Journal of Dentistry 140 (2024) 104793

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent

Impact of artificial intelligence on dentists' gaze during caries detection: A randomized controlled trial

Lubaina T. Arsiwala-Scheppach [a,b,*], Nora J. Castner [c], Csaba Rohrer [a], Sarah Mertens [a], Enkelejda Kasneci [d], Jose Eduardo Cejudo Grano de Oro [a], Falk Schwendicke [a,b]

[a] Department of Oral Diagnostics, Digital Health and Health Services Research, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Aßmannshauser Straße 4-6, 14197, Berlin, Germany
[b] ITU/WHO Focus Group AI on Health, Topic Group Dental Diagnostics and Digital Dentistry, Germany
[c] Carl Zeiss Vision International GmbH, Tübingen, Germany
[d] Human-Centered Technologies for Learning, Technical University Munich, Germany

ARTICLE INFO

Keywords:
Artificial intelligence
Dental radiography
Caries
Eye tracking
Medical image inspection
Scanpath

ABSTRACT

Objectives: We aimed to understand how artificial intelligence (AI) influences dentists by comparing their gaze behavior when using versus not using an AI software to detect primary proximal carious lesions on bitewing radiographs.

Methods: 22 dentists assessed a median of 18 bitewing images resulting in 170 datasets from dentists without AI and 179 datasets from dentists with AI, after excluding data with poor gaze recording quality. We compared time to first fixation, fixation count, average fixation duration, and fixation frequency between both trial groups. Analyses were performed for the entire image and stratified by (1) presence of carious lesions and/or restorations and (2) lesion depth (E1/2: outer/inner enamel; D1–3 outer-inner third of dentin). We also compared the transitional pattern of the dentists' gaze between the trial groups.

Results: Median time to first fixation was shorter in all groups of teeth for dentists with AI versus without AI, although $p>0.05$. Dentists with AI had more fixations (median=68, IQR=31, 116) on teeth with restorations compared to dentists without AI (median=47, IQR=19, 100), $p = 0.01$. In turn, average fixation duration was longer on teeth with caries for the dentists with AI than those without AI; although $p>0.05$. The visual search strategy employed by dentists with AI was less systematic with a lower proportion of lateral tooth-wise transitions compared to dentists without AI.

Conclusions: Dentists with AI exhibited more efficient viewing behavior compared to dentists without AI, e.g., lesser time taken to notice caries and/or restorations, more fixations on teeth with restorations, and fixating for shorter durations on teeth without carious lesions and/or restorations.

Clinical significance: Analysis of dentists' gaze patterns while using AI-generated annotations of carious lesions demonstrates how AI influences their data extraction methods for dental images. Such insights can be exploited to improve, and even customize, AI-based diagnostic tools, thus reducing the dentists' extraneous attentional processing and allowing for more thorough examination of other image areas.

1. Introduction

Artificial intelligence (AI) has been successfully employed in dentistry from managing workflow in the clinic, e.g., booking and coordinating appointments, to assisting with clinical diagnosis and treatment planning [1,2]. Its efficacy in clinical diagnosis primarily stems from its performance in analyzing diverse types of dental imagery, including photographs, radiographs, transillumination images, and 3-D computed tomography scans. AI methods have been effectively employed for a wide range of tasks on dental imagery, such as tooth classification and segmentation, cephalometric landmark detection, caries identification or predicting the risk of dental complications following third molar extraction [3,4]. Integrating AI into clinical practices may increase the chances of early detection of pathologies and appropriate treatments, resulting in improved patient outcomes and reduced dental care expenditures [5].

* Corresponding author.
E-mail address: lubaina.arsiwala@charite.de (L.T. Arsiwala-Scheppach).

https://doi.org/10.1016/j.jdent.2023.104793
Received 28 August 2023; Received in revised form 15 November 2023; Accepted 24 November 2023
Available online 26 November 2023
0300-5712/

While accuracy gains by using AI have been demonstrated by a wealth of studies [4], further aspects like the impact on care processes or costs have been investigated to a much lesser degree [6,7]. One question of interest, for example, how AI leads impacts the diagnostic process, for example by directing attention towards relevant areas of interest (AOI) and reducing attention towards irrelevant features and areas devoid of pertinent information, or vice versa. A deeper understanding in this direction could help to improve AI systems for diagnostic support, but also to safeguard users from diagnostic biases when using AI or addressing these biases during dental education. To assess the effects AI has on the diagnostic process, mere diagnostic accuracy studies comparing AI against dentists are not sufficient: Instead, prospective clinical studies are needed.

In a recent randomized controlled trial, we compared AI-assisted detection of proximal caries on bitewing radiographs with that of non-assisted dentists and demonstrated a significant increase in sensitivity when using AI [6]. In parallel to recording accuracy estimates, we also employed eye tracking to precisely determine where dentists focus on during image analysis and to record their eye movements while detecting caries [8]. Previous work using eye tracking has shown that dentists employ a task-dependent gaze known as *scanpath* which comprises of 'fixations' (attentional information) and 'saccades' (transitions to attentional areas) [9,10]. Moreover, different types of radiographs are assessed differently. For instance, when examining panoramic radiographs, a holistic representation of the content is formed at a glance [11] and then a systematic spiraling scan pattern [12,13] or a circular scan pattern has been observed [14–16]. For intraoral periapical radiographs, dentists commonly adopt a tooth-by-tooth viewing approach [17]. Using data from the control group of the randomized trial (i.e., dentists not assisted by AI), we showed that the dentists employed a heightened focus on certain image areas, with respect to their task [8]. Also, they generally examined the entire image in a systematic tooth-by-tooth pattern for caries detection [8].

The present study aimed to compare gaze pattern and scanpaths of dentists detecting caries on bitewing radiographs when they are assisted or not assisted by an AI software in the aforementioned randomized controlled trial. Our results may offer valuable insights into the interaction between dentists and AI, and how it may impact their diagnostic performance in a real-world setting. We did not aim to compare the vectorial differences in the gaze patterns. We hypothesized that dentists using AI would exhibit shorter viewing durations, heightened focus on relevant AOI, especially inconspicuous carious lesions, and, in general, more efficient gaze patterns compared to those without AI. The assumption underlying this hypothesis was that the dentists would incorporate the findings of the AI software into their caries detection strategies thus resulting in more efficient gaze patterns.

