Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Presentation of AI-Based Decision Support Biases Clinician Decision Making

Aaron Glick
The University of Texas Health Science Center at Houston
Ali Al-Hatem
The University of Texas Health Science Center at Houston
Nikola Angelov
The University of Texas Health Science Center at Houston
Jennifer Chang
Jennifer.Chang@uth.tmc.edu
The University of Texas Health Science Center at Houston

Research Article

Keywords: Artificial Intelligence, Clinical Decision Support System, Decision Making, Human-Computer Interaction, Diagnostic Errors

Posted Date: June 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9717563/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Medical Informatics and Decision Making on October 7th, 2026. See the published version at https://doi.org/10.1186/s12911-026-03871-w.

Abstract

With the advances in technology in the healthcare field, artificial intelligence (AI) continues to be a growing component to be incorporated in diagnosis, treatment planning, and rendering of care. Historically, dentists' clinical judgement has low inter-rater reliability when viewing similar radiographs. AI has the potential to be used as a means for consensus or identification of pathology that would otherwise be missed by a clinician. Our study aimed to evaluate clinical experience (dental students compared with faculty) and their interactions with different methods of delivering AI information. The dental pathology that was investigated was radiographic furcation where bone loss eventually leads to tooth loss. Each participant was randomly assigned to 1) control, 2) visual survey, or 3) descriptive conditions. Control group received radiographs without AI assistance; the visual group received the same radiographs as the controls with AI-labeled radiographs; and the descriptive group received the same radiographs with a statement indicating whether the AI had detected pathology. A total of 60 clinical faculty and 62 predoctoral students completed the surveys. The differences in accuracy of responses were not significant except for one radiograph out of five, where visual demonstration compared with descriptive demonstration improved accuracy in both experienced and non-experienced dental clinicians. The human-computer interaction is complex particularly in difficult diagnostic situations and can affect clinical decision making based on the presentation of information.

1 Introduction

With the advances in technology in the healthcare field, artificial intelligence (AI) continues to be a growing component to be incorporated in diagnosis, treatment planning, and rendering of care. In dentistry, AI has ongoing research and applications that involve using deep learning models to aid in clinical decision making particularly regarding visual identification of pathologic conditions that would require a potential treatment decision.

Historically, dentists' clinical judgement has low inter-rater reliability when viewing similar radiographs (Flack 1996; Langlais 1987; Meusburger 2023). The low agreement between dentists can be complicated by differences in training, clinical expertise, and radiographic artifacts among other biases. AI has the potential to be used as a means for consensus or identification of pathology that would otherwise be missed by a dentist observer. The AI could be particularly robust if histologic or other extremely reliable gold standard measures were used in large datasets. However, not only are these large and reliable datasets difficult to capture, there are inherent biases when implementing AI clinically. For instance, when presenting an AI generated recommendation visually the clinical experience of the dentist has been shown to bias their clinical judgment leading to potentially incorrect diagnoses (Glick 2022). Diagnostic inaccuracies can lead to incorrect treatment or delays in treatment all potentially having serious consequences for patients (Committee 2015).

As dental systems move towards a future of data dependence with multiple clinical decision support tools available, the presentation and visualization of layered AI-generated information can affect clinical

decisions. Most current research is focused on improving accuracies of AI models, yet less research focuses on the clinical implementation of these models. Accuracies derived from validation and testing phases are an important component to an overall functioning AI-based approach for a decision support system. However, the clinician will ultimately use data derived from the AI-based system to decide on proper treatment. In fact, current clinical accuracies for AI-based systems are gaining FDA approval for sensitivities of 88% and specificities of 87% (Khan 2023). Therefore, AI-based decision support tools today are likely to provide errors in recommendations. The clinician-AI interaction is important to dictate final decision accuracies.

The ultimate clinical decision can and should be based on patient, clinic, and provider experience. Traditional methods of viewing diagnostic information is with primary data of x-ray information. Currently, secondary data can be layered on x-rays to show interpretability/explainability. Limited data exists to demonstrate how the information presented affects clinical decision-making. For instance, presentation of binary AI-based decision or graphical representation of AI-based decision.

2 Materials and Methods

2.1 Study Design

Our study aimed to evaluate and compare faculty and predoctoral students with clinical experience based on three surveys each administered to different faculty and students in terms of: (1) performance, (2) time spent, (3) confidence level in detecting radiographic furcation lesions, (4) and perception regarding the usefulness of AI in improving confidence in detecting radiographic furcation lesions.

