GMR
The Original

"COMPARISON OF DIAGNOSTIC AND TREATMENT RELIABILITY OF ARTIFICIAL INTELLIGENCE (AI) SOFTWARE WITH HUMAN EVALUATION USING PANORAMIC RADIOGRAPHS"

Chinta Brahmaja, MDS[1], Dr. Challagulla Anusha, MDS[2*], A. Ratnaditya, MDS[3], Hari Pranathi Mallavarapu, MDS[4], Vyshnavi mididoddi, MDS[5], K. Aishwarya, MDS[6]

[1]Chinta Brahmaja, MDS, Department of Pediatric and preventive dentistry, Meghna institute of dental sciences, Nizamabad, mallaram Telangana, EMAIL: brahmajachowdary@gmail.com
[2*]Dr. Challagulla Anusha, MDS, Professor, Department of Pediatric and Preventive Dentistry, Meghna Institute of Dental Sciences, Mallaram(V), Nizamabad, Telangana, Email: challagulla.anusha4@gmail.com
[3]A. Ratnaditya, MDS, Professor and Head of the department, Department of Pediatric and preventive dentistry, Meghna institute of dental sciences, Nizamabad, mallaram Telangana, Email:adityahospital3@gmail.com
[4]Hari Pranathi Mallavarapu, MDS, Department of Pediatric and preventive dentistry, Meghna institute of dental sciences, Nizamabad, mallaram Telangana, Email:mharipranathi@gmail.com
[5]Vyshnavi mididoddi, MDS, Department of Pediatric and preventive dentistry, Meghna institute of dental sciences, Nizamabad, mallaram Telangana, Email:vyshnavisanju.28@gmail.com
[6]K. Aishwarya, MDS, Department of Pediatric and preventive dentistry, Meghna institute of dental sciences, Nizamabad, mallaram Telangana, Email:koorakulaaishwarya@gmail.com

ABSTRACT:
Background/Aim: AI in pediatric dentistry is used for the purpose of making an accurate diagnosis and assisting clinicians, dentists, and pediatric dentists in clinical decision making, developing preventive strategies, and is also a time-saving procedure. The current state-of-the-art artificial intelligence-based natural language modelling has become so convincing that readers cannot tell the difference between human and machine-written texts. The present study was conducted to compare AI software with expert visual examination in diagnosing dental conditions in pediatric patients.
Methodology: A total of 208 panoramic radiographs were collected from patients attending the Department of Pediatric and Preventive Dentistry at Meghna Institute of Dental Sciences. Each radiograph was evaluated independently by an expert team (Pediatric Dentist), who will provide detailed diagnostic reports. The same radiographs will also be analysed using an advanced AI software system. This generated a comprehensive report, including sensitivity and specificity metrics for comparison with expert evaluations.
Results: The inter-examiner kappa values were consistently high across all diagnostic categories, ranging from 0.866 to 0.982, indicating excellent agreement among human examiners. AI also demonstrated good to excellent agreement, with kappa values ranging from 0.722 to 0.955. The highest agreement for AI was observed in the diagnosis of dental caries ($\kappa = 0.955$), closely approaching human agreement levels, while the lowest agreement was seen in periodontal pathology ($\kappa = 0.722$), reflecting moderate to good reliability.
Conclusion: The present study compared the diagnostic performance of human examiners and an AI-based software system in detecting multiple dental pathologies on panoramic radiographs using expert consensus as a reference standard. Overall, human evaluators demonstrated consistently higher sensitivity across all assessed conditions, including dental caries, periodontal pathology, periapical lesions, root fragments or impacted tooth and prosthesis or missing teeth.

KEYWORDS: Artificial intelligence (AI); machine learning; digital dentistry, orthopantomography, pediatric radiograph, mixed dentition.

INTRODUCTION

Accurate diagnosis is essential for the prevention and appropriate management of oral diseases in children. Conventional diagnostic methods in Pediatric dentistry include medical and dental history, clinical examination, caries-risk assessment, and appropriately indicated radiographic investigations. Clinical and radiographic examinations help identify dental caries, developmental abnormalities, eruption disturbances, missing or supernumerary teeth, and other pathological conditions. However, diagnosis in children can be challenging because of developing dentition, anatomical variations, mixed dentition, image interpretation difficulties, and limited patient cooperation. (1)
Artificial intelligence (AI) has emerged as a promising adjunct to conventional diagnostic methods in dentistry. AI-based systems, particularly machine learning and deep learning algorithms, can analyse dental radiographs and identify patterns associated with various dental conditions. Its application is particularly relevant in

