Dentomaxillofacial Radiology (2020) 49, 20190204

birpublications.org/dmfr

SYSTEMATIC REVIEW

Accuracy of computer-assisted image analysis in the diagnosis of maxillofacial radiolucent lesions: A systematic review and meta-analysis

[1]Virginia K S Silva, [2]Walbert A Vieira, [3]Ítalo M Bernardino, [4]Bruno A N Travençolo, [5]Marcos A V Bittencourt, [6]Cauane Blumenberg, [7]Luiz R Paranhos and [8]Hebel C Galvão

[1]Department of Dentistry, Postgraduate Program in Health Sciences, Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil; [2]Postgraduate Program in Dentistry, Endodontics Division, Piracicaba Dental School, State University of Campinas, Piracicaba, São Paulo, Brazil; [3]Department of Dentistry, Postgraduate Program in Dentistry, State University of Paraíba, Campina Grande, Paraíba, Brazil; [4]Center for Exact Sciences and Technology, School of Computing, Federal University of Uberlândia, Uberlândia, Minas Gerais, Brazil; [5]Department of Pediatric and Community Dentistry, School of Dentistry, Federal University of Bahia, Salvador, Bahia, Brazil; [6]Department of Social Medicine, Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Rio Grande do Sul, Brazil; [7]Department of Preventive and Community Dentistry, School of Dentistry, Federal University of Uberlândia, Uberlândia, Minas Gerais, Brazil; [8]Department of Dentistry, Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil

Objectives: This study aimed to search for scientific evidence concerning the accuracy of computer-assisted analysis for diagnosing maxillofacial radiolucent lesions.
Methods: A systematic review was conducted according to the statements of Preferred Reporting Items for Systematic Reviews and Meta-analyses Protocols and considering 10 databases, including the gray literature. Protocol was registered at the International Prospective Register of Systematic Reviews (CRD42018089945). The population, intervention, comparison and outcome strategy was used to define the eligibility criteria and only diagnostic test studies were included. Their risk of bias was assessed by the Joanna Briggs Institute Critical Appraisal tool. Random-effects model meta-analysis was performed and heterogeneity among the included studies was estimated using the I^2 statistic. The grade of recommendation, assessment, development, and evaluation (GRADE) tool assessed the quality of evidence and strength of recommendation across included studies.
Results: Out of 715 identified citations, four papers, published between 2009 and 2017, fulfilled the criteria and were included in this systematic review. A total of 191 lesions, classified as periapical granuloma and cyst, dentigerous cyst or keratocystic odontogenic tumor, were analyzed. All selected articles scored low risk of bias. The pooled accuracy estimation, regardless of the classification method used, was 88.75% (95% CI = 85.19-92.30). Heterogeneity test reached moderate values (I^2 = 57.89%). According to the GRADE tool, the analyzed outcome was classified as having low level of certainty.
Conclusions: The overall evaluation showed all studies presented high accuracy rates of computer-aided diagnosis systems in classifying radiolucent maxillofacial lesions compared to histopathological biopsy. However, due to the moderate heterogeneity found among the studies included in this meta-analysis, a pragmatic recommendation about the use of computer-assisted analysis is not possible.
Dentomaxillofacial Radiology (2020) **49**, 20190204. doi: 10.1259/dmfr.20190204

Cite this article as: Silva VKS, Vieira WA, Bernardino ÍM, Travençolo BAN, Bittencourt MAV, Blumenberg C, et al. Accuracy of computer-assisted image analysis in the diagnosis of maxillofacial radiolucent lesions: A systematic review and meta-analysis. *Dentomaxillofac Radiol*

Correspondence to: Dr Marcos A V Bittencourt, E-mail: alan_orto@yahoo.com.br

Received 26 May 2019; revised 19 August 2019; accepted 05 November 2019

2020; **49**: 20190204.

Keywords: Computer-assisted diagnosis; Computer-assisted image analysis; Computer-assisted image processing; periapical cyst; Maxillofacial pathology.

Introduction

Radiographies, together with clinical and histological exams, play an important role in the diagnosis of bone lesions in craniofacial area.[1] Traditionally, two fundamental strategies exist for the diagnosis of periapical lesions with the use of radiographs. The first is to base the diagnosis on local features in the area of interest, and the second is to base it on the global structures in the radiograph.[2] Technology has provided major changes in this segment, and the type and use of digital radiographic equipment has already been subject of several studies in multiple countries.[3–5] These studies have shown an increased availability of dental radiographic units throughout the years, with an exponential increase in digitalization.[6] This evolution is noticeable and has provided clinicians with greater diagnostic possibilities since the use of digital periapical radiographs[7,8] till the use of CBCT scans[9–11] and MRI.[12,13]

Despite this digital revolution on imaging technologies, an accurate diagnosis is still dependent on the experience and judgment of the specialist,[14] and this is time-consuming and prone to human error.[15] With the aim of minimizing the subjectivity of personal evaluation and substantiating the radiologists, semi-automatic solutions using computational algorithms for diagnosis and measurement of maxillofacial lesions have been developed.[15–17] This computer-aided diagnosis (CAD) has been developed to perform an automated quantitative analysis of the textural features, voxel intensity, location, shape, area and entropy of lesions, using mathematical formulas and algorithms.[18]

As CAD is a non-invasive tool for diagnosis and with instant results,[19] it has been successful in many high-impact clinical areas, providing medical professionals with a valuable second opinion, being a great complementary tool.[20–22] However, the field of dentistry has benefited little from the advancements of medical image analysis, despite how common the dental practice is to our daily life. According to Yilmaz et al,[23] research conducted on dental images in the field of computer vision is extremely challenging. As radicular cysts are the main reason of chronic swelling in maxillofacial region,[24] some studies have been performed for their detection in radiographs or CBCT scans. Mol et al[2] have determined that it is feasible to use texture analysis to identify the presence of trabecular bone pattern in radiographs and to detect a periapical bone lesion based on a local absence of this pattern. Later, Flores et al[15] and Okada et al[25] combined graph-based random walks segmentation with machine learning-based boosted classifiers in order to diagnose these lesions. In Banumathi et al,[16] an algorithm was developed to identify dental cysts using morphological descriptors of the shape, margin and area at an early stage. A combination of support vector machine based on texture features and a novel approach involving logistic regression was proposed by Nurtanio et al[1,26] and Yilmaz et al,[23] aiming to identify cyst and tumor lesions. Finally, automatic segmentation of maxillofacial structure based on asymmetry analysis of pixels was proposed by Abdolali et al[27] also for diagnosis of cysts and tumors.

All the above-mentioned authors have tried to create computer learning-based methodologies to simplify the task of establishing a non-invasive diagnosis for maxillofacial lesions. However, as an accurate and fast classification of these lesions is clinically very important, improving the quality of patient care, the present systematic review and meta-analysis aim to investigate and exposure if there is scientific evidence that supports the validity of computer image processing in accurately performing the differential diagnosis of radiolucent maxillofacial lesions, comparing to the histological biopsy diagnosis.