Online surveys were distributed to 60 clinical faculty and 60 predoctoral students with clinical patient experience from the University of Texas Health Science Center at Houston School of Dentistry. Approval from the Committee for Protection of Human Subjects of University of Texas Health Science Center at Houston was obtained (HSB-DB-20-1358). Participants' confidentiality was maintained, with only their roles as dental students or faculty members documented, and no personally identifiable information was collected.

2.2 Survey Items

The survey aimed to assess how dental clinicians interact with AI generated diagnostic recommendations. Further, we specifically were investigating the impact of the delivery of this AI generated information and if this delivery method biased clinical decision making based on clinical expertise.

Each participant was randomly assigned to 1) control, 2) visual survey, or 3) descriptive conditions (Table 1). Control group received radiographs without AI assistance; the visual group received the same radiographs as the controls with AI-labeled radiographs; and the descriptive group received the same radiographs with a statement indicating whether the AI had detected pathology (furcation lesion).

A total of five teeth (#2, 3, 18, 9, 30) were used in the surveys (Table 1). These represented all bifurcated teeth with the potential for radiographic furcation lesion. A previously written AI software was used as the AI labeled radiographs. The AI software detected furcation lesions with an overall accuracy of proximal furcation involvement of 80% and buccal/lingual furcation involvement 85% (Chang 2024). All participants in visual and descriptive groups were informed that the accuracy of the AI was 80%.

2.3 Data Analysis

Data was analyzed using Chi-square, Fischer's exact, analysis of variance, or Kruskal-Wallis tests, with the aid of R statistical software (R Core Team 2020). The survey consisted of 9 questions: academic status of participant, pre-survey question inquiring how confident the participant felt in their ability to accurately identify radiographic furcation lesions, 5 questions with maxillary/mandibular radiographs asking to identify the presence/absence of furcation lesions, post-survey question inquiring how confident the participant the participant felt regarding their confidence in their ability to accurately identify radiographic furcation lesions, and post-survey question inquiring how useful the participant believes computer programs are in helping identify radiographic furcation lesions. Answer choices for the questions asking about participant confidence levels / their perception of computer program usefulness was based on a Likert-scale.

The study enrolled a minimum number of 20 subjects within each group as calculated by a power analysis for 80% statistical power to detect differences in correct response percentages of 20% or less between groups. The analysis maintained a two-sided 5% significance level for each test. Analyses were formulated without assuming interaction between groups, while accounting for a within-subject correlation reaching up to 0.8. This framework harnessed a generalized linear mixed effects model tailored to binary outcomes, utilizing a logit link.

To evaluate variations in responses and time allocation for furcation identification, Chi-square tests were used. Additionally, Kruskal-Wallis tests were employed to gauge disparities in confidence levels across the groups.

3 Results

A total of 60 clinical faculty and 62 predoctoral students completed the surveys (20 faculty and 21 students completed the control test, 20 faculty and 20 students completed the visual test, and 20 faculty and 21 students completed the descriptive test). Analysis of deviance table using Chi-squared test was then performed to evaluate the presence of an interaction between the status of a participant (faculty vs student) with the type of test administered.

3.1 Performance

There were five teeth total and out of these three of the cases (#3, 19, and 31) did not show statistically significant differences between clinical experience levels or data presentation style. Tooth #30 showed a

statistically significant difference between faculty and students with 95% of students among all groups answering correctly versus 100% faculty answering correctly despite AI labelled furcation involvement ($p < 0.05$).

Tooth #18 in a previous article (Glick et al 2022) has been shown to pose as a more complex clinical decision. In this case, there was a statistically significant difference in the faculty and students with the students' answers more likely to be congruent with the AI based decision regardless of presentation ($p < 0.05$). Additionally, there were statistically significant differences in the descriptive and visual presentation where both faculty and students' answers were more likely to be congruent with the AI decision in the visually presented condition versus the descriptive condition ($p < 0.05$).

3.2 Time Efficiency

No statistically significant differences existed in the time the task took to be completed between clinical experience levels or data presentation groups for teeth #18, 30, and 31. However, for teeth #3 and 19 with high performance students spent significantly less time allocated to these questions than faculty (Table 2a, b).