Pediatric dentistry, where early and accurate diagnosis can facilitate timely preventive and interceptive treatment and reduce the need for invasive procedures. (2)

AI-assisted analysis of panoramic radiographs has gained increasing attention for the detection and classification of dental caries, missing and supernumerary teeth, developmental anomalies, tooth identification, and other dental conditions. Panoramic radiographs provide a comprehensive view of the developing dentition and are therefore particularly suitable for AI-based assessment. AI may also reduce diagnostic workload, improve consistency, and provide decision-support to dental practitioners, although it should complement rather than replace clinical expertise. (3,4)

However, further evaluation of AI systems against human practitioners using clinically relevant radiographs is required. The present study aimed to compare the diagnostic performance of an AI system in detecting dental conditions using panoramic radiographs, compare its diagnostic performance with expert visual examination, and evaluate the clinical suitability of AI-generated diagnostic outputs.

Methodology: Ethical clearance was obtained from the Institutional Ethics Committee (Ref. No.: 02/MIDS/MDS/PEDO/IEC/2024) before the commencement of the study. This retrospective, cross-sectional, comparative observational study was conducted to evaluate and compare the diagnostic performance of expert visual assessment and artificial intelligence–based analysis of panoramic radiographs in Pediatric dental patients. The study was carried out in the Department of Paediatric and Preventive Dentistry at Meghna Institute of Dental Sciences, Nizamabad, Telangana, over a period of 24 months, from February 2024 to February 2026. A total of 208 panoramic radiographs of Pediatric patients aged 6 to 13 years were retrieved from departmental digital radiographic archives. The inclusion criteria comprised panoramic radiographs of Pediatric patients aged 6–13 years, images of diagnostically acceptable quality, and radiographs demonstrating dentition appropriate for the patient's chronological age. Exclusion criteria included radiographs with motion artefacts or distortion, a history of maxillofacial trauma or surgery, and patients with syndromic conditions. All selected radiographs underwent a two-phase evaluation process consisting of expert visual assessment followed by artificial intelligence–based analysis. During the expert evaluation phase, the evaluators were blinded to the AI results. The radiographs were independently assessed by two experienced paediatric dentists, each with a minimum of 10 years of clinical experience. The expert evaluators assessed the radiographs for the presence of dental caries, periapical lesions, furcation involvement, impacted or unerupted teeth, missing teeth, and other relevant dentoalveolar findings, using predefined radiographic diagnostic criteria. All findings were systematically recorded in a structured data collection proforma, and a consensus diagnosis was established as the reference standard for comparative analysis. The same set of panoramic radiographs was subsequently analysed using the AI Dent software (Version V1.0.3; IDBG AI Dent Global Private Limited). The system, based on neural network algorithms, automatically detected dental pathologies through an independent analysis without access to expert findings. It first performed region of interest (ROI) segmentation, followed by pathology detection using artificial neural networks, generating structured diagnostic outputs. These results were compared with the reference standard to assess sensitivity, specificity, predictive values, and accuracy. Diagnostic agreement between AI and expert evaluation was further determined using Cohen's kappa statistics.

Statistical Analysis: Statistical analysis was performed using SPSS software, version 22. Descriptive statistics were calculated for all variables. Diagnostic performance parameters, including sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy, were computed. The AI assessment was analysed using Cohen's kappa statistic. Receiver's operating characteristic (ROC) analysis was performed to evaluate diagnostic accuracy. A p- value<0.05 was considered statistically significant at a 95% confidence interval. Inter-examiner reliability was assessed by reevaluating 20% of radiographs after a 2-week interval. All obtained data is entered into Microsoft Excel and cross-verified before statistical analysis.