2. Materials and methods

2.1. Study design

A randomized, controlled, non-blinded, clustered cross-over, superiority trial with an allocation ratio of 1:1 was conducted [6], with an aim to assess the impact of an AI software on the radiographic detection of carious lesions. The trial was conducted using retrospectively sampled imagery material, which was randomly assessed by recruited dentists with and without assistance from the AI software. The trial was registered at Deutsches Register Klinischer Studien (DRKS00022357). Ethical approval was provided by Charité – Universitätsmedizin Berlin (EA/144/20). Written informed consent was obtained from all participating dentists. Reporting of the trial follows the CONSORT-AI checklist [14] as well as the Checklist for Artificial Intelligence in Dental Research [15].

2.2. Participants, sample size, and image data

Recruitment of participants and study conduct took place between October 2020 and January 2021. Participants were dentists employed at dental hospital of Charité – Universitätsmedizin Berlin or in private practices in Berlin, Germany. The study was performed either in dental hospital of Charité – Universitätsmedizin Berlin or at the private practice of the participants. All participating dentists had more than two years of clinical experience (i.e., had finished postgraduate education according to German insurance law). Exclusion criteria for the participants were not being clinically active any longer or having no regular experience with caries detection (e.g., orthodontists or oral surgeons). A total of 22 dentists were recruited, each assessing 20 bitewings. Sample size estimation for the primary outcome of accuracy in this trial (not evaluated in the current study) and randomization of the images and participants have been described in detail elsewhere [6].

Bitewings of permanent teeth taken between 2016 and 2018 at Charité – Universitätsmedizin Berlin were used for the study under an ethics approved protocol (EA4/080/18) [18]. Radiographs of primary teeth or those where assessment was deemed impossible were excluded. This resulted in one hundred and forty bitewing radiographs of the permanent dentition, with at least the crowns of one dental arch being detectable, being included. Most of the images (63 %) were generated using radiographic machines from the manufacturer Dentsply Sirona (Bensheim, Germany), mainly Orthophos XG; the rest using Dürr Dental machines (Bietigheim-Bissingen, Germany). The bitewings radiographs were recorded using dental digital RadioVisioGraphy sensors.

The establishment of the reference test used for confirming the presence of carious lesions and denoting their depth is described elsewhere [18]; a brief description has been provided in the Appendix. The teeth on the bitewing images were stratified according to the reference test for the statistical analysis Lesion depth was defined by two independent reviewers in agreement as follows: E1 denoted lesions into the outer half of the enamel, E2 those into the inner enamel half but not extending into the dentin, D1 those not extending deeper than the outer 1/3rd of the dentin, D2 those not extending deeper than the outer 2/3rd of the dentin, and D3 those extending beyond the outer 2/3rd of the dentin.

2.3. Trial intervention

The intervention was an AI-based application consisting of a radiographic viewing software linked to cloud-based machine learning models for detecting and classifying teeth and segmenting restorations and carious lesions on bitewing radiographs (dentalXrai Pro 1.0.4, dentalXrai GmbH, Berlin, Germany). The software allowed the participant to view the native radiograph as well as its augmented version where detections by the AI software were shown as pixel overlays; the participant could also add, remove, or change findings and generate an automated report (see Fig. 1 for examples of augmented radiographs). At least one week prior to the study, all participants received a handbook of the AI software and were advised to experience the software in advance on a minimum of four independent bitewing radiographs. The control group consisted of conventional radiographic proximal caries detection without any AI assistance.

The intervention was applied as follows: First, a subset of 20 bitewing radiographs was randomly chosen from the overall pool and the sequence to examine this subset was randomly determined, too (for sequence generation, please see [6] and Fig. 1). The participating dentists were then asked to upload the specific bitewing radiograph into the AI viewing software (without any AI support at this stage). Prior to uploading, the participants drew a slip of paper from a pool of 20 slips contained in a sealed opaque envelope (ten indicating to use the AI software and ten not) to determine whether to use the AI software (intervention group) or not (control group). All dentists, irrespective of the allocation, first assessed the native image in the viewing function of

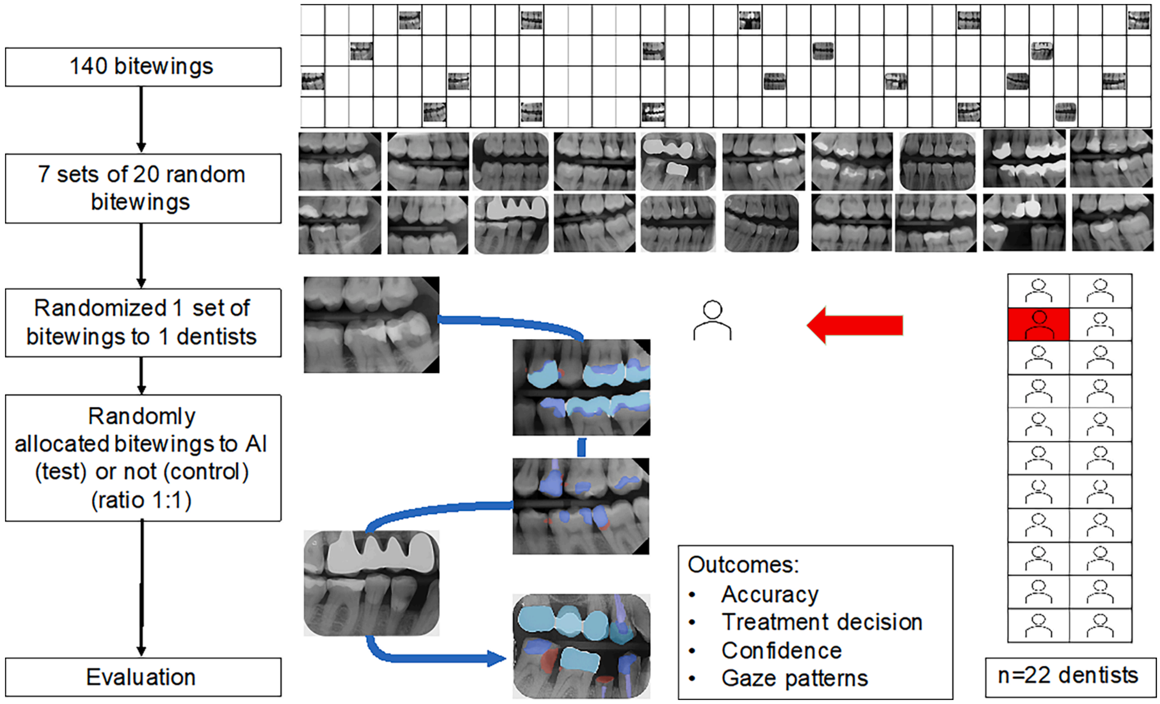

Fig. 1. Flowchart of the randomized clinical trial. From 140 bitewing radiographic images, seven blocks of 20 images were randomly generated. Each of the 22 dentists randomly assessed one block, with images being randomly allocated to the intervention (with artificial intelligence software) or control group in a 1:1 allocation ratio. Different colors on the bitewing images indicate different findings, e.g., blue indicates fillings, crowns, or root-canal fillings, while red indicates carious lesions. Abbreviation: AI, artificial intelligence. Source: [6].

the AI tool, which allowed to adjust image contrast and saturation. In the AI group, dentists could then enable or disable the AI augmentations as needed. Dentists verbally reported any proximal caries detections and their corresponding treatment decisions to the study assistant. Eventually, the dentists concluded the examination of the image and the next one could be uploaded, prior of which another slip of paper was drawn.