Table 2

a. Responses and time taken for radiographic furcation identification of tooth 3

Subject status	Test group	Responses		Time taken in seconds
		No	Yes	(Mean ± SD)
Faculty	**Control**	0	20	12.31 ± 8.25*
	Visual	1	19	13.09 ± 9.34*
	Descriptive	1	19	13.51 ± 11.52*
Student	**Control**	1	20	22.72 ± 14.82*
	Visual	0	20	21.03 ± 34.60*
	Descriptive	1	20	28.64 ± 13.19*

* Students spent significantly more time on the task in comparison to the faculty ($p < 0.05$).

Table 2
b. Responses and time taken for radiographic furcation identification of tooth 19

Subject status	Test group	Responses		Time taken in seconds
		No	Yes	(Mean ± SD)
Faculty	**Control**	1	19	6.37 ± 5.35*
	Visual	0	20	12.16 ± 17.79*
	Descriptive	0	20	9.67 ± 10.23*
Student	**Control**	1	20	4.44 ± 3.85*
	Visual	1	19	6.01 ± 2.63*
	Descriptive	0	21	5.03 ± 5.96*
* Students spent significantly less time on the task in comparison to the faculty ($p < 0.05$).				

3.3 Confidence Levels

Overall, no statistically significant alterations were observed in confidence levels within and across faculty and student groups, except when comparing the visual groups ($p < 0.05$) (Table 3). Notably, within the visual groups, faculty confidence levels decreased post-survey by an average of 0.55 ± 0.22, whereas student confidence levels increased by 0.05 ± 0.17.

Table 3
Participant confidence levels and perceived usefulness of AI

Subject status	**Test group**	**Number of subjects**	**Mean pre-test self-confidence**	**Mean post-test self-confidence**	**Self-confidence change (mean ± SD)**	**Usefulness level of AI (mean ± SD)**
Faculty	**Control**	20	4.05	4.05	0 ± 0.30	4.15 ± 0.22
	Visual	20	4.35	3.80	-0.55 ± 0.22	3.50 ± 0.20*
	Descriptive	20	3.80	3.55	-0.25 ± 0.12	3.75 ± 0.20
Student	**Control**	20	3.19	3.29	0.10 ± 0.18	4.38 ± 0.13
	Visual	21	3.25	3.30	0.05 ± 0.17	4.05 ± 0.25*
	Descriptive	21	2.71	2.86	0.14 ± 0.16	4.19 ± 0.11
Confidence levels were assessed on a scale of 1 (lowest) to 5 (highest).						
*Significant difference between groups ($p < 0.05$).						

3.4 Usefulness of AI

The general perception across both faculty and student groups exhibited positivity regarding the potential of CNN in enhancing clinical confidence for furcation detection. Employing a scoring system spanning 1 (very unlikely to be helpful) to 5 (very likely to be helpful), faculty members from the visual test group expressed the lowest average score of 3.50 ± 0.20, while the students from the control group recorded the highest average score of 4.38 ± 0.13 (Table 3). No statistically significant differences were noted within and between faculty and student groups, except for the visual test groups ($P < 0.05$), where students held a more favorable perspective on AI's helpfulness (average score of 4.05 ± 0.25) compared to the faculty (3.50 ± 0.20) following the furcation identification task.

4 Discussion

This study assessed the performance and perceptions of faculty members and predoctoral students with clinical experience on their effectiveness in detecting radiographic furcation lesions, the time spent, their confidence levels, and their opinions on the potential of AI to enhance confidence in this area. The findings of this investigation shed light on the dynamic relationship between clinical proficiency and the integration of AI in the realm of identifying radiographic furcation lesions.

When examining the responses related to radiographic furcation detection, it was observed that, for the most part, there were no differences between the groups, regardless of participant expertise (faculty members versus students) or presentation condition. However, a notable exception emerged with regard to tooth 18, where the AI model identified a radiolucency over the furcation area. There were statistically significant variations in both faculty and students. Moreover, distinctions were also statistically significant based on the type of data presentation group. In our prior research, this case has been shown to require more complex clinical decision making (Glick et al. 2022).

In most cases with minimal clinical decision-making mental load the mode of presentation (visual or text-based) is irrelevant to the final clinical identification of the provider. However, given a more taxing mental load for the radiographic discrimination, a clinical provider is more likely to be affected by the context of the information. In general, less experienced clinicians are more likely to over rely on AI-based recommendations (Glick et al. 2022). Yet, despite clinical expertise a visual identifier was more likely to lead to an agreement with AI-based recommendation than textual identifier.