Table 1: Diagnostic Performance of Human and AI Assessment Using Expert Consensus as Reference Standard

Dental parameters		**Correctly Diagnosed (true positive)**	**Mis Diagnosed (false negative)**	**Sensitivity**	**False Negative rate**
				TP/ (TP+FN)	**FN/TP**
Dental Caries	Human	108 (98.1%)	2	108/110* 100 = 98.2%	2/108 *100= 1.8%
	AI	105 (95.4%)	5	105/110* 100 =95.5%	5/105*100= 4.5%
Periodontal Furcation lesion Periodontal bone loss Dental calculus	Human	6 (85.7%)	1	6/7 *100= 85.6%	1/7*100= 14.4%
	AI	5 (71.4%)	2	5/2*100= 71.4%	1/7*100= 28.6%

Periapical Lesion	Human	18 (94.7%)	1	18/19*100= 94.8%	1/19*100= 5.2%
	AI	16 (84.2%)	3	18/19*100= 84.3%	3/19*100=15.7%
Root Fragment Impacted teeth	Human	33 (94.2%)	2	33/35*100= 94.2%	2/35*100=5.8%
	AI	31 (88.5%)	4	31/35*100= 88.5%	4/35*100= 11.5%
Artificial teeth missing teeth	Human	13 (92.8%)	1	13/14*100= 92.8%	1/14*100= 7.2%
	AI	12 (85.7%)	2	12/14*100= 85.6%	2/14*100= 14.3%

The assessment of true negatives and false positives was not achieved; therefore, sensitivity and predictive values could not be calculated using the AI dent software.

Interpretation: For dental caries, human diagnosis showed very high accuracy with a sensitivity of 98.2% and only two misdiagnosed cases, while AI achieved a slightly lower sensitivity of 95.5% with five misdiagnoses. A similar pattern was observed for periodontal pathology, where human examiners showed higher sensitivity (85.6%) compared to AI (71.4%), indicating a greater tendency of AI to miss periodontal disease. In periapical pathology, humans again outperformed AI with sensitivities of 94.8% and 84.3%, respectively.

For root fragments or impacted teeth, both methods performed well; however, human diagnosis maintained a higher sensitivity (94.2%) than AI (88.5%). Likewise, in detecting prosthesis and missing teeth, human examiners showed superior sensitivity (92.8%) compared to AI (85.6%)

Table 2: Human Inter Examiner Kappa or AI Vs Reference Standard Kappa Values

DIAGNOSIS	**INTER EXAMINER**	**AIDENT OBSERVER**
Dental Caries	0.982	0.955
Periodontal Pathology	0.866	0.722
Periapical Pathology	0.946	0.846
Root Fragment / impacted teeth	0.943	0.885
Prosthesis and missing teeth	0.926	0.857

Interpretation: Table 2 presents the agreement analysis using Cohen's kappa values for both inter-examiner (human) reliability and AI performance. The inter-examiner kappa values were consistently high across all diagnostic categories, ranging from 0.866 to 0.982, indicating excellent agreement among human examiners. AI also demonstrated good to excellent agreement, with kappa values ranging from 0.722 to 0.955.

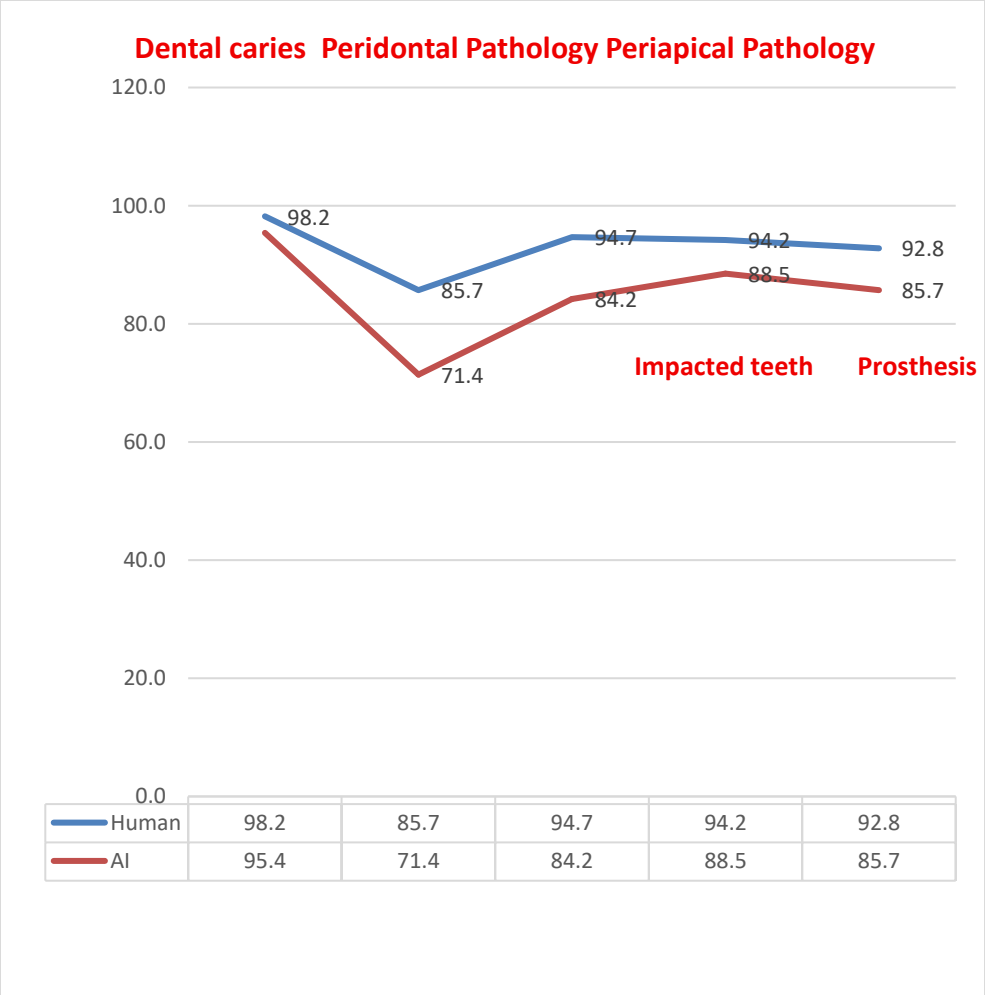
The highest agreement for AI was observed in the diagnosis of dental caries ($\kappa = 0.955$), closely approaching human agreement levels, while the lowest agreement was seen in periodontal pathology ($\kappa = 0.722$), reflecting moderate to good reliability. Overall, although AI shows substantial agreement and promising diagnostic consistency, human examiners continue to demonstrate superior agreement and reliability across all diagnostic categories.