Methods and materials

Protocol and registration

This systematic review was performed according to the statements of the Preferred Reporting Items for Systematic Reviews and Meta-analyses Protocols (PRISMA-P),[28] with guidance from the Cochrane Handbook for Systematic Reviews.[29] Systematic review protocol was registered at the International Prospective Register of Systematic Reviews (PROSPERO) database, under the number CRD42018089945 (http://www.crd.york.ac.uk/PROSPERO).

Study design and eligibility criteria

Using the population, intervention, comparison and outcome (PICO) strategy to define the eligibility criteria, the present research aimed to answer the following question: "Is the CAD of radiolucent maxillofacial lesions as accurate as histological biopsy?"

Only studies that evaluated the accuracy of CAD of radiolucent maxillofacial lesions were included. No language or publication year were imposed. The following exclusion criteria were applied: (1) studies in which the subject of interest was not addressed, (2) abstracts or indexes, (3) letters to editors, (4) literature reviews, (5) personal or short communications, (6) book chapters, (7) patents, (8) studies in which the histological biopsy was not used to confirm the diagnosis, and (9) studies with high risk of bias or poor methodological quality.

Search and information sources

In order to identify relevant studies, a systematic search was conducted in the following electronic databases: PubMed (including MedLine), Scopus, Embase, SciELO, Web of Science, LILACS and IEEE Xplore. A partial grey literature search was performed through OpenGrey, OpenThesis and Open Access Thesis and Dissertations (OATD). It was also conducted a hand search of cross-references from original articles to identify additional studies that could not be located in the electronic databases. These procedures were performed to avoid potential selection and publication biases.

Descriptors were selected using the Descriptors in Health Science (DeCS), the Medical Subject Headings (MeSH) and the Embase Subject Headings (Emtree). Boolean operators (AND and OR) were used to combine the descriptors and potentiate the search strategy by means of different combinations. This search was performed in January 2019. The full electronic search strategy is illustrated in the Table 1. All references obtained from PubMed, Scopus, Embase, SciELO, Web of Science, LILACS and IEEE Xplore were exported to EndNote Web™ (Thomson Reuters™, Toronto, Canada) software, in which duplicate records were removed. Additionally, references obtained from OpenGrey, OpenThesis and OATD were exported to Microsoft Word™ 2010 (Microsoft™ Ltd., Washington, USA) software, in order to manually remove possible duplicate records.

Study selection

The data collection was independently performed by two reviewers (VKSS and WAV), in three different phases. First, reviewers discussed the eligibility criteria applying them in 20% of the references, aiming to assess potential errors in the method. The inter-rater concordance was evaluated by κ statistic, obtaining a strong agreement ($K \geq 0.81$) and confirming the reproducibility and reliability of the evaluation. Then, titles were carefully read to exclude articles out of the scope of this research. Reviewers were not blinded for authorship information or name of the journals. Studies in which the subject of interest could not be addressed were excluded.

In Phase 2, abstracts of the remaining studies were independently analyzed by the two reviewers. At this stage, abstracts in which the subject of interest could not be addressed, literature reviews, case reports, congress abstract and studies with no histological biopsy were excluded. Those whose titles matched the eligibility criteria but did not have abstracts available were obtained and had their full-text posteriorly analyzed in Phase 3.

In Phase 3, remaining articles had their full-text evaluated and their reference lists carefully read in order to identify studies that could not be located. Then, articles were assessed to verify whether they fulfilled the other eligibility criteria. Studies in which diagnosis of radiolucent lesions was not performed with the aid of computer image processing, or that did not compare computer images to histological biopsy were excluded. When mutual agreement between the two reviewers was not reached, a third reviewer (LRP) was involved to make a final decision. Rejected studies and reasons for their exclusion were recorded.

Data collection process and data items

After the screening was performed, texts of selected articles were reviewed and data were extracted in a systematic way, considering the authorship, year of publication and country of origin of the article; the study population (size and type of lesion); the image processing features (type of exam, segmentation method, extracted features); and the way the results were achieved (classification method, validation method, rate of accuracy).

In order to ensure the consistency among reviewers, a calibration exercise was performed with both reviewers (VKSS and WAV), in which information were extracted jointly from an eligible study. After, one author (VKSS) collected all the above information and a second author (WAV) cross-checked them to confirm the quality of extracted data. Any disagreement between the reviewers was solved by discussion with a third author (MAVB). When additional assistance was necessary to make a final decision, a fourth author (LRP) was consulted.

Risk of bias in individual studies

The risk of bias of the selected articles was investigated using the Joanna Briggs Institute (JBI) Critical Appraisal tool[30] for use in JBI systematic review studies involving diagnosis accuracy.[31,32] Each domain related to the potential risk of bias was independently evaluated by two reviewers (VKSS and WAV), according to the PRISMA-P recommendations.[28] The following questions were used for this evaluation: (1) was the study based on a consecutive or random sample?; (2) was a case control design avoided?; (3) did the study avoid inappropriate exclusions?; (4) were the index test results interpreted without knowledge of the results of the reference standard?; (5) if a threshold was used, was it pre-specified?; (6) is the reference standard likely to correctly classify the target condition?; (7) were the reference standard results interpreted without knowledge of the results of the index test?; (8) was there an appropriate interval between index test and reference standard?; (9) did all patients receive the same reference standard?; (10) were all patients included in the analysis? Then, according to the tool,[30] the risk of bias should be rated as high, when the study reached up to 49% score "yes", moderate, when it reached 50 to 69% score "yes", and low, when it reached more than 70% score "yes".