2.4. Eye tracker

To record gaze data, the remote eye tracker *SmartEye Aurora* running at 60 Hz was positioned under a monitor which had full high-definition resolution of 1920 * 1080 pixels; Figure S1. The eye tracker used a standard method of Video-based P(pupil)–CR(corneal reflection/ ”glints”) eye tracking, which gathers gaze data by camera sensors that record the participant’s head, and using infrared illumination detects glints on the corneal surface and combines this information with the pupil center detection to estimate the gaze relative to the monitor [19, 20]. The study room was dimly lit, and the participants were unconstrained and positioned approximately 70 cm from the tracker. For the participants from private clinics, the study investigator brought the monitor to their clinic and the experiment was carried out in a dimly lit room in the clinic. An initial 9-point calibration and validation were performed. Recalibration was done if the software indicated that the calibration quality was poor. Gaze data was collected for the whole duration of the study and then pre-processed using the *iMotions* software (version 8.2.22899.4). Event detection was performed using the *iMotions* implementation of the I-VT algorithm, with a minimum fixation duration of 60 milliseconds (ms) and a velocity threshold of 30°/second. The current analysis used the fixations reported from the software, which are interpolated between the left and the right eye. We interpret fixations as the areas of attentional focus related to the stimuli presented on the screen.

2.5. Gaze data and its preprocessing

Data collection resulted in 445 datasets. As five participants unintentionally examined one image twice, we excluded the first time they viewed the image, as it was too short for proper investigation. To ensure gaze data quality, we removed datasets with an average reported gaze signal quality lower than 0.60 (on a scale of 0.0 being the lowest and 1.0 being the highest quality) [19]. Datasets with missing gaze values, though still above the signal quality threshold of 0.60, treated those values as “not a number” and were excluded. These missing data points could be brief moments where the pupil is not detected by the algorithm. This can be due to blink, extreme gaze directions not referenced to the monitor, or other factors [19]. Overall, 80 datasets were excluded by this criterion. Stimulus presentation error resulted in the additional removal of 11 datasets. These exclusion criteria adhere to the recent standard guidelines used in eye tracking research on data quality control [19] and thus these exclusions did not introduce a bias in the study results. Overall and finally, 349 datasets were included for analysis: 170 datasets from dentists without AI and 179 datasets from dentists with AI. Each dentist viewed a median of 18 bitewing radiograph images (nine images in each trial group) and each image was viewed by a median of three dentists.

2.6. Outcomes and covariates

We analyzed fixation behavior of dentists while visually inspecting bitewing radiographic images, stratified by trial group. A range of outcomes and units of measurement were employed. Time to first fixation indicates the amount of time that it takes for a dentist to fixate on a specific AOI from the onset of stimulus and is measured in ms [21]. Fixation count provides information on how many times a dentist returned their gaze to a particular AOI and is measured as a numeric count [21]. Average fixation duration quantifies how long on average a fixation lasted for and is measured in ms [21].

General fixation metrics like average fixation duration offer insights into how professionals holistically process bitewing radiographs. Additionally, the fixation count related to the time per specific task indicates how quickly the relevant information is extracted. Since viewing times per image were variable, an additional measure of fixation frequency per second i.e., number of fixations per second, was calculated. Time to first fixation as well as fixation count and average fixation duration in relevant regions (e.g., AOIs) are indicators of efficient information retrieval. The relevant regions for this study were the teeth, proximal carious lesions, and restorations visible on the bitewing images. For marking the teeth, an (unpublished) in-house tooth detection model, whose findings had been validated by an experienced dentist for each bitewing image, was employed. The carious lesions and restorations were established by the reference test devised for this study, as laid out earlier.

To further investigate the procedural aspect of dentists' gaze, we looked at the transitions of the scanpaths, i.e., how often the gaze transitioned to a neighboring tooth (e.g., from tooth 24 to tooth 25) or somewhere else instead (e.g., from tooth 26 to tooth 37). To account for image dependent patterns, transition matrices were created for six images that were viewed by at least three dentists.

2.7. Statistical methods

Participant characteristics such as age and gender were recorded and used for descriptive analyses. All analyses were performed for the total dataset (i.e., overall) and stratified by presence of carious lesions and/or restorations, and carious lesion depth (E1/2: outer/inner enamel; D1–3: outer to inner third of dentin). The unit for statistical analysis was per image per dentist. The relevant variables exhibited non-normal distributions and thus were summarized using median and inter-quartile range (IQR) and were analyzed using non-parametric tests. Differences in each gaze metric between relevant groups were tested using the Wilcoxon rank sum test and Kruskal-Wallis test, as appropriate, where level of significance was set to $p<0.05$. Missing data was not imputed. No deviation from the intended to the provided intervention occurred.

2.7.1. Gaze stimulus feature preprocessing

To account for any possible spatial offsets in the gaze data, AOI were given an extra pixel padding based on their relative pixel area. The teeth were large enough to simply have the bounding boxes. Based on our study setup, the average size of a bounding box for a tooth was 325.98 by 234.97 pixels, which is approximately 9.2 cm by 6.6 cm on the monitor relative to the participant. The bounding boxes of restorations were given a pixel padding of 3° of visual angle (which approximates to 129 pixels for our specific setup), and bounding boxes of carious lesions were given 3-, 7-, or 10-degrees padding based on whether their area was on the larger, medium, or smaller side of the lesion area distribution, respectively. For fixation behavior analysis, we counted fixations that land in overlapping AOI as a *hit* in both AOI. All statistical analyses and data management were performed using Python (version 3.8 and above) and R (version 4.0.3, www.r-project.org).

3. Results

The trial design is summarized in Fig. 1. Six female and 16 male dentists participated; their mean age was 38 years (range: 27 years to 60 years).

3.1. Gaze patterns

Upon evaluating the fixation frequency in context with the average fixation duration and dispersion, separately, both trial groups showed similar patterns, i.e., longer fixation durations corresponded to slower fixation frequencies whereas shorter durations corresponded to faster fixation frequencies (Figure S2) and that for slower fixation frequencies, lesser image area was inspected, and more image area was covered at faster frequencies (Figure S3).