Table 3: Overall pooled sensitivity comparing Human and AI

		Pooled sensitivity	**Accuracy**	**P value**
Human	True Positive	201	96.6%	0.210
	False Negative	7		
AI	True Positive	192	91.4%	
	FalseNegative	16		

Interpretation: The overall pooled sensitivity analysis demonstrates that human evaluators outperformed the AI system in detecting positive cases. Humans identified 201 true positives and only 7 false negatives, yielding an accuracy of 96.6%. In contrast, the AI detected 192 true positives but had a higher number of false negatives (16), yielding a comparatively lower accuracy of 91.4%. The greater number of false negatives in the AI group indicates that it missed more positive cases than human evaluators. However, the difference was statistically not significant (p=0.210).

GRAPH 1: COMPARATIVE ASSESSMENT CURVE SHOWING DIFFERENT DIAGNOSES AND THEIR RATE OF ACCURACY BETWEEN HUMAN AND AI OBSERVERS



DISCUSSION

The present study evaluated the diagnostic performance of artificial intelligence software AIDENT compared with human examiners for detecting dental pathologies, structures, and treatment-related findings on panoramic radiographs. Overall, AIDENT demonstrated promising performance, particularly for dental caries, crowns, restorations, and endodontic treatments, supporting its potential as an adjunctive tool in routine radiographic assessment.

For dental caries, AIDENT demonstrated high sensitivity, approaching that of human examiners. AI-assisted interpretation can improve the detection of subtle carious lesions that may be overlooked during conventional assessment. (5) AI-assisted detection of dental caries has considerable potential to enhance radiographic diagnosis, particularly for lesions that present with subtle or early radiographic changes. Carious lesions, especially in the proximal surfaces of posterior teeth, may be difficult to identify because of anatomical superimposition, variations in image quality, and the limited visual contrast between early demineralization and surrounding tooth structures. AI-based analysis can provide a systematic assessment of the radiograph and draw the clinician's attention to suspicious areas, thereby reducing the likelihood of overlooked lesions. In Pediatric dentistry, where early recognition of caries is particularly important for preventing lesion progression and preserving tooth structure, such assistance may be clinically valuable. The ability of AI to provide consistent and reproducible image analysis may also support clinicians in screening multiple teeth efficiently and in identifying findings that warrant closer clinical evaluation. Thus, rather than replacing clinical judgement, AI

may function as a supportive diagnostic aid that complements visual and radiographic examination and contributes to earlier and more comprehensive caries detection.

