applied sciences

MDPI

Systematic Review

The Application of Eye-Tracking Technology in the Assessment of Radiology Practices: A Systematic Review

Elizabeth Arthur and Zhonghua Sun *

Discipline of Medical Radiation Science, Curtin Medical School, Curtin University, Perth 6845, Australia
* Correspondence: z.sun@curtin.edu.au; Tel.: +61-8-92667509

Abstract: The aim of this review is to provide an in-depth analysis of literature pertaining to the use of eye-tracking equipment in the evaluation of radiological image interpretation by professionals in clinical practice. A systematic search of current literature was conducted through the databases of CINAHL, Medline, ProQuest, PubMed, Scopus, Web of Science and Wiley Online Library. A total of 25 articles were included in the final analysis. The literature gathered referenced four main discussions, which were competency assessment, educational tools, visual search behaviour and assistive aid evaluations. The majority of articles (68%) referenced to the competency assessment of professional groups yet appeared to have conflicting results within the categories of speed and eye-metrics. Significant conclusions could be made pertaining to confidence (100%) and accuracy measurements (56%), which suggested a background of higher experience correlates to a higher rate of accuracy and a higher confidence level. Other findings regarding the main themes focused on eye-tracking as an educational tool, where the literature suggests that such equipment may be useful in improving educational repertoire and interpretation technique. Literature pertaining to the visual search behaviour analysis and the evaluation of assistive aids did not provide strong conclusions due to research limitations. Whilst the use of eye-tracking in the analysis of radiological practices is a promising new venture to quantify the interpretation patterns of professionals, undertaking future research is recommended to solidify conclusions and provide greater insight.

Keywords: eye-tracking; radiographic interpretation; healthcare interpretation patterns; medical assessment

Citation: Arthur, E.; Sun, Z. The Application of Eye-Tracking Technology in the Assessment of Radiology Practices: A Systematic Review. *Appl. Sci.* **2022**, *12*, 8267. https://doi.org/10.3390/app12168267

Academic Editors: Yingke Xu and Yubing Han

Received: 7 July 2022
Accepted: 17 August 2022
Published: 18 August 2022

1. Introduction

The development of eye-tracking technology has allowed for the quantification of visual search behaviour, cognitive thought and the analysis of human performance in the interpretation of stimuli [1,2]. Progressing from the 19th century, the advancement of computer-based eye-trackers prioritise a non-invasive passive approach to monitoring participants, facilitating a wide variety of studies across many disciplines [2,3]. This recently has included the healthcare and medical fields, where stimuli interpretations can be studied—for example the interpretation of electrocardiograms by Davies et al. [3,4].

In a separate discussion, the use of radiological images is an essential component of the medical industry, providing insight into patient wellbeing, medical conditions and potential treatment plans. Diagnostic error, defined as an "incorrect, delayed or missed diagnos(i)s" has the capacity to adversely impact patient wellbeing, safety and care [5,6]. Rates of diagnostic error are consistent at around 3–5% annually (persistent irrespective of time and intervention), resulting in a global estimation of 40 million diagnostic errors annually [6]. As a major contributor to diagnostic processes, the radiological space can play a considerable role in diagnostic error, as shown by various intervention and literature into the field [5–7].

Despite the research, documentation and attempted intervention of many techniques to reduce radiologic error [6,7], there are some interesting discussions from a review by Brunyè et al. [8] suggesting a place for eye-tracking technology in the advancement of

professional experience and expertise in radiographic interpretation. There appears to be a lack of competency-based assessment in medical education for health professionals, as there have been challenges in developing educational material which has the capacity to "create meaningful, relevant and repeatable outcome-based assessments for use in graduate medical education, residency and fellowships" [8]. This suggested use of gaze-monitoring technology appears to be one potential method for the further understanding of radiology. Potential for continuous aptitude assessment, assessment of visual search behaviour and the establishment of assistive tool use are also considerable avenues which will be considered during this review. This thematic analysis is supported by Ganesan et al. [9], who also consider eye-tracking technology as a potential tool for strength and weakness analysis of radiological interpretation.

This systematic review aims to provide a primary understanding of the utilisation of eye-tracking and its potential to be used in the radiological space. The goal of this paper is to provide a review of the current literature and to contribute a summative analysis, providing an opportunity to discuss developments, future research areas and potential quantification of radiological interpretation styles and techniques.

2. Methodology

This systematic review was performed in accordance with the preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidelines [10,11]. No ethics approval was required.

2.1. Search Strategy

A systematic search of the literature was conducted within the first quarter of 2022, utilising relevant databases of; CINAHL, Medline, ProQuest, PubMed, Scopus, Web of Science and Wiley Online Library. A developed search strategy was used across all databases.

In the development of the search strategy, specific key terms were included or excluded. This is evident in phrase 1.0, where numerous synonyms were included for "eye-tracking" to encompass the variation in terminology across literature. In addition, there was a restriction for phrase 2.0 in which the terms "assessment" or "judgement" were omitted due to preliminary search techniques resulting in literature referencing computerised AI interpretation analysis rather than human interpretation analysis. Similar exclusions were made for phrase 3.0, in which "medical image" and other synonyms were absent to avoid inclusion of literature focusing on other medical stimuli, such as electrocardiograms.

The search strategy was utilised across the abstract of articles within the databases, as preliminary findings found title searches too narrow and specified. On the contrary, the "keyword" and full text searches produced literature results which were too numerous and irrelevant for an adequate review.

Primary specifications to the literature search included a date range, language and type of production. The date range from 2007 authorised inclusion of research from a fifteen-year time period, which was designed to establish a degree of relevance and currency. Another consideration was the use of the English language, as the authors are monolingual. The exclusion of literature which did not fit "original primary research criteria" was also essential for the formulation of the review. Peer reviewed articles, and those which were full texts were selected for analysis.

In the final stage of eligibility audits, the literature was screened by title, abstract and full text in three separate stages (Table 1). In the analysis of articles, a critique was made of key elements. This is demonstrated in Section 3.

Table 1. Use of a combination of search terms to locate relevant studies.

Boolean Operator	Term	Field
1.0	"eye tracking" OR "eye monitor *" OR "eye-tracking" OR "gaze monitor *" OR "gaze tracking" OR "eye gaze *"	Abstract
2.0 AND	Interpret * OR competenc * OR performance *	Abstract
3.0 AND	Radiograph * OR xray * OR x-ray * OR "general radiograph *" OR "computed tomography" OR "magnetic resonance *"	Abstract

2.2. Data Extraction and Analysis

The following details from eligible studies were extracted: authors and year of publication, sample size, study design and methodology and key findings with regard to the use of eye-tracking device in interpreting radiographic/radiological images. Referencing searching and data extraction was performed by one assessor (E.A) with results validated by another assessor (Z.S).

3. Results

From the search strategy and applied criteria, 348 articles were initially considered for analysis. After screening titles, abstracts as well as full-texts of the relevance, a total of 25 articles were found to meet our selection criteria and thus included in the final analysis (13–16, 18–38) (Figure 1).