On comparing the trial groups, we noted that median time to first fixation was shorter in all groups of teeth for dentists with AI versus without AI, although none of the observed differences were statistically significant; Table 1.

Analysis of fixation count showed that dentists with AI had higher median fixation count on teeth with carious lesions and/or restorations (median = 163, IQR = 104, 234) as compared to the dentists without AI (median = 138, IQR = 87, 204), $p = 0.004$; Table 1. Focusing on the teeth with lesions and/or restorations, we noted that this difference stemmed from teeth with restorations where dentists with AI had higher median fixation count (median = 68, IQR = 31, 116) as compared to the dentists without AI (median = 47, IQR = 19, 100), $p = 0.01$; Table 1. For the teeth with carious lesions, dentists with AI had lesser fixations on teeth with D2 lesions (median = 17, IQR = 9, 32), as compared to the dentists without AI (median = 43, IQR = 20, 51), $p = 0.03$; Table 1.

Among dentists with AI, the median fixation duration was longer for teeth with caries (median = 412 ms, IQR = 245, 692) than teeth with restorations (median = 292 ms, IQR = 221, 369), $p < 0.001$; Table 1. Comparing the trial groups with each other, we observed that dentists with AI fixated longer on teeth with carious lesions and/or restorations and shorter on teeth without - specifically, they fixated longer on teeth with caries - as compared to dentists without AI, although these differences were not statistically significant; Table 1.

Scanpath and Transitional Behavior

We also investigated the procedural aspect of the dentists' scanpaths, i.e., how often their gaze transitioned to a neighboring tooth versus non-neighboring tooth, which is indicative of a systematic visual inspection. We noted that both trial groups showed that the highest proportion of transitions were to the neighboring tooth. Often, there were transitions to a tooth in the opposite jaw but to a lesser extent; Fig. 2. The two trial groups were similar in terms of the highest proportion of gaze transitions being to the neighboring tooth but the distributions of the transitions, when formally tested, were different from each other; $p < 0.001$. In order to confirm this lateral tooth-by-tooth visual inspection, transition matrices were created for exemplary six images (dentists without AI: images A to C and dentists with AI: images D to F) that were viewed by at least three dentists to control for image-dependent scanpath patterns; Fig. 3. The transition matrices showed that dentists generally examined the images tooth by tooth (lighter colors along the diagonal in the image indicated higher number of transitions to the neighboring tooth).

Based on the semantic information from Fig. 3, the individual scanpaths for two of these images were qualitatively examined. Generally, the search strategy employed by dentists with AI was less systematic with a lower proportion of lateral tooth-by-tooth transitions and higher proportion of transitions to teeth in the opposite jaw as compared to dentists without AI; while overall, both groups employed a rather systematic search; Fig. 4.

4. Discussion

The diagnostic value of AI tools for clinical dental tasks has been researched extensively but the mechanistic pathway of how it impacts the diagnostic performance of a dentist is not well understood. The present study is one of the few randomized controlled clinical trials on AI in dentistry and it aimed to quantify the impact of an AI software on the gaze behavior of dentists while detecting proximal carious lesions on bitewing radiographs. As hypothesized, the dentists with AI exhibited more efficient viewing behavior (i.e., faster to notice certain image features) for the assigned task of caries detection as compared to their counterparts without any AI support. This finding is corroborated by our previous study which reported that the use of AI increased the dentists' diagnostic accuracy, mainly via increasing their sensitivity for detecting enamel carious lesions [6]. First, the use of AI assistance made the dentists fixate quicker on the teeth with relevant features, i.e., caries and/or restorations, as compared to dentists without AI (note that these differences were not statistically significant). In fact, for enamel carious

Table 1
Distribution of gaze characteristics in the randomized clinical trial, stratified by trial groups.

Gaze metrics		Trial groups: Dentists without artificial intelligence software ($n = 22$)	Trial groups: Dentists with artificial intelligence software ($n = 22$)	p-value
Time to First Fixation, median (IQR), milliseconds, [n]	Teeth with carious lesions and/or restorations	359 (181, 674) [$n = 129$]	319 (146, 652) [$n = 144$]	0.77
	Teeth without carious lesions and/or restorations	384 (236, 612) [$n = 41$]	367 (189, 604) [$n = 35$]	0.58
	Teeth with caries	6598 (2945, 20,669)[a] [$n = 148$]	6586 (2766, 17,672)[a] [$n = 158$]	0.86
	Teeth with restorations	1275 (501, 4075)[a] [$n = 144$]	1217 (498, 3303)[a] [$n = 155$]	0.57
	Teeth with E1 caries	17,128 (8813, 21,540) [$n = 8$]	6722 (4790, 11,515) [$n = 11$]	0.11
	Teeth with E2 caries	9398 (3850, 33,388) [$n = 53$]	7690 (3217, 17,751) [$n = 49$]	0.20
	Teeth with D1 caries	8390 (2955, 20,420) [$n = 37$]	6450 (4157, 20,945) [$n = 45$]	0.85
	Teeth with D2 caries	5146 (3021, 6987) [$n = 16$]	4442 (2183, 22,844) [$n = 22$]	0.90
	Teeth with D3 caries	3300 (2567, 9082) [$n = 17$]	1953 (1567, 11,443) [$n = 16$]	0.53
Total Fixation Count, median (IQR), [n]	Teeth with carious lesions and/or restorations	138 (87, 204)[a] [$n = 170$]	163 (104, 234)[a] [$n = 179$]	0.004
	Teeth without carious lesions and/or restorations	32 (15, 66)[a] [$n = 170$]	25 (6, 51) [$n = 179$][a]	0.01
	Teeth with caries	17 (6, 32)[a] [$n = 155$]	17 (7, 38)[a] [$n = 165$]	0.31
	Teeth with restorations	47 (19, 100)[a] [$n = 146$]	68 (31, 116)[a] [$n = 156$]	0.01
	Teeth with E1 caries	5 (1, 37) [$n = 10$]	21 (14, 45) [$n = 11$]	0.15
	Teeth with E2 caries	10 (2, 22) [$n = 55$]	13 (6, 28) [$n = 52$]	0.12
	Teeth with D1 caries	15 (10, 27) [$n = 39$]	17 (7, 43) [$n = 46$]	0.76
	Teeth with D2 caries	43 (20, 51) [$n = 16$]	17 (9, 32) [$n = 24$]	0.03
	Teeth with D3 caries	25 (18, 31) [$n = 17$]	35 (21, 56) [$n = 16$]	0.09
Average Fixation Duration, median (IQR), milliseconds, [n]	Teeth with carious lesions and/or restorations	337 (251, 413) [$n = 170$]	347 (261, 424)[a] [$n = 179$]	0.50
	Teeth without carious lesions and/or restorations	308 (227, 367) [$n = 155$]	292 (234, 365)[a] [$n = 147$]	0.70
	Teeth with caries	407 (242, 591)[a] [$n = 148$]	412 (245, 692)[i] [$n = 158$]	0.29
	Teeth with restorations	289 (216, 337)[a] [$n = 144$]	292 (221, 369)[a] [$n = 155$]	0.18
	Teeth with E1 caries	530 (468, 664) [$n = 8$]	502 (254, 639 [$n = 11$]	0.78
	Teeth with E2 caries	381 (227, 614) [$n = 53$]	478 (298, 764) [$n = 49$]	0.10
	Teeth with D1 caries	447 (199, 604) [$n = 37$]	385 (226, 687) [$n = 45$]	0.87
	Teeth with D2 caries	486 (285, 569) [$n = 16$]	424 (329, 852) [$n = 22$]	0.55
	Teeth with D3 caries	310 (255, 417) [$n = 17$]	322 (232, 420) [$n = 16$]	0.68