Table 1 Electronic databases and applied search strategy

Database	***Search Strategy (January 2019)***
PubMed https://www.ncbi.nlm.nih.gov/pubmed/	("Odontogenic Cysts"[MeSH Term] OR "Odontogenic Cysts"[All Fields] OR "Odontogenic Cyst"[All Fields] OR "Periapical Cyst"[All Fields] OR "Dental Apical Lesion"[All Fields] OR "Radicular Cyst"[MeSH Term] OR "Radicular Cyst"[All Fields] OR "Follicular Cyst"[All Fields] OR "Apical Periodontal Cyst"[All Fields] OR "Periapical Lesion"[All Fields] OR "Maxillofacial Cyst"[All Fields]) AND ("Computer-Aided Diagnosis"[MeSH Term] OR "Computer-Aided"[All Fields] OR "Cone-Beam Computed Tomography"[MeSH Term] OR "Cone-Beam Computed Tomography"[All Fields] OR "Image Processing Computer-Assisted"[MeSH Term] OR "Image Processing Computer-Assisted"[All Fields] OR "CBCT"[All Fields] OR "Computer-Assisted Surgery"[All Fields] OR "Dental Panoramic Images"[All Fields] OR "Digital Radiography"[All Fields])
Scopus http://www.scopus.com/	("Odontogenic Cysts" OR "Apical Lesion" OR "Radicular Cyst" OR "Maxillofacial Cyst") AND ("Computer-Aided Diagnosis" OR "Cone-Beam Computed Tomography" OR "Image Processing Computer-Assisted" OR "CBCT" OR "Dental Panoramic Images")
Embase http://www.embase.com	('odontogenic cysts'/exp OR 'odontogenic cysts' OR 'periapical cyst' OR 'apical lesion' OR 'radicular cyst'/exp OR 'radicular cyst' OR 'maxillofacial cyst') AND ('computer-aided diagnosis' OR 'cone-beam computed tomography'/exp OR 'cone-beam computed tomography' OR 'image processing, computer-assisted'/exp OR 'image processing, computer-assisted' OR 'cbct' OR 'dental panoramic images' OR 'digital radiography'/exp OR 'digital radiography')
SciELO www.scielo.org/	Odontogenic Cysts AND CBCT Odontogenic Cysts AND Cone-Beam Computed Tomography Odontogenic Cysts AND Computer-Aided diagnosis Odontogenic Cysts AND Digital Radiography Apical lesion AND Digital Radiography Periapical Cyst AND Digital Radiography Periapical Cyst AND Cone-Beam Computed Tomography Maxillofacial cyst AND Digital Radiography Maxillofacial cyst AND Computer-Aided diagnosis Maxillofacial cyst AND Cone-Beam Computed Tomography Tumor Odontogênico AND Tomografia [Portuguese] Lesão Periapical AND Tomografia [Portuguese]
Web of Science http://apps.webofknowledge.com/	("Odontogenic Cysts" OR "Periapical Cyst" OR "Apical Lesion" OR "Radicular Cyst" OR "Maxillofacial Cyst") AND ("Computer-Aided Diagnosis" OR "Cone-Beam Computed Tomography" OR "Image Processing Computer-Assisted" OR "CBCT" OR "Dental Panoramic Images" OR "Digital Radiography")
LILACS lilacs.bvsalud.org/	"Odontogenic Cysts" AND "CBCT" "Odontogenic Cysts" AND "Cone-Beam Computed Tomography" "Apical Lesion" AND "Digital Radiography" "Periapical Cyst" AND "Digital Radiography" "Maxillofacial cyst" AND "Digital Radiography" "Maxillofacial cyst" AND "Computer-Aided Diagnosis" "Tumor Odontogênico" AND "Tomografia" [Portuguese] "Lesão Periapical" AND "Tomografia" [Portuguese]
IEEE Xplore https://ieeexplore.ieee.org/	("Odontogenic Cysts" OR "Periapical Cyst" OR "Apical Lesion" OR "Radicular Cyst" OR "Maxillofacial Cyst") AND ("Computer-Aided Diagnosis" OR "Cone-Beam Computed Tomography" OR "Image Processing Computer-Assisted" OR "CBCT" OR "Dental Panoramic Images" OR "Digital Radiography")
OpenGrey http://www.opengrey.eu/	("Odontogenic Cysts" OR "Apical Lesion" OR "Radicular Cyst" OR "Maxillofacial Cyst") AND ("Computer-Aided Diagnosis" OR "Cone-Beam Computed Tomography" OR "Image Processing Computer-Assisted" OR "CBCT" OR "Dental Panoramic Images" OR "Digital Radiography")
OpenThesis http://www.openthesis.org/	("Odontogenic Cysts" OR "Apical Lesion")
Open Access Thesis and Dissertations https://oatd.org/	("Odontogenic Cysts" OR "Apical Lesion")

Data analysis
A meta-analysis using random-effects model was performed to estimate pooled accuracy rates of diagnosis based on computer image processing compared to the histological biopsy. In addition to a pooled measure, accuracy was also assessed according to the different methods of classifying lesions used in the studies. The random-effects model was used aiming to minimize the influence of the heterogeneity among the included articles.[33] Heterogeneity among the included studies was estimated using the I^2 statistic.[34] Analyses were performed using Stata 15.1 (StataCorp LLC, College Station, TX, USA) software.

Quality of the evidence
GRADEpro GDT software (http://gdt.guidelinedevelopment.org) was used to summarize the results. The quality of the evidence and the strength of recommendations were evaluated using the Grade of Recommendation, Assessment, Development, and Evaluation (GRADE) tool, based on the study design, methodological limitations, inconsistency, indirect evidence, imprecision and other considerations, being rated as high, moderate, low and very low.[35]

Results

Study selection
The systematic search performed within 10 electronic databases, including the grey literature, resulted in 715 references, of which 266 were duplicates. After removing duplicates, 449 references had their titles carefully read. A total of 392 references were excluded. Then, in Phase 2, the remaining references had their abstracts analyzed and 53 were excluded. References which did not have abstracts available had their full-texts evaluated in Phase 3. In this phase, only four articles[17,23,25,36] were selected for full-text reading and assessment of the eligibility criteria. Their reference lists were also carefully read in order to identify studies that could not be located in the search. After reading the reference lists and performing a manual search of cross-references of these four selected articles, three new titles were selected.[15,16,26] After reading the full-texts of the seven articles, the study by Wiener et al[36] was excluded because did not use computer image processing, and the studies by Banumathi et al[16] and Nurtanio et al[26] were excluded because they did not perform histological biopsy. Finally, both in qualitative and quantitative analysis, four articles were included.[15,17,23,25] A flowchart depicting the selection process based in the PRISMA diagram[37] is provided in Figure 1.

Characteristics of the selected studies
Selected articles were published between 2009 and 2017, all of them written in English.[15,17,23,25] Studies were conducted by research groups from four different countries, namely United States,[15,25] Iran and Japan,[17] and Turkey.[23] Table 2 provides a summary of their characteristics.

Two out of the four eligible studies mentioned about the approval by the ethical committee.[15,25] Altogether, images from 191 patients were analyzed and, based on histopathological diagnosis, lesions were classified as periapical cyst or periapical granuloma in two studies,[15,25] radicular cyst, dentigerous cyst or keratocystic odontogenic tumor in one,[17] and as periapical cyst or keratocystic odontogenic tumor in the other study.[23]

CAD was performed using CBCT in all the studies,[15,17,23,25] although other methods of lesion segmentation have been used. Manual segmentation was performed in one research,[23] semiautomatic segmentation with minimal user interaction combined to the graph-based random walks algorithm was used in two,[15,25] and automatic segmentation based on symmetry analysis and active contours to improve the accuracy of segmentation was performed in the remaining study.[17] In one article,[25] the random walks segmentation was probabilistic extended by using likelihood ratio test formalism. As textural information is often utilized in the analysis of medical images, texture was the main feature extracted of lesions in all studies.[15,17,23,25] Two of them[15,25] considered a set of eight features computed from the lesion's intensity distribution, namely maximum, minimum, mean, median, standard deviation, skewness, kurtosis and entropy. In one study,[23] five of those features (median, standard deviation, skewness, kurtosis, and entropy) were used. In this study,[23] the gray-level co-occurrence matrix (GLCM) proposed by Haralick[38] was also used to obtain the textural data considering 12 different features: energy, entropy, correlation, contrast, variance, sum mean, inertia, cluster shade, cluster tendency, homogeneity, maximum probability and inverse variance. These features were obtained from each GLCM. As a result, 12 features were extracted from the order statistics and 624 from the GLCM, and then the feature vector consisted of 636 different values, including the order statistics and the textural feature information. Finally, in one study, texture and shape features were extracted using contourlet and orthogonalized spherical harmonics (SPHARM) coefficients.[17]