Human examiners showed higher sensitivity for periodontal and periapical pathology. This may reflect the greater complexity of these conditions, which often require assessment of subtle radiographic changes and correlation with clinical findings. Nevertheless, AI demonstrated useful detection capability and may assist clinicians by highlighting potentially abnormal areas. (6). AIDENT showed particularly strong performance in detecting crowns, restorations, and endodontic treatments, consistent with previous research demonstrating the effectiveness of AI for identifying dental structures and treatment-related findings on radiographs (7,8). This capability may facilitate rapid radiographic assessment, treatment documentation, and clinical decision-making. For prostheses and missing teeth, human examiners demonstrated higher sensitivity; however, AIDENT showed potential for automated identification of these findings. Previous studies have demonstrated the feasibility of AI-based tooth detection and classification on panoramic radiographs, supporting its application in automated dental charting and assessment (9,10)

The inter-examiner agreement among human examiners was excellent, indicating strong consistency of the reference assessments. AI also demonstrated good to excellent agreement across the diagnostic categories. The highest AI agreement was observed for dental caries, whereas the lowest was observed for periodontal pathology. This suggests that AIDENT may perform particularly consistently when identifying relatively distinct radiographic features, while more complex pathological findings remain challenging.

Overall, AIDENT demonstrated promising diagnostic performance and good reliability across several dental conditions. Its ability to rapidly identify multiple structures and treatments may improve the efficiency and consistency of radiographic interpretation and help clinicians identify findings that could otherwise be overlooked. However, human expertise remains essential, particularly for complex pathological conditions.

The strengths of the study include the use of expert consensus and a reference standard, inclusion of commonly encountered dental conditions in the Pediatric population, and evaluation of multiple diagnostic categories using panoramic radiographs. The high agreement among human examiners further supports the reliability of the reference assessment.

The study has several limitations. True-negative and false-positive cases were not assessed for certain diagnostic categories, preventing calculation of specificity and predictive values for those conditions. The single-centre design, use of a single imaging system, and relatively small sample size may limit generalizability and precision of the findings. In addition, only two-dimensional panoramic radiographs were evaluated. Future multicentre studies involving larger and more diverse populations, multiple imaging systems, and three-dimensional imaging modalities such as cone-beam computed tomography are warranted to further validate AI performance and establish its clinical utility.

CONCLUSION

AIDENT demonstrated promising performance in the detection and classification of dental conditions and treatment-related findings on Pediatric panoramic radiographs. Its ability to identify caries, restorations, crowns, and endodontic treatments highlights its potential to support efficient and consistent radiographic assessment. Although human examiners remained superior for some pathological conditions, AI may serve as a valuable adjunct to clinical judgement by highlighting findings that may otherwise be overlooked. Further multicentre studies with larger datasets and diverse imaging modalities are warranted to validate its clinical applicability.

REFERENCES

1. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology. 2017 Dec 1;2(4).
2. Jaiswal M, Sharma M, Khandnor P, Goyal A, Belokar R, Harit S, Sood T, Goyal K, Dua P. Deep learning models for classification of deciduous and permanent teeth from digital panoramic images. Cureus. 2023 Dec 4;15(12): e49937.
3. Schwendicke FA, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. Journal of Dental Research. 2020 Jul;99(7):769-74.
4. Mertens S, Krois J, Cantu AG, Arsiwala LT, Schwendicke F. Artificial intelligence for caries detection: randomized trial. Journal of dentistry. 2021 Dec 1; 115:103849.
5. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. Dentomaxillofacial Radiology. 2020 Jan 1;49(1):20190107.
6. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. Journal of dentistry. 2018 Oct 1; 77:106-11.
7. Issa J, Jaber M, Rifai I, Mozdziak P, Kempisty B, Dyszkiewicz-Konwińska M. Diagnostic test accuracy of artificial intelligence in detecting periapical periodontitis on two-dimensional radiographs: a retrospective study and literature review. Medicina. 2023 Apr 15;59(4):768.
8. Orhan K, Belgin CA, Manulis D, Golitsyna M, Bayrak S, Aksoy S, Sanders A, Önder M, Ezhov M, Shamshiev M, Gusarev M. Determining the reliability of diagnosis and treatment using artificial intelligence software with panoramic radiographs. Imaging science in dentistry. 2023 Aug 2;53(3):199.

9. Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. Quintessence Int. 2020 Mar 1;51(3):248-57.
10. Kazimierczak N, Sultani N, Chwarścianek N, Krzykowski S, Janiszewska-Olszowska J, Serafin Z, Kazimierczak W. Detection accuracy of an AI platform for dental treatment features on panoramic radiographs–tooth-and patient-level analyses. Scientific Reports. 2025 Dec 15.