The lesion depth was defined as follows; E1 denoted lesions into the outer half of the enamel, E2 those into the inner enamel half but not extending into the dentin, D1 those not extending deeper than the outer 1/3rd of the dentin, D2 those not extending deeper than the outer 2/3rd of the dentin, and D3 those extending beyond the outer 2/3rd of the dentin.
The p-values are from the comparisons of the medians between the two trial groups and apply to the entire table row. P-values<0.05 are in bold.
Abbreviation: IQR, inter-quartile range.
[a] Each superscript letter denotes two groups with significant differences between them.

lesions, which were most likely to be underdiagnosed by this cohort of dentists [6], the median time taken to fixate on them first by dentists with AI was nearly one-third of that taken by dentists without AI. This is indicative of a heightened focus on relevant areas being employed by dentists when assisted by AI.

Second, the presence of AI assistance prompted dentists to fixate more on teeth with restorations compared to their counterparts without any AI support. It may seem counter-intuitive for the given task that the dentists fixated more on teeth with restorations than with carious lesions which can be explained by the fact that the average fixation duration was longer on teeth with caries for the dentists with AI than the dentists without AI (although this difference was not statistically significant). Hence the dentists with AI examined the teeth with caries for longer durations than those without AI and thus needed to look at them less frequently. This viewing behavior is also indicative of careful inspection of the teeth that the dentists determined to have carious lesions. A second potential reason for this observation may be that the AI overlays for the restorations were usually larger than that of caries and in blue color whereas the overlays for caries were usually smaller in size and in red color. Thus, the AI overlays of restorations likely drew more attention owing to their larger size and greater contrast of the blue color in the grayscale bitewing image. Another potential reason may be that since the dentists were assisted by the AI software in detecting proximal caries on the bitewing radiographs, i.e., the primary task assigned to them in this trial, they may have had more chances to closely inspect other areas of the radiographs. This is further supported by the observation that dentists with AI had lower number of fixations on teeth with D2 caries (larger lesions) as compared to dentists without AI. Studies have showed that obvious features on images demand lower cognitive load from experts [22,23] and the present results suggest that the AI assistance helped to further ease this cognitive load. These findings in combination with the better diagnostic performance of the dentists with AI as compared to those without AI in this trial [6] suggest that the use of AI helped the dentists to reduce extraneous attentional processing for the assigned task and thus allowed for more thorough examination of other areas of the radiograph.

Third, AI support allowed the dentists to fixate for longer durations on teeth with carious lesions and/or restorations and for shorter

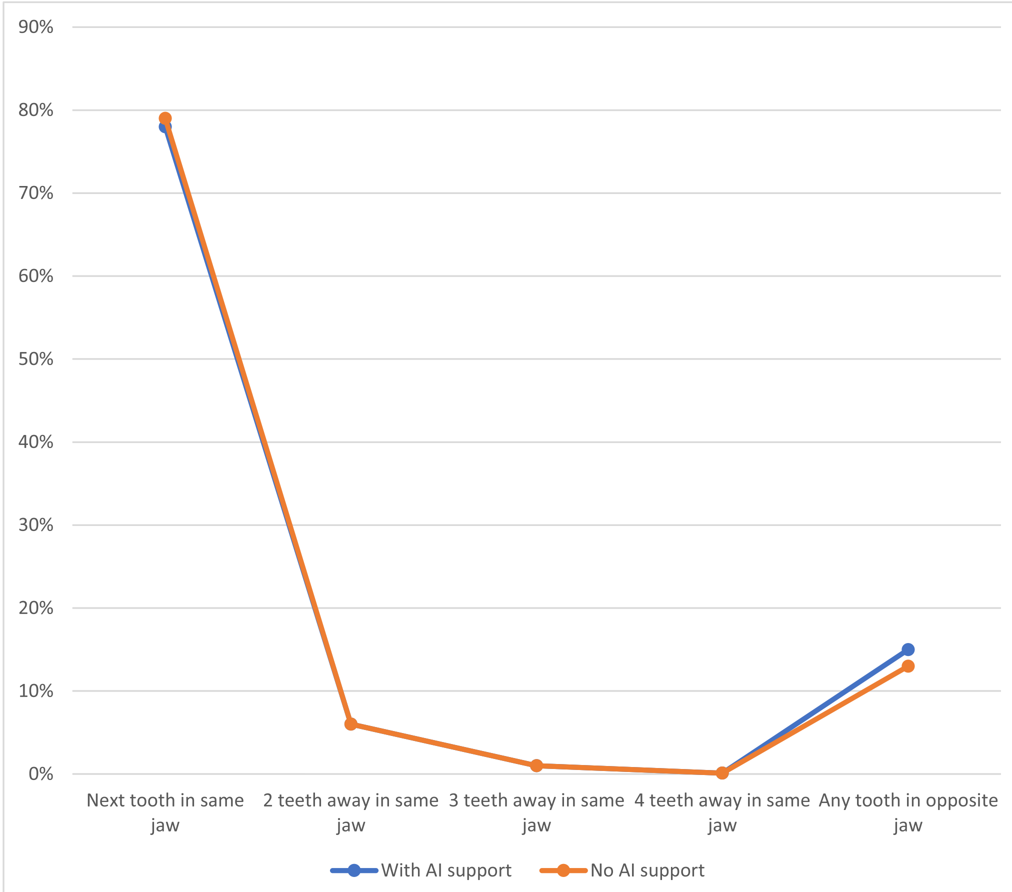

Fig. 2. The proportion of transitions of dentists' gaze across the teeth in bitewing radiographs compared between dentists without and those using artificial intelligence while detecting proximal carious lesions on bitewing radiographic images. Abbreviation: AI, artificial intelligence.

durations on teeth without carious lesions and/or restorations, compared to the dentists without AI (note that these differences were not statistically significant). These findings also allude to the potential role of AI in reducing noise in the cognitive processing of the dentists in the context of radiographic areas not relevant to the assigned task. Additionally, it ties into the primary task of the trial which was to detect proximal caries and not monitor the status of existing restorations. As highlighted earlier, it is also indicative of thorough inspection of the teeth that the dentists determined to have carious lesions.