The eligible studies used different classification methods for the lesions based on the extracted features. The linear discriminant analysis (LDA) and the adaptive boosting (AdaBoost) were used in two articles,[15,25] while the support vector machine (SVM) was utilized in the other two.[17,23] Other nine classification methods were considered by the studies, namely simple threshold,[15] two types of LDA-AdaBoost combinations,[25] sparse discriminant analysis (SDA)[17] and k-nearest neighbors, naive Bayes, decision trees, random forest and neural network.[23]

Although more than one validation method have been used in two eligible studies,[23,25] the leave-one-out cross-validation (LOOCV) was used to validate the

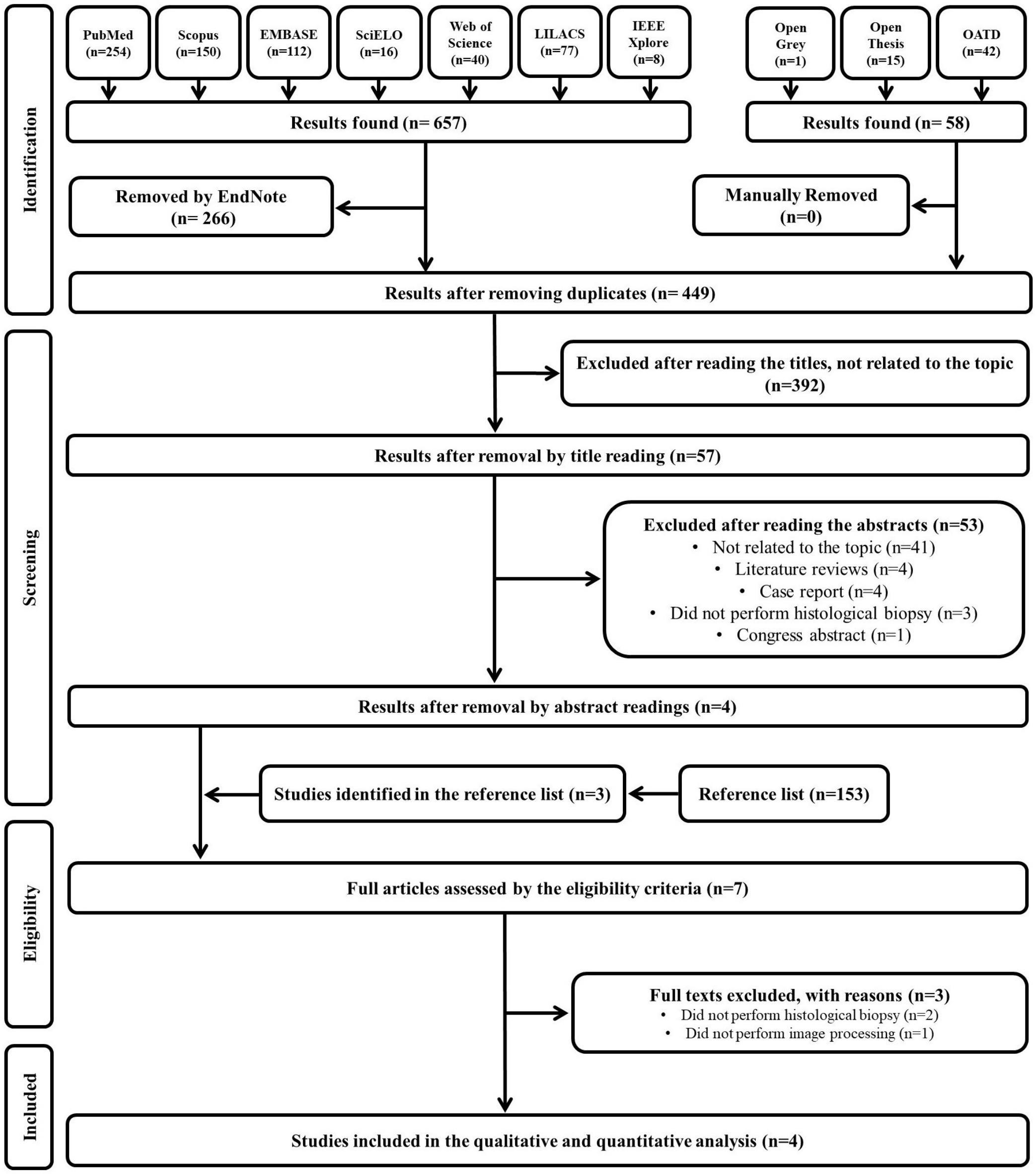

Figure 1 Flowchart showing the results of the search process.

classification experiments in three studies.[15,23,25] In the only eligible article that have not used LOOCV,[17] a three-fold cross-validation method was used for verification of the classifier performances. In addition to LOOCV, the seven-fold cross-validation was used in one research,[25] while the ten-fold cross-validation and the split sample validation were use in other one.[23]

All studies analyzed their system performance through accuracy measurements. Two studies visually assessed the diagnostic ability of the classifier system using the receiver operating characteristic (ROC) curve, created by plotting the sensitivity against the specificity values.[15,25] In addition to the sensitivity and specificity values, the positive and negative predictive values were statistical measures also used to describe the performance of different classifiers in the other two articles.[17,23] In one study,[23] besides accuracy, another metric was considered to analyze the classifier performance, the F1-score, obtaining similar results compared to the accuracy evaluation.

Risk of bias within the studies

The methodological risk of bias evaluation using the Joanna Briggs Institute (JBI) Critical Appraisal tool[30] for use in JBI systematic review studies involving diagnosis accuracy[31,32] is shown in Table 3, where it is possible to depict the answers to the 10 questions. None of the included articles fulfilled all the criteria from the checklist. All studies scored low risk of bias.[15,17,23,25] Question

Table 2 Characteristics of studies included in the systematic review and meta-analysis

Authorship Year of publication Country of origin	*N*	*Lesion Type*	*Dataset*	*Lesion Segmentation*	*User Interaction*	*Extracted Features*	*Classification Method*	*Validation Method*	*Performance Measures*
Flores et al.[15] 2009 USA	17	Periapical cyst Periapical granuloma	3D images (CBCT)	Graph-based random walks algorithm	Semiautomatic with very few interactions	Intensity statistics (maximum, minimum, mean, median, standard deviation, skewness, kurtosis, entropy)	Simple threshold LDA AdaBoost	LOOCV	Accuracy ROC curve (Sensitivity, 1-Specificity)
Okada et al.[25] 2015 USA	28	Periapical cyst (14) Periapical granuloma (14)	3D images (CBCT)	LRT-extended graph-based random walks algorithm	Semiautomatic with very few interactions	Intensity statistics (maximum, minimum, mean, median, standard deviation, skewness, kurtosis, entropy)	LDA AdaBoost LDA-AdaBoost combination	Seven-fold cross-validation LOOCV	Accuracy ROC curve (Sensitivity, 1-Specificity)
Abdolali et al.[17] 2017 Iran and Japan	96	Radicular cyst (38) Dentigerous cyst (36) KCOT (22)	3D images (CBCT)	Symmetry analysis combined with active contours	Automatic	Contourlet transform SPHARM Orthogonalized SPHARM	SVM SDA	Three-fold cross-validation	Accuracy Sensitivity Specificity Positive predictive value Negative predictive value
Yilmaz et al.[23] 2017 Turkey	50	Periapical cyst (25) KCOT (25)	3D images (CBCT)	Manual segmentation	Manual	Intensity statistics (median, standard deviation, skewness, kurtosis, entropy) GLCM (624 texture features)	K-nearest neighbors Naive Bayes Decision trees Random forest Neural network SVM	Ten-fold cross-validation Split sample validation LOOCV	Accuracy Sensitivity Specificity Positive predictive value Negative predictive value F1-score

