral Maxillofac Surg Med Pathol* 2022;34(5):553–560. <https://doi.org/10.1016/j.ajoms.2022.02.004>.
51. Setzer FC, Shi Z, Zhang Y, et al. Artificial Intelligence for the Computer-aided Detection of Periapical Lesions in Cone-beam Computed Tomographic Images. *J Endod* 2020;46(7):987–993. <https://doi.org/10.1016/j.joen.2020.03.025>.
52. 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. <https://doi.org/10.1016/j.joen.2019.03.016>.
53. Tobias RS, Patel S, Mohanty S, et al. Diagnostic capability of artificial intelligence tools for detecting and classifying odontogenic cysts and tumors: a systematic review and meta-analysis. *Sci Direct* 2024. <https://www.sciencedirect.com/science/article/pii/S2212440324001214> (Accessed: 9 June 2024).
54. Priyanshu Kumar Shrivastava, Sharma R, Sayed Abdul N, et al. Accuracy of machine learning in the diagnosis of odontogenic cysts and tumors: a systematic review and meta-analysis. *Oral Radiol* 2024;40(3):342–356. <https://doi.org/10.1007/s11282-024-00745-7>.