Fourth, the dentists with AI exhibited a less systematic search pattern while examining the bitewing radiographs as compared to those without AI, i.e., lower proportion of transitions of the gaze were to the neighboring teeth and higher proportion of transitions were to the teeth in the opposite jaw. These observed differences, albeit small in magnitude, suggest that the dentists with AI may have felt a lesser need to systematically search the bitewing images for proximal caries owing to the assistance they received from the AI software, as compared to the dentists without any AI support. Thus, dental clinicians are less likely to use their natural visual search strategies when being assisted with AI technology and rather adapt, quite successfully, to accommodate the AI system into their visual exploration strategies. Also, we must note that the statistical difference in gaze transitions noted between the two trial groups did not account for the effect of time on the distributions. Interestingly, there was no preference observed for left to right inspection or right to left inspection of the bitewing image which would be illustrated by one side of the diagonal in the transition matrix being lighter than the other side of the diagonal.

Additional factors that may influence a dentist's gaze while looking at dental radiographic images include expertise of the dentist, satisfaction of search, visual fatigue, confidence of the dentist in reporting abnormalities, training received, and prior knowledge of the dentist. We have discussed each of these ahead. First, regarding the expertise of the dentist, the inclusion criteria for this study for the participating dentists was to have more than two years of clinical experience (i.e., have finished postgraduate education according to the German insurance law). Exclusion criteria for the participants were not being clinically active any longer or having no regular experience with caries detection (e.g., orthodontists or oral surgeons). Thus, we regard this cohort of dentists as experts in reading carious lesions on bitewing radiographs. Second, we acknowledge that satisfaction of search may have occurred in the study, which occurs when the reader fails to continue to search for subsequent abnormalities after identifying an initial one. We tried to counteract this by not imposing a time limit on the dentists for evaluating the radiographs and recording their findings. Also, the dentists generally turned on the AI software in the second half of their viewing period suggesting that they first performed a thorough evaluation of the radiograph by themselves. Then they used the AI software to confirm their findings and thus did not terminate the investigation earlier. Third, to prevent visual fatigue in the study participants, we assigned only 20 bitewing radiographs to each dentist. Additionally, the randomization of the order in which the images were presented for each dentist controlled for the visual fatigue effect, too, in the last couple of images evaluated. Fourth, the study also recorded the outcome of the dentists' confidence into their detections per image (measured on a visual analogue scale) [6]. However, this is not reported in the present evaluation. Fifth, all participating dentists received standardized instructions before the commencement of the study to prevent any differences in their introduction to the AI software. At least

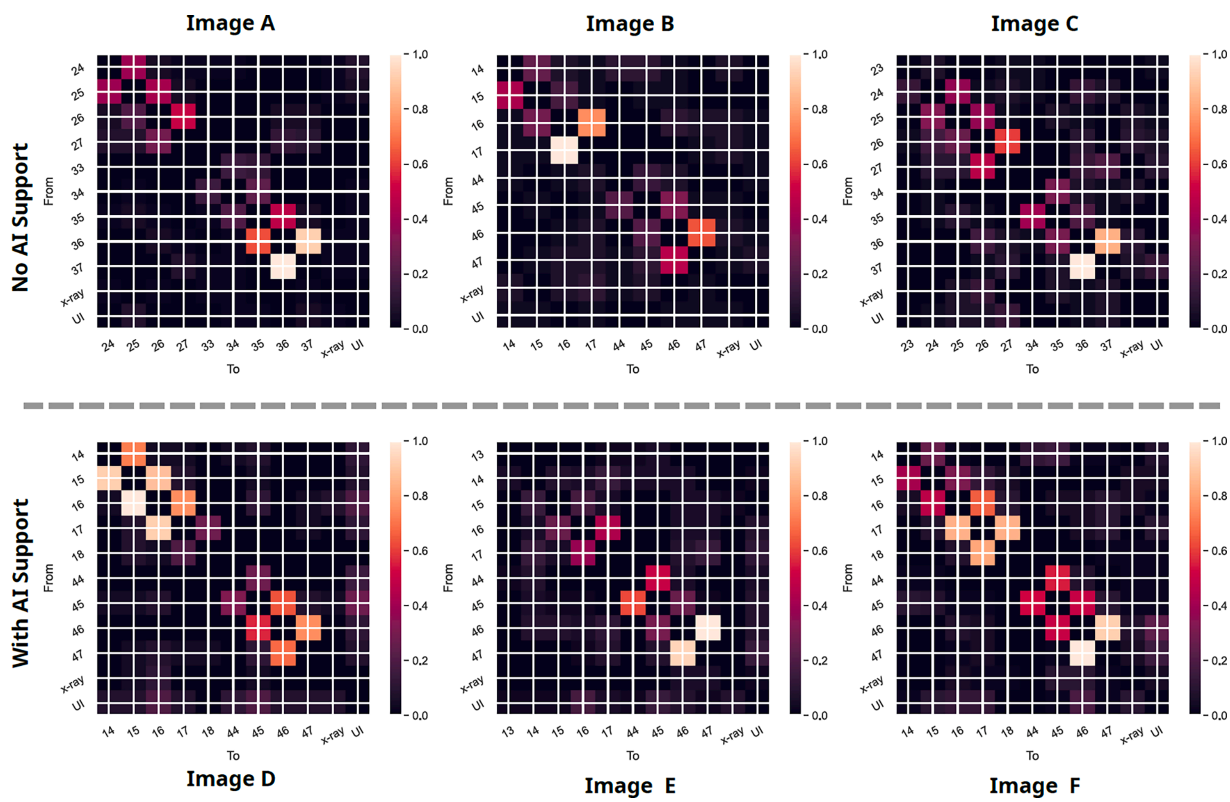

Fig. 3. Transition matrices of the dentists' gaze for six exemplary bitewing images while detecting proximal carious lesions, stratified by trial group. Note that the dentists also looked outside of the bitewing image when using the digital viewing platform and its functionalities. Abbreviation: AI, artificial intelligence; UI, user interface.

one week prior to the study, all participants received a handbook of the AI software and were advised to experience the software in advance on a minimum of four independent bitewing radiographs. Last, as noted earlier, the included dentists had more than two years of clinical experience (i.e., have finished postgraduate education according to German insurance law), were clinically active, and had regular experience with radiographic detection of caries. Additionally, all dentists received a uniform introduction to the AI software; none of them had prior experience with it.