N, Sample size; CBCT, Cone beam computed tomography; LDA, Linear discriminant analysis; AdaBoost, Adaptive boosting; LOOCV, Leave-one-out cross-validation; ROC, Receiver operating characteristic; LRT, Likelihood radio test; KCOT, Keratocystic odontogenic tumor; SPHARM, Spherical harmonics; SVM, Support vector machine; SDA, Sparse discriminant analysis; GLCM, Gray-level co-occurrence matrix.

four was answered as "no" because histopathological diagnosis was already known before image processing, and question eight was considered "unclear" because no study defined the interval between histopathological exams and CAD.

Table 3 Risk of bias assessed by the Joanna Briggs Institute (JBI) Critical Appraisal tool[30] for use in JBI systematic review studies involving diagnosis accuracy.[31,32]

Authors	*Q1*	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*	*Q7*	*Q8*	*Q9*	*Q10*	*% Yes*	*Risk of Bias*
Flores et al.[15]	√	√	√	--	√	√	√	*U*	√	√	80%	Low
Okada et al.[25]	√	√	√	--	√	√	√	*U*	√	√	80%	Low
Abdolali et al.[17]	√	√	√	--	√	√	√	*U*	√	√	80%	Low
Yilmaz et al.[23]	√	√	√	--	√	√	√	*U*	√	√	80%	Low

Q1) Was the study based on a consecutive or random sample?; Q2) Was a case control design avoided?; Q3) Did the study avoid inappropriate exclusions?; Q4) Were the index test results interpreted without knowledge of the results of the reference standard?; Q5) If a threshold was used, was it pre-specified?; Q6) Is the reference standard likely to correctly classify the target condition?; Q7) Were the reference standard results interpreted without knowledge of the results of the index test?; Q8) Was there an appropriate interval between index test and reference standard?; Q9) Did all patients receive the same reference standard?; Q10) Were all patients included in the analysis?; √) Yes; --) No; U) Unclear.

Table 4 Description of the main outcomes of selected studies

Authorship *Year of publication* *Country of origin*	*Outcome*
Flores et al.[15] 2009 USA	The individual classifier with highest accuracy was AdaBoost, using a combination of skewness, kurtosis and entropy as the feature vector, resulting in a performance of 88.2% success rate using the LOOCV statistics.
Okada et al.[25] 2015 USA	The new proposed classifier LDA-AdaBoost combination achieved the best result, yielding a performance of the proposed method of 85.7% of accuracy using the seven-fold cross-validation method and of 78.9% accuracy when the LOOCV method was applied.
Abdolali et al.[17] 2017 Iran and Japan	The proposed framework based on orthogonalized SPHARM and contourlet features to obtain shape and texture information, using SDA and SVM classifiers, achieved diagnosis accuracies of 96.48 and 94.29%, respectively, applying the three-fold cross-validation method.
Yilmaz et al.[23] 2017 Turkey	The individual classifier with highest accuracy was SVM, and its performance was increased by reducing the size of the feature vector using the forward feature selection algorithm. The best results were achieved as 100% accuracy using the 10-fold cross-validation method, as 96% accuracy when a split sample validation was used, and 94% accuracy when LOOCV method was applied.

AdaBoost, Adaptive boosting; LDA, Linear discriminant analysis; SPHARM, Spherical harmonics; SDA, Sparse discriminant analysis; SVM, Support vector machine; LOOCV, Leave-one-out cross-validation.

Results of individual studies and meta-analysis

The main outcomes of the selected articles are described in Table 4. The overall evaluation showed that all studies presented high accuracy rates of computer image processing methods in diagnosing radiolucent maxillofacial lesions. Although two studies[23,25] have used more than one validation method, the LOOCV statistics was the one preferred for obtaining the accuracy measures, since it was applied in three of the four eligible studies.[15,23,25] As can be seen in Figure 2, considering the 12 classification methods used by different authors, the lowest accuracy rate was 69.12% (95% CI = 45.45–85.74), when the simple threshold method was applied,[15] and the highest rate was 96.48 (95% CI = 91.62–98.56) using the individual classifier SDA.[17] Heterogeneity between studies was considered moderate (I^2 = 57.89%). The pooled accuracy estimation, regardless of the classification method used, was 88.75% (95% CI = 85.19–92.30).

Certainty of evidence

The quality of evidence and the strength of recommendation of the main outcomes evaluated by the GRADE tool[35] was rated as low, which means that confidence in the effect estimate is limited and that the true effect may be substantially different from the estimate of the effect (Table 5).

Discussion

This systematic review and meta-analysis aimed to evaluate the accuracy of computer image processing for performing the diagnosis of radiolucent maxillofacial lesions in comparison with the histological biopsy diagnosis. In recent years, many studies have been performed in the areas of medical imaging and signal processing,[23] including lung nodule detection[39] and breast tumor diagnosis.[40] However, each study considers a specific method depending on the characteristics of the disease of interest. In head and neck areas, only few published articles exist that address computer-aided detection and classification of lesions related to dental structures.[1,2,15–17,23,25,26,41,42] In most of them, the high accuracy of the novel approaches demonstrates the effectiveness of CAD systems in classifying lesions. Despite this, a close variation of 30% was observed between the most and least accurate classification methods of the eligible studies, indicating the need for standardization and search for even more accurate methods.

According to Yilmaz et al,[23] lesion classification is still a problem in the field of computer vision. In their pioneer work, Simon et al[41] demonstrated high correlation between CBCT-based predictions and histological biopsy, suggesting its potential to be an effective, safe and non-invasive diagnostic tool for periapical lesions. However, their procedure requires an expert to manually search the whole lesion for the minimum intensity voxel corresponding to the cavity, which is time-consuming due to the large amount of data to be analyzed. In addition, the simple threshold classifier used in their method may be unreliable due to the fact that CBCT image's gray scale values can be modified according to the patient. Flores et al[15] proposed a semiautomatic solution for the differential diagnosis of two periapical lesions (cyst and granuloma). They adapted the CAD approach to exploit advanced computational algorithms, improving the usability of the system by reducing the amount of user interaction. In their study,[15] lesion classification was performed by three machine learning-based classifiers, the simple threshold, the LDA and the AdaBoost. Due to the small training data set, LOOCV method was used to validate the classification experiments, and

