Clinical update quiz: your free CPD hour

The detection of apical radiolucencies in periapical radiographs: A comparison between an artificial intelligence platform and expert endodontists with CBCT serving as the diagnostic benchmark.

Allihaibi M, Koller G, & Mannocci F. (2025). International Endodontic Journal.

Reviewed and edited by Professor Harcourt and Dr Condon. Compiled by Francis Chan.

This article was originally published by Wiley Online Library and has been edited for brevity and clarity.

Introduction

Apical periodontitis is a common chronic inflammatory condition affecting an estimated 52% of individuals globally. Periapical radiographs are widely used for identifying periapical radiolucent lesions (PARLs) associated with apical periodontitis but have limitations that affect accuracy. These include the two-dimensional nature of these images, anatomical noise, and geometric distortion.

Cone beam computed tomography (CBCT) addresses many of these shortcomings. It exhibits higher sensitivity for the detection of apical periodontitis compared with periapical radiographs, using histology as the gold standard. However, its use is limited by higher associated costs, increased radiation exposure, and limited availability. Therefore, there is a need for improved diagnostic accuracy within the scope of 2D imaging modalities.

Artificial intelligence (AI) has emerged as a promising tool to enhance diagnostic accuracy in various medical fields, including dentistry. Commercially available AI-driven platforms (e.g., Diagnocat) use neural networks trained on large datasets to assist in dental diagnosis. These are designed as screening and decision-support tools. Earlier studies show promising results in periapical radiograph analysis but are limited by small sample sizes and reliance on expert opinion of periapical radiographs as the reference standard, rather than using CBCT.

The aim of this study was to:

1. Determine the accuracy of Diagnocat for detecting apical radiolucencies on periapical radiographs of untreated teeth, compared with CBCT.
2. Compare the performance between Diagnocat and expert clinicians for detecting apical radiolucencies, using CBCT as the reference standard.

Materials and Methods

This was a retrospective non-interventional observational study conducted using existing periapical radiographs and CBCT images drawn from four earlier prospective clinical trials conducted between 2012 and 2022 at Guy's and St Thomas' NHS Foundation Trust (London, UK). The study material consisted of 339 teeth (796 roots).

The inclusion criteria were teeth indicated for primary root canal treatment with a confirmed diagnosis of irreversible pulpitis or pulp necrosis, where both periapical radiographs and CBCT scans were available. Cases were excluded if there was missing radiographic data or inadequate image quality preventing definitive assessment.

For both CBCT and PA radiographs, a PARL was defined as a radiolucency measuring at least twice the width of the periodontal ligament space. For multi-rooted teeth, a lesion was categorised as present if it was detected in at least one root. CBCT scans were taken with a small field-of-view (4 x 4cm) using a 3D Accuitomo CBCT scanner (J. Morita, Kyoto, Japan). The scans were assessed by two endodontists who had access to full CBCT datasets. CBCT scans served as the reference standard for PARLs.

PA radiographs were taken using a standardised paralleling technique and independently evaluated by two different endodontists. Brightness and contrast of all images were adjusted to enhance lesion visualisation. Assessments were blinded to clinical information, Diagnocat results, and CBCT findings.

Diagnocat was used for automatic analysis of the digital periapical radiographs. The default setting (highlighting area with >50% probability of being a PARL) and the "low probability detection" setting (> 30% probability) were used.

Results were recorded independently and compared with assessments from human examiners.

Parameters assessed included diagnostic accuracy (sensitivity, specificity, accuracy) and inter-rater agreement. Sensitivity measures a test's ability to correctly identify those who have a disease (true positives), while specificity measures its ability to correctly identify those who do not have a disease (true negatives). Accuracy is the overall measure of how well a test correctly classifies both positive and negative cases. Receiver-operator characteristic (ROC) curves were plotted to compare Diagnocat and clinician performance. Exploratory subgroup analyses examined accuracy based on tooth type and root canal complexity.

Due to the retrospective design, an a priori sample size calculation was not conducted; however, a post-hoc power analysis was carried out.