The applications of gaze pattern analysis in dentistry are multifold since the field heavily relies on imagery; from dental education to inferring about the cognitive strategies of dentists [8,22]. Empirical evidence from the present trial on how dental professionals extract data from images [8] coupled with how AI, which has boosted the diagnostic performance of these dental professionals, impacts this data extraction may serve in building better AI-supported diagnostic tools and take us a step further in the direction of explainable AI. While AI systems could contribute to diagnostic support, future studies should explore the potential for AI systems to introduce diagnostic biases. For instance, confirmation bias may result from the dentists over-trusting AI systems. Further research should investigate how can AI contribute to this phenomenon, despite the human users being aware that AI systems may not be perfect. The interplay between AI support and image complexity on the gaze patterns of dental clinicians, while also accounting for their temporal characteristics, should also be further explored.

The present study has a number of strengths and limitations. First, it is one of the few randomized controlled clinical trials in dentistry and uses a range of outcomes to comprehensively characterize the gaze patterns of dentists. Second, and as a limitation, the cohort of bitewing radiographs and dentists was limited and selected. As described, the imagery stemmed from two machines and one clinical center and thus the generalizability of the results on other imagery cannot be expected. Also, a small and selected sample of dentists was enrolled; the sample was younger than the average German dentist and mainly situated in an urban clinic or practice environment. Third, the reference test of whether a tooth had a carious lesion and/or restoration was constituted by five experts; no further validation (e.g., histology) was performed. It can be expected that even five experts and their verdict may not always yield "the ground truth", a caveat we accepted. Similarly, the lesion depth was determined by two reviewers in agreement which may come with limited robustness as well.

Conclusions

In the present randomized controlled clinical trial, analysis of dentists' gaze patterns demonstrated how they were influenced by AI during visual inspection of dental bitewing radiographs. The assessment of one's gaze pattern is an efficient and non-invasive method to collect objective data on the complex interplay of one's cognition and education/training for accomplishing a given task and what implications it may have while interacting with AI-based software.

Role of funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Lubaina T. Arsiwala-Scheppach: Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Nora J. Castner:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – review & editing, Visualization. **Csaba Rohrer:** Methodology, Software, Formal analysis, Data curation. **Sarah Mertens:** Investigation, Project administration. **Enkelejda Kasneci:** Resources.

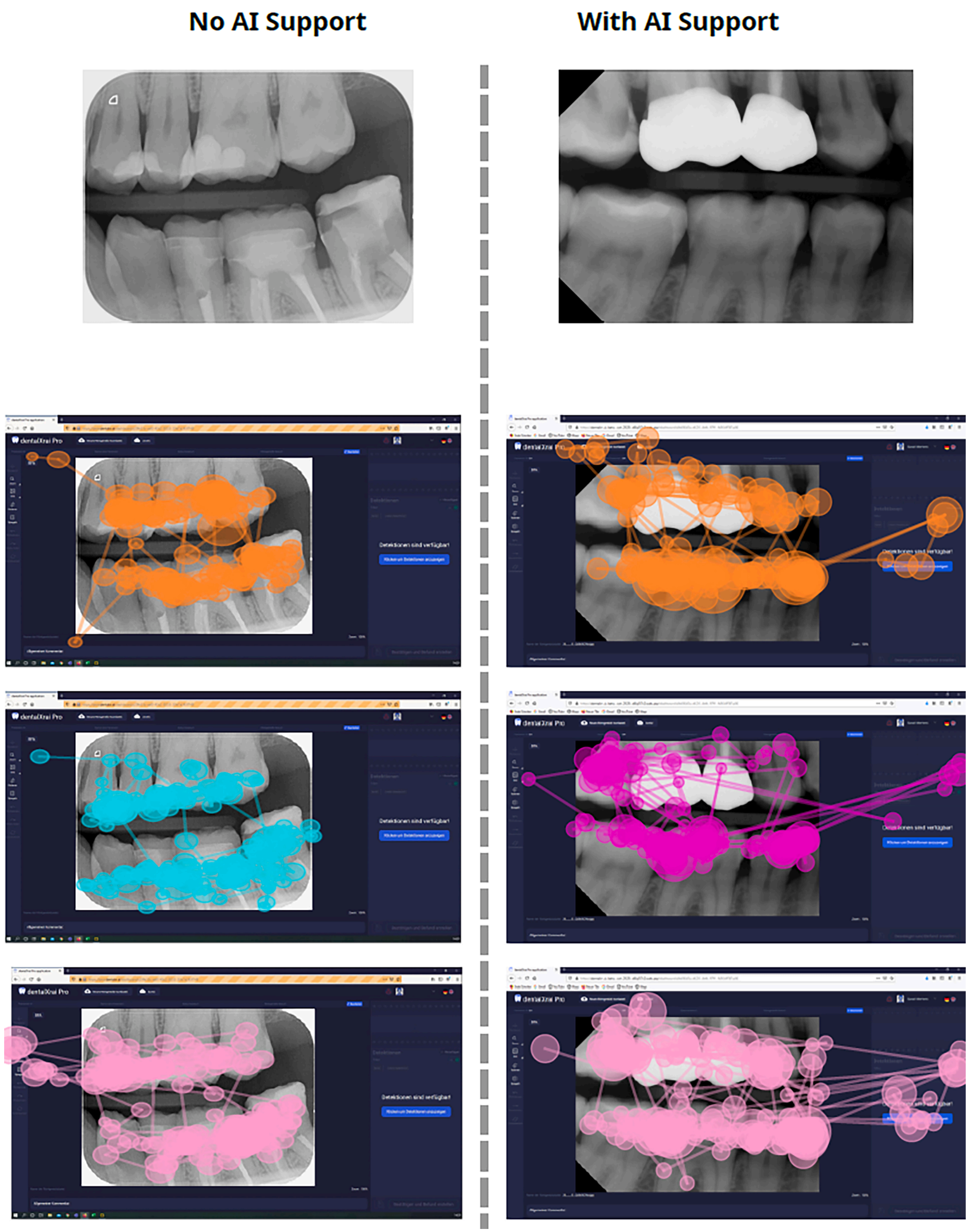

Fig. 4. Visual scanpaths of three dentists for two exemplary bitewing images while detecting proximal carious lesions, stratified by trial group. Note that the dentists also looked outside of the bitewing image when using the digital viewing platform and its functionalities. Abbreviation: AI, artificial intelligence.