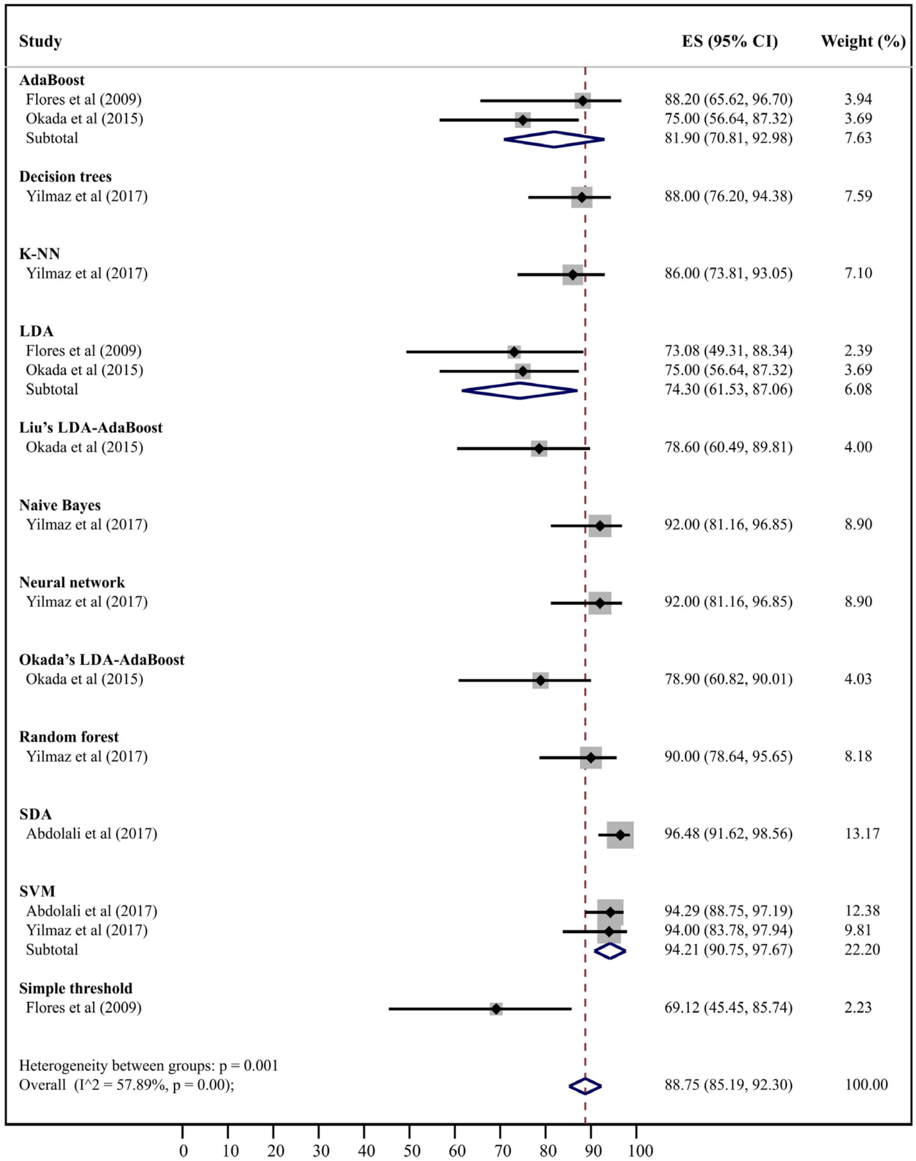

Figure 2 Forest plot with accuracy estimates according to classification method, and pooled estimate using random effects modeling.

the individual classifier with the highest accuracy was AdaBoost, resulting in 88.2% success rate.

Six years later, the same group of researchers[25] extended their previous study,[15] employing a larger dataset, also with the aim of obtaining differential diagnosis between cyst and granuloma. Besides LDA and AdaBoost classifiers, they used the LDA-AdaBoost combination proposed by Liu et al[43] and a new LDA-AdaBoost combination created by them. The newly proposed LDA-AdaBoost combination achieved the highest accuracy, yielding a performance of 85.7% using the seven-fold cross-validation method, and of 78.9% when the LOOCV statistics was applied.

Radiolucent images of periapical cysts were also evaluated in the other two eligible studies.[17,23] According to Shear and Speight,[24] the most common type of maxillofacial cysts is the periapical cyst, being the main reason of chronic swelling in mandible. Still according to them,[24] the second prevalent cyst is dentigerous cyst, followed by the keratocystic odontogenic tumor, which is of great importance due to the high possibility of recurring. Considering this, Abdolali et al[17] proposed a hybrid method based on surface and texture information for accurate and fast classification of these three types of maxillofacial cysts. In their research, an automatic lesion segmentation using symmetry analysis and active contour models, one of the most successful variational models in medical image segmentation,[27] improved the accuracy and speed of this step. This method outperforms the semiautomatic approach used by Flores et al[15] and Okada et al[25] in terms of speed. The random walker algorithm applied by Flores et al[15] and Okada et al[25] is

Table 5 Grade of Recommendation, Assessment, Development, and Evaluation (GRADE) based on the characteristics of studies included in the systematic review and meta-analysis

Quality Assessment							*Summary of Findings*			*Importance*
Computer-assisted image analysis in the diagnosis of maxillofacial radiolucent lesions										
Number of Studies	**Study Design**	**Risk of Bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication Biases**	**Number of Participants**	**Effects**	**General Quality**	
								Random (95% CI)		
4	Diagnostic test accuracy	Not serious	Serious[a]	Not serious	Not serious	Not serious	191	88.75 (95% CI = 85.19–92.30)	⊕⊕ LOW	Critical

High, Very confident that the true effect lies close to that of the estimate of the effect; Low, Limited confidence in the effect estimate, the true effect may be substantially different from the estimate of the effect.; Moderate, Moderately confident in the effect estimate, the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different; Very Low, Little confidence in the effect estimate, the true effect is likely to be substantially different from the estimate of effect.
GRADE approach results in an assessment of the quality of a body of evidence
[a] *Due to the moderate heterogeneity (I^2 = 57.89%), it was downgraded in one level.*

too slow for real-time applications, taking around 1 h to complete for each dataset, whereas segmentation using Abdolali's et al[17] approach is faster, taking around 10 min.

An effective feature extraction approach using a combination of SPHARM coefficients, the most widely used methods in studying anatomical structures such as brain and liver,[44,45] and contourlet features, which provide texture information for physicians in finalizing their diagnosis process, was employed by Abdolali et al.[17] In fact, the authors also improved the accuracy of SPHARM using an adaptive algorithm for performing an extra orthogonalization, which reduced the bias of linear dependency between submatrices in iterative residual fitting estimation. Then, two classifiers were applied, the SVM, which is widely used in pattern classification applications, and the SDA, a modified version of LDA used by Flores et al[15] and Okada et al.[25] According to Abdolali et al,[17] one of the main challenges in medical data processing is the small number of available samples, with SDA being more indicated than LDA. The three-fold cross-validation method was used to validate both classifiers, and SDA performance was superior to SVM in terms of diagnosis accuracy, achieving 96.48 and 94.29% success rates, respectively.