Table 1: Correct classification of periapical status (PARL absent or present) by clinicians and Diagnocat, using CBCT as the reference standard

Classification	Tooth level	Root level
Periapical status correctly classified by clinician and Diagnocat	260 teeth (76.7%)	610 roots (76.8%)
Periapical status correctly classified by clinician only	32 teeth (9.4%)	63 roots (7.9%)
Periapical status correctly classified by Diagnocat only	6 teeth (1.8%)	20 roots (2.5%)
TOTAL	**298 teeth (87.9%)**	**693 roots (87.3%)**

Table 2: Diagnostic accuracy of clinicians and Diagnocat for identification of PARLs

Level	Clinician	Diagnocat
Tooth level	Sensitivity: 65.3% Specificity: 97.7% Accuracy: 86.1%	Sensitivity: 47.9% Specificity: 95.4% Accuracy: 78.5%
Root level	Sensitivity: 55.9% Specificity: 97.1% Accuracy: 84.8%	Sensitivity: 39.5% Specificity: 96.4% Accuracy: 79.4%

Results

A total of 76 cases were excluded from the eligible dataset (69 due to missing PA or CBCT radiographs, 7 due to poor image quality). The final sample comprised 339 teeth (796 roots): 137 maxillary teeth (394 roots) and 202 mandibular teeth (402 roots). Most of the sample were molars (96.5% of teeth, 98.2% of roots), with few anterior teeth and premolars.

CBCT identified PARLs in 121 teeth (35.7%) or 240 roots (30.2%). CBCT also identified 2 roots (0.3%) that were undetectable on PA radiographs. Post-hoc power analysis using McNemar's test indicated a statistical power of 0.99.

The intra-examiner Kappa scores (agreement scores) were 0.90 for CBCT and 0.81 for PA.
The interexaminer score was 0.71 for CBCT and 0.73 for PA radiographs. Overall agreement between clinicians and Diagnocat for detecting PARLs was high (tooth level 88.8%, Cohen's κ=0.68; root level 89.5%, Cohen's κ=0.62).
Pooling classifications of clinicians and Diagnocat (default setting), the periapical status was correctly assigned in around 87% of teeth/roots (Table 1). Clinicians consistently demonstrated a higher overall accuracy and sensitivity compared with Diagnocat at both the tooth and root levels (Table 2). Specificity was comparable between clinicians and Diagnocat. Overdiagnosis occurred in less than 1% of roots. When the "low probability detection setting" was used for Diagnocat, the sensitivity increased but the specificity and accuracy decreased. The performance remained inferior to that of experienced clinicians.

Discussion

There was a high level of agreement between clinicians and Diagnocat for the assessment of PARLs on PA radiographs. This likely reflects Diagnocat's training on expert-annotated

PA radiographs, enabling it to replicate human diagnostic patterns while inheriting similar limitations. However, when performance was evaluated against the CBCT reference standard, clinicians outperformed Diagnocat in both accuracy and sensitivity.

This demonstrates clinicians' superior ability to identify true positives, which is critical for reducing missed diagnoses. Specificity was comparable between the two groups, indicating similar effectiveness in ruling out disease. Nevertheless, the low sensitivity observed in both groups underscores the challenges of detecting PARLs with 2D imaging alone.

For maxillary first molars, clinicians exhibited significantly higher sensitivity (65.7% vs 40%) and accuracy (83.5% vs 73.4%) compared with Diagnocat. No statistically significant difference was found for maxillary second molars. There was overall low sensitivity for detection of PARL in upper molars. This may be related to anatomical complexities, such as divergent or convergent anatomy of multi-rooted teeth causing geometric distortion and magnification, particularly in shallow palatal vaults. Sensitivity was lowest for the palatal roots, likely due to proximity to the maxillary sinus floor and superimposition of the buccal roots over the palatal root.

For mandibular molars, the performance gap between clinicians and Diagnocat was smaller. This may be due to easier and more consistent positioning of the image receptor, fewer anatomical obstructions, and less root overlap producing clearer imaging. The lowest sensitivity overall for detection of PARLs was associated with second lower molars (tooth level 36.8% for clinicians, 21.1% for Diagnocat), possibly due to the dense overlying cortical plate obscuring PARLs.

While clinicians identified many more PARLs than Diagnocat, the AI did detect some lesions that clinicians missed. This suggests Diagnocat could be useful as an adjunctive tool rather than a standalone diagnostic method. However, combining both approaches presents challenges. Clinicians can explain their diagnostic reasoning, but the algorithmic processes behind AI systems remain opaque, making it difficult to understand why certain lesions are flagged.