Jose Eduardo Cejudo Grano de Oro: Methodology, Software, Formal analysis. **Falk Schwendicke:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision.

Declaration of Competing Interest

FS is a co-founder of a startup company called dentalXrai GmbH. dentalXrai GmbH had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Data availability

Relevant data is available from the authors upon reasonable request.

Acknowledgments

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jdent.2023.104793.

References

[1] Y.W. Chen, K. Stanley, W. Att, Artificial intelligence in dentistry: current applications and future perspectives, Quintessence Int. 51 (3) (2020) 248–257.
[2] N. Ahmed, M.S. Abbasi, F. Zuberi, W. Qamar, M.S.B. Halim, A. Maqsood, M. K. Alam, Artificial intelligence techniques: analysis, application, and outcome in dentistry-a systematic review, Biomed. Res. Int. 2021 (2021), 9751564.
[3] F. Schwendicke, T. Golla, M. Dreher, J. Krois, Convolutional neural networks for dental image diagnostics: a scoping review, J. Dent. (2019), 103226.
[4] L.T. Arsiwala-Scheppach, A. Chaurasia, A. Muller, J. Krois, F. Schwendicke, Machine learning in dentistry: a scoping review, J. Clin. Med. 12 (3) (2023).
[5] F. Schwendicke, J.G. Rossi, G. Göstemeyer, K. Elhennawy, A.G. Cantu, R. Gaudin, A. Chaurasia, S. Gehrung, J. Krois, Cost-effectiveness of artificial intelligence for proximal caries detection, J. Dent. Res. (2020), 22034520972335.
[6] S. Mertens, J. Krois, A.G. Cantu, L.T. Arsiwala, F. Schwendicke, Artificial intelligence for caries detection: randomized trial, J. Dent. 115 (2021), 103849.
[7] M. Nagendran, Y. Chen, C.A. Lovejoy, A.C. Gordon, M. Komorowski, H. Harvey, E. J. Topol, J.P.A. Ioannidis, G.S. Collins, M. Maruthappu, Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies, BMJ 368 (2020) m689.
[8] L.T. Arsiwala-Scheppach, N. Castner, C. Rohrer, S. Mertens, E. Kasneci, J. de Oro, J. Krois, F. Schwendicke, Gaze patterns of dentists while evaluating bitewing radiographs, J. Dent. (2023), 104585.
[9] D. Noton, L. Stark, Scanpaths in saccadic eye movements while viewing and recognizing patterns, Vision Res. 11 (9) (1971) 929–942.
[10] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, J. van de Weijer, Eye Tracking: A Comprehensive Guide To Methods And Measures, 1 ed., Oxford University Press, 2011.
[11] N. Castner, J. Frankemölle, C. Keutel, F. Huettig, E. Kasneci, LSTMs can distinguish dental expert saccade behavior with high "plaque-urracy", 2022 Symposium on Eye Tracking Research and Applications, Association for Computing Machinery, Seattle, WA, USA, 2022, p. 8.
[12] H.L. Kundel, C.F. Nodine, E.F. Conant, S.P. Weinstein, Holistic component of image perception in mammogram interpretation: gaze-tracking study, Radiology 242 (2) (2007) 396–402.
[13] E.A. Krupinski, Visual scanning patterns of radiologists searching mammograms, Acad. Radiol. 3 (2) (1996) 137–144.
[14] T. Grünheid, D.A. Hollevoet, J.R. Miller, B.E. Larson, Visual scan behavior of new and experienced clinicians assessing panoramic radiographs, J. World Fed. Orthod. 2 (1) (2013) e3–e7.
[15] T. Drew, M.L. Vo, A. Olwal, F. Jacobson, S.E. Seltzer, J.M. Wolfe, Scanners and drillers: characterizing expert visual search through volumetric images, J. Vis. 13 (10) (2013).
[16] E. Mercan, L.G. Shapiro, T.T. Brunye, D.L. Weaver, J.G. Elmore, Characterizing diagnostic search patterns in digital breast pathology: scanners and drillers, J. Digit. Imaging 31 (1) (2018) 32–41.
[17] B.P. Hermanson, G.C. Burgdorf, J.F. Hatton, D.M. Speegle, K.F. Woodmansey, Visual fixation and scan patterns of dentists viewing dental periapical radiographs: an eye tracking pilot study, J. Endod. 44 (5) (2018) 722–727.
[18] A.G. Cantu, S. Gehrung, J. Krois, A. Chaurasia, J.G. Rossi, R. Gaudin, K. Elhennawy, F. Schwendicke, Detecting caries lesions of different radiographic extension on bitewings using deep learning, J. Dent. 100 (2020), 103425.
[19] K. Holmqvist, S.L. Orbom, I.T.C. Hooge, D.C. Niehorster, R.G. Alexander, R. Andersson, J.S. Benjamins, P. Blignaut, A.M. Brouwer, L.L. Chuang, K. A. Dalrymple, D. Drieghe, M.J. Dunn, U. Ettinger, S. Fiedler, T. Foulsham, J.N. van der Geest, D.W. Hansen, S.B. Hutton, E. Kasneci, A. Kingstone, P.C. Knox, E.M. Kok, H. Lee, J.Y. Lee, J.M. Leppanen, S. Macknik, P. Majaranta, S. Martinez-Conde, A. Nuthmann, M. Nystrom, J.L. Orquin, J. Otero-Millan, S.Y. Park, S. Popelka, F. Proudlock, F. Renkewitz, A. Roorda, M. Schulte-Mecklenbeck, B. Sharif, F. Shic, M. Shovman, M.G. Thomas, W. Venrooij, R. Zemblys, R.S. Hessels, Eye tracking: empirical foundations for a minimal reporting guideline, Behav. Res. Methods (2022).
[20] Smart Eye Aurora, 2023. https://imotions.com/products/hardware/smart-eye-aurora/#product-accessories. (Accessed 05 October, 2023).
[21] B. Farnsworth, 10 most used eye tracking metrics and terms, 2022. https://imotions.com/blog/learning/10-terms-metrics-eye-tracking/. (Accessed 07 May, 2023).
[22] N. Castner, T. Kübler, K. Scheiter, J. Richter, T. Eder, F. Hüttig, C. Keutel, E. Kasneci, Deep semantic gaze embedding and scanpath comparison for expertise classification during OPT viewing, 2020, p. arXiv:2003.13987.
[23] D.P. Turgeon, E.W. Lam, Influence of experience and training on dental students' examination performance regarding panoramic images, J. Dent. Educ. 80 (2) (2016) 156–164.