In Yilmaz et al[23] research, CAD was performed between periapical cysts and keratocystic odontogenic tumors. As they evaluated images obtained using the Kodak 9500 CBCT imaging unit and analyzed them with software bundled with the device, which did not have the necessary tools to mark the volume of interest in dental images, they developed software containing both a volume of interest marker and feature extraction tools. Contrary to the other selected studies,[15,17,25] manual lesion segmentation was employed, approximately 127 sections per lesion. Segmentation of radiolucent lesions is technically challenging because the low-dose CBCT images tend to be noisy and the interface between the lesion and the other soft tissues is often extremely vague.[25] Hence, the manual approach is time-consuming, prone to human error and difficult to reproduce.[46] Six different experiment groups were designed by the authors[23] for the diagnosis of lesions, using the 636 different values of textural features obtained from each order statistics and GLCM, combined to the six different selected classifiers. Different parameter combinations were used for each classifier to obtain the best classification performance in these groups. SVM was the individual classification method with the highest accuracy, achieving 100%, 96%, and 94% success rates using the 10-fold cross-validation, the split sample validation, and the LOOCV methods, respectively.

One important point that should be highlighted is the small variety of lesions addressed in the eligible studies. Similar CAD systems have been widely and successfully applied in many medical areas, providing physicians with a valuable tool. However, despite the commonality of dental procedures, the field of dentistry has not yet fully benefited from the advancements of these technologies, and this fact is reflected in the low number of studies eligible for this systematic review. On the other hand, the present study is original and has as an important positive point the low risk of bias of the selected articles, which provides more secure and reliable results, in addition to a level 3 degree of evidence according to Oxford Centre for Evidence-Based Medicine.[47]

Recently, a great interest has been focused on the deep learning-based methods, which have been extensively used for solving complex problems in medical radiology.[48] These methods overcome some limitations of conventional ones, and the most established algorithm among various deep learning models is convolutional neural network (CNN), a class of artificial neural networks that has been a dominant method in computer vision tasks, including various radiological tasks.[49] Despite there still be some challenges in applying CNN to radiological tasks, since 2009, Banumathi et al[16] have aimed to develop an algorithm with radial basis function neural network to diagnose and measure severity of dental cysts with high classification accuracy in dental radiographies, based on gray level properties, circularity and area. While still an incipient method, they concluded that it provided a clear idea to perform an accurate diagnosis of the lesion. Then, further studies on the advantages and limitations of CNN are essential to leverage its potential in diagnostic radiology, with the goal of augmenting the performance of radiologists and improving patient care.

One final aspect that should be addressed is the quality of evidence evaluated using the GRADE tool,[35] which was rated as low (Table 5). The first reason that contributes to this rating is the methodological model presented by the studies, which justifies the decrease in one level. In addition, heterogeneity between studies was responsible for the decrease in one more level, and this is justified by the different methods of lesion segmentation applied by them, and the variety of lesions that were assessed. Future studies are encouraged, with well-designed methodologies, including standardized and reproducible methods of segmentation and analysis.

Conclusion

All individual studies selected for this systematic review suggested the great potential of CAD system for accurately classifying radiolucent maxillofacial lesions, compared to the histological biopsy diagnosis. However, due to the low certainty of evidence found among the studies included in this meta-analysis, a pragmatic recommendation about its use is not possible. Further standardized studies are needed to increase the strength of evidence and to confirm the accuracy of computer-assisted analysis for diagnosing maxillofacial lesions.

Funding

All authors disclose any financial and personal relationships with other people or organizations that could inappropriately influence this research, and declare they have no conflict of interest.

References

1. Nurtanio I, Purnama IKE, Hariadi M, Purnomo MH. Cyst and tumor lesion segmentation on dental panoramic images using active contour models. *JTS* 2011; **22**: 152–8. doi: https://doi.org/10.12962/j20882033.v22i3.66
2. Mol A, Dunn SM, van der Stelt PF. Diagnosing periapical bone lesions on radiographs by means of texture analysis. *Oral Surgery, Oral Medicine, Oral Pathology* 1992; **73**: 746–50. doi: https://doi.org/10.1016/0030-4220(92)90022-I
3. Gijbels F, Debaveye D, Vanderstappen M, Jacobs R. Digital radiographic equipment in the Belgian dental office. *Radiat Prot Dosimetry* 2005; **117**(1-3): 309–12. doi: https://doi.org/10.1093/rpd/nci761
4. İlgüy D, İlgüy M, Dinçer S, Bayırlı G. Survey of dental radiological practice in turkey. *Dentomaxillofac Radiol* 2005; **34**: 222–7. doi: https://doi.org/10.1259/dmfr/22885703
5. Svenson B, Ståhlnacke K, Karlsson R, Fält A. Dentists' use of digital radiographic techniques: Part I – intraoral X-ray: a questionnaire study of Swedish dentists. *Acta Odontol Scand* 2018; **76**: 111–8. doi: https://doi.org/10.1080/00016357.2017.1387930
6. Dölekoğlu S, Fişekçioğlu E, İlgüy M, İlgüy D. The usage of digital radiography and cone beam computed tomography among Turkish dentists. *Dentomaxillofac Radiol* 2011; **40**: 379–84. doi: https://doi.org/10.1259/dmfr/27837552
7. Xambre PA, Valerio CS. E Alves Cardoso Ca, Custódio al, Manzi Fr. The use of digital periapical radiographs to study the prevalence of alveolar domes. *Imaging Sci Dent* 2016; **46**: 179–84.
8. Snel R, Van De Maele E, Politis C, Jacobs R. Digital dental radiology in Belgium: a nationwide survey. *Dentomaxillofac Radiol* 2018; **47**: 20180045. doi: https://doi.org/10.1259/dmfr.20180045
9. El-Beialy AR, El Nigoumi A, Kaddah A, Afify H. Using a single cone-beam computed tomography scan to obtain full occlusal details, with the mandible in centric relation and maximum intercuspation. *American Journal of Orthodontics and Dentofacial Orthopedics* 2018; **153**: 741–6. doi: https://doi.org/10.1016/j.ajodo.2017.09.014
10. Rozema R, Doff MHJ, van Ooijen PMA, Postmus D, Westerlaan HE, Boomsma MF, et al. Diagnostic reliability of low dose multidetector CT and cone beam CT in maxillofacial trauma—an experimental blinded and randomized study. *Dentomaxillofac Radiol* 2018; **47**: 20170423. doi: https://doi.org/10.1259/dmfr.20170423
11. Sönmez G, Koç C, Kamburoğlu K. Accuracy of linear and volumetric measurements of artificial ERR cavities by using CBCT images obtained at 4 different voxel sizes and measured by using 4 different software: an ex vivo research. *Dentomaxillofac Radiol* 2018; **47**: 20170325. doi: https://doi.org/10.1259/dmfr.20170325
12. Idiyatullin D, Corum C, Moeller S, Prasad HS, Garwood M, Nixdorf DR. Dental magnetic resonance imaging: making the invisible visible. *J Endod* 2011; **37**: 745–52. doi: https://doi.org/10.1016/j.joen.2011.02.022
13. Wanner L, Ludwig U, Hövener J-B, Nelson K, Flügge T. Magnetic resonance imaging—a diagnostic tool for postoperative evaluation of dental implants: a case report. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2018; **125**: e103–7. doi: https://doi.org/10.1016/j.oooo.2018.01.005
14. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. *BMJ Qual Saf* 2017; **26**: 484–94. doi: https://doi.org/10.1136/bmjqs-2016-005401
15. Flores A, Rysavy S, Enciso R, Okada K. Non-Invasive differential diagnosis of dental periapical lesions in cone-beam CT. *Medical Physics* 2009; **42**: 566–9.
16. Banumathi A, Mallika JP, Raju S, Kumar VA. Automated diagnosis and severity measurement of cyst in dental X-ray images using neural network. *Int J Biomed Soft Comp Hum Sci* 2009; **14**: 105–10.
17. Abdolali F, Zoroofi RA, Otake Y, Sato Y. Automated classification of maxillofacial cysts in cone beam CT images using contourlet transformation and spherical harmonics. *Comput Methods Programs Biomed* 2017; **139**: 197–207. doi: https://doi.org/10.1016/j.cmpb.2016.10.024
18. Doi K. Computer-Aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics* 2007; **31**(4-5): 198–211. doi: https://doi.org/10.1016/j.compmedimag.2007.02.002
19. Doi K. Current status and future potential of computer-aided diagnosis in medical imaging. *Br J Radiol* 2005; **78**(suppl_1): s3–19. doi: https://doi.org/10.1259/bjr/82933343
20. Nakamoto T, Taguchi A, Ohtsuka M, Suei Y, Fujita M, Tsuda M, et al. A computer-aided diagnosis system to screen for osteoporosis using dental panoramic radiographs. *Dentomaxillofac Radiol* 2008; **37**: 274–81. doi: https://doi.org/10.1259/dmfr/68621207