Previous studies reporting higher Diagnocat performance may have been influenced by using clinician assessments as the reference standard and smaller sample sizes. Strengths of the present study include use of CBCT as a high-accuracy reference standard, a large sample size focused on complex cases, and blinded evaluations by experienced endodontists. Limitations include the predominance of molars, which may restrict generalisability to other tooth types, and the exclusive use of Diagnocat, which limits extrapolation to other AI systems using different algorithms and training datasets. The retrospective design also introduces potential selection and information biases.

Further prospective research with more balanced tooth-type distribution, larger sample sizes, and inclusion of less experienced operators is recommended to improve generalisability of the findings.

Conclusion

Clinicians outperformed Diagnocat in both accuracy and sensitivity for detecting PARLs on PA radiographs, while specificity was comparable. The high specificity of both suggests Diagnocat may have value in ruling out disease, but at present it should be regarded as an adjunctive tool rather than a replacement for clinician assessment.

Clinical Update Quiz – October 2025

YOUR FREE CPD HOUR

Complete the Clinical Update quiz online at adavb.org/clinicalupdate (member login required). Each new Clinical Update will be available online from the first business day of each month, as will correct answers from the previous quiz.

Answer at least eight out of 10 questions correctly and you will be awarded ONE FREE CPD hour. Or select your answers below, make a copy and send to CPD Coordinator. Mail: ADAVB, PO Box 9015, South Yarra, Vic 3141.
Email: cpd@adavb.org.

This is an ADAVB member-only service.

Submit your answers by 31 Oct 2025.

Name:______________________________

Member #: __________________________

1. A B C D
2. A B C D
3. A B C D
4. A B C D
5. A B C D
6. A B C D
7. A B C D
8. A B C D
9. A B C D
10. A B C D

August QUIZ 2025 answers

1. B
2. A
3. FALSE
4. B
5. C
6. D
7. TRUE
8. C
9. A
10. C

Questions

1. The inclusion criteria included:

A. Root-filled teeth
B. Teeth diagnosed with irreversible pulpitis or pulp necrosis
C. Teeth with previously initiated endodontic treatment
D. All of the above

2. What was the field-of-view used for the CBCT scans?

A. 4 x 4 cm
B. 6 x 6 cm
C. 8 x 8 cm
D. 12 x 12 cm

3. What proportion of teeth had incorrect classification of the periapical status by both clinicians and Diagnocat?

A. 3.3%
B. 13.3%
C. 23.3%
D. 33.3%

4. What was the accuracy of clinicians in detecting PARLs compared with CBCT at the tooth level?

A. 78.5%
B. 65.3%
C. 86.1%
D. 95.4%

5. Which of the following best describes Diagnocat's performance relative to clinicians in this study?

A. Diagnocat was significantly more accurate than clinicians
B. Diagnocat showed higher sensitivity but lower specificity
C. Diagnocat had comparable specificity but lower sensitivity and accuracy
D. Diagnocat and clinicians performed identically

6. When the "low probability detection setting" was used for Diagnocat:

A. Sensitivity for detection of PARL decreased
B. Sensitivity for detection of PARL increased
C. Accuracy for detection of PARL increased
D. Specificity for detection of PARL increased

7. What were the limitations of 2D radiography for detection of PARLs in maxillary molars?

A. Geometric distortion
B. Overlap of roots
C. Proximity of the maxillary sinus
D. All of the above

8. Sensitivity for detection of PARLs for the maxillary molars was:

A. Lowest for the mesio-buccal root
B. Lowest for the disto-buccal root
C. Lowest for the palatal root
D. Similar for all three roots

9. Which tooth type was associated with the lowest detection sensitivity of PARL on PA radiographs?

A. Maxillary first molars
B. Maxillary second molars
C. Mandibular first molars
D. Mandibular second molars

10. What was the main role of Diagnocat for detection of PARLs as suggested in this paper?

A. Primary tool for the detection of PARL on PA radiographs
B. Adjunctive tool to clinician assessment to rule out PARL on PA radiographs.
C. Primary tool for the detection of PARL on CBCT scans
D. Adjunctive tool to clinician assessment to rule out PARL on CBCT scans.