While most AI systems are tested and validated within a closed, lab-based setting. Few studies investigate the clinical accuracy of the AI system based on human-computer interactions. This study among others, identifies risks of clinical implementation bias that could affect patient care and safety. As the FDA publishes new guidelines for clinical decision support software to ensure public safety, currently missing is the concept that presentation of information will inadvertently bias healthcare providers (FDA 2022). In addition, few manufacturers validate algorithms in true clinical situations that are heavily influenced by computer-human interactions.

Some limitations of the study is the narrow focus on radiographic furcation lesion identification that could potentially limit the broader applicability of the findings. Additionally, the reliance on the output of

a single CNN model prompts the need for further generalization of explainable AI presentations. The study was administered remotely resulting in varied experimental conditions and it was decided to limit the number of survey questions (limited radiographs) to improve participant engagement and standardization of timing to reduce survey fatigue.

As AI becomes more ubiquitous in clinical practice, it is important to understand the unconscious biases that occur during the AI human-machine interaction. These biases can lead to diagnostic inaccuracies that can ultimately have serious treatment consequences for patients (Committee Diagnostic Error 2015).

5 Conclusion

It was observed that both inexperienced and experienced clinicians tend to over-rely on AI when presented visually compared with textually, yet only in cases that require more demanding clinical decision making. Additionally, inexperienced clinicians are more likely to be more time efficient in cases that require minimal clinical decision making than experienced clinicians.

Declarations

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

AG and AA were responsible for study conception and design, data collection, data analysis, and manuscript drafting. NA provided guidance in data analysis and manuscript development. JC provided mentorship in study design, data collection, data analysis, and manuscript development.

Funding

None.

Acknowledgments

We would like to acknowledge the statistic analysis of Dr. Nat Holland and the study respondents who participated in this project.

Ethics and Consent to Participate Declarations

All participants provided informed consent prior to participation in the survey. The study was conducted in accordance with the ethical standards of the institutional research committee and with the principles of the Declaration of Helsinki. Approval from the Committee for the Protection of Human Subjects at The University of Texas Health Science Center at Houston Institutional Review Board (IRB) was obtained prior to the start of the stuy (HSB-DB-20-1358).

Data Availability

The datasets are available from the corresponding author upon reasonable request.

Consent for Publication

Not applicable.

References

1. Langlais RP, Skoczylas LJ, Prihoda TJ, Langland OE, Schiff T. Interpretation of bitewing radiographs: application of the kappa statistic to determine rater agreements. Oral Surgery, Oral Medicine, Oral Pathology. 1987;64(6):751-6.
2. Flack VF, Atchison KA, Hewlett ER, White SC. Relationships between clinician variability and radiographic guidelines. J Dent Res. 1996;75(2):775–82.
3. Glick A, Clayton M, Angelov N, Chang J. Impact of explainable artificial intelligence assistance on clinical decision-making of novice dental clinicians. JAMIA open. 2022;5(2):ooac031.
4. Meusburger T, Wülk A, Kessler A, Heck K, Hickel R, Dujic H, Kühnisch J. The Detection of Dental Pathologies on Periapical Radiographs—Results from a Reliability Study. J Clin Med. 2023;12(6):2224.
5. Committee on Diagnostic Error in Health Care, Board on Health Care Services. Institute of Medicine, & The National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care. National Academies Press (US); 2015. pp. 217–61.
6. Khan A, Khan KJ, Ghaza MA, Dave T, Shahnoor S, Khan AM, Oduoye MO, Nafula WP, Ubechu SC. Celebrating breakthrough in dental diagnostics: FDA approval of an AI model for diagnosis of periodontal diseases: A correspondence. Health Sci Rep. 2023;6(9):e1573.
7. Chang J, Chang MF, Angelov N, Hsu CY, Meng HW, Sheng S, Glick A, Chang K, He YR, Lin YB, Wang BY. Application of deep machine learning for the radiographic diagnosis of periodontitis. Clin Oral Invest. 2022;26(11):6629–37.

8. Chang J, Bliss L, Angelov N, Glick A. Artificial intelligence-assisted full-mouth radiograph mounting in dental education. J Dent Educ. 2024.
9. US Food and Drug Administration. Clinical Decision Support Software. Guidance for Industry and Food and Drug Administration Staff; 2022.

Table

Table 1 is available in the Supplementary Files section.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 2026SupplementaryMaterial.docx
- Table1.docx