21. Eadie LH, Taylor P, Gibson AP. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *Eur J Radiol* 2012; **81**: e70–6. doi: https://doi.org/10.1016/j.ejrad.2011.01.098
22. Qin C, Yao D, Shi Y, Song Z. Computer-Aided detection in chest radiography based on artificial intelligence: a survey. *Biomed Eng Online* 2018; **17**: 113. doi: https://doi.org/10.1186/s12938-018-0544-y
23. 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. doi: https://doi.org/10.1016/j.cmpb.2017.05.012
24. Shear M, Speight P. *Cysts of the oral and maxillofacial regions*. 4th edn. Australia: Blackwell Munksgaard; 2007.
25. Okada K, Rysavy S, Flores A, Linguraru MG. Noninvasive differential diagnosis of dental periapical lesions in cone-beam CT scans. *Med Phys* 2015; **42**: 1653–65. doi: https://doi.org/10.1118/1.4914418
26. Nurtanio I, Astuti ER, Purnama IK, Hariadi M, Purnomo MH. Classifying cyst and tumor lesion using support vector machine based on dental panoramic images texture features. *IAENG Int J Comput Sci* 2013; **40**: 29–37.
27. Abdolali F, Zoroofi RA, Otake Y, Sato Y. Automatic segmentation of maxillofacial cysts in cone beam CT images. *Comput Biol Med* 2016; **72**: 108–19. doi: https://doi.org/10.1016/j.compbiomed.2016.03.014
28. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015; **4**: 1. doi: https://doi.org/10.1186/2046-4053-4-1
29. Leeflang MMG, Deeks JJ, Takwoingi Y, Macaskill P. Cochrane diagnostic test accuracy reviews. *Syst Rev* 2013; **2**: 82. doi: https://doi.org/10.1186/2046-4053-2-82
30. The Joana Briggs Institute. *Joanna Briggs Institute Reviewers' Manual: 2014 edition*. Adelaide: The Joanna Briggs Institute; 2014.
31. Whiting PFet al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–36. doi: https://doi.org/10.7326/0003-4819-155-8-201110180-00009
32. Campbell JM, Klugar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, et al. Diagnostic test accuracy: methods for systematic review and meta-analysis. *Int J Evid Based Healthc* 2015; **13**: 154–62.
33. Higgins JPT, Green S. Cochrane Handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration. [updated March 2011]. 2011. Available from: http://handbook.cochrane.org.
34. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539–58. doi: https://doi.org/10.1002/sim.1186
35. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. Grade guidelines: 3. rating the quality of evidence. *J Clin Epidemiol* 2011; **64**: 401–6. doi: https://doi.org/10.1016/j.jclinepi.2010.07.015
36. Wiener F, Laufer D, Ribak A. Computer-Aided diagnosis of odontogenic lesions. *Int J Oral Maxillofac Surg* 1986; **15**: 592–6. doi: https://doi.org/10.1016/S0300-9785(86)80065-5
37. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009; **62**: e1–34. doi: https://doi.org/10.1016/j.jclinepi.2009.06.006
38. Haralick RM, Shanmugam K, Dinstein Its'Hak, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973; **SMC-3**: 610–21. doi: https://doi.org/10.1109/TSMC.1973.4309314
39. Kanazawa K, Kubo M, Niki N. Computer aided diagnosis system for lung cancer based on helical CT images. *Proc Int Conf Pattern Recognition* 1996; **3**: 381–5.
40. Huang Y-L, Wang K-L, Chen D-R. Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines. *Neural Comput & Applic* 2006; **15**: 164–9. doi: https://doi.org/10.1007/s00521-005-0019-5
41. Simon JHS, Enciso R, Malfaz J-M, Roges R, Bailey-Perry M, Patel A. Differential diagnosis of large periapical lesions using cone-beam computed tomography measurements and biopsy. *J Endod* 2006; **32**: 833–7. doi: https://doi.org/10.1016/j.joen.2006.03.008
42. Mikulka J, Gescheidtová E, Kabrda M, Peřina V. Classification of jaw bone cysts and necrosis via the processing of orthopantomograms. *Radioengineering* 2013; **22**: 114–22.
43. Liu X, Zhang L, Li M, Zhang H, Wang D. Boosting image classification with LDA-based feature combination for digital photograph management. *Pattern Recognit* 2005; **38**: 887–901. doi: https://doi.org/10.1016/j.patcog.2004.11.008
44. Gerardin E, Chételat G, Chupin M, Cuingnet R, Desgranges B, Kim H-S, et al. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *Neuroimage* 2009; **47**: 1476–86. doi: https://doi.org/10.1016/j.neuroimage.2009.05.036
45. Mofrad FB, Zoroofi RA, Tehrani-Fard AA, Akhlaghpoor S, Sato Y. Classification of normal and diseased liver shapes based on spherical harmonics coefficients. *J Med Syst* 2014; **38**: 20. doi: https://doi.org/10.1007/s10916-014-0020-6
46. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: a survey. *Med Image Anal* 2015; **24**: 205–19. doi: https://doi.org/10.1016/j.media.2015.06.012
47. Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B. . Oxford Centre for Evidence-based Medicine: Levels of evidence. Grades of recommendation. 2009. Available from: http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ [accessed on 24 June 2018].
48. LeCun Y, Bengio Y, Hinton G, learning D. Deep learning. *Nature* 2015; **521**: 436–44. doi: https://doi.org/10.1038/nature14539
49. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018; **9**: 611–29. doi: https://doi.org/10.1007/s13244-018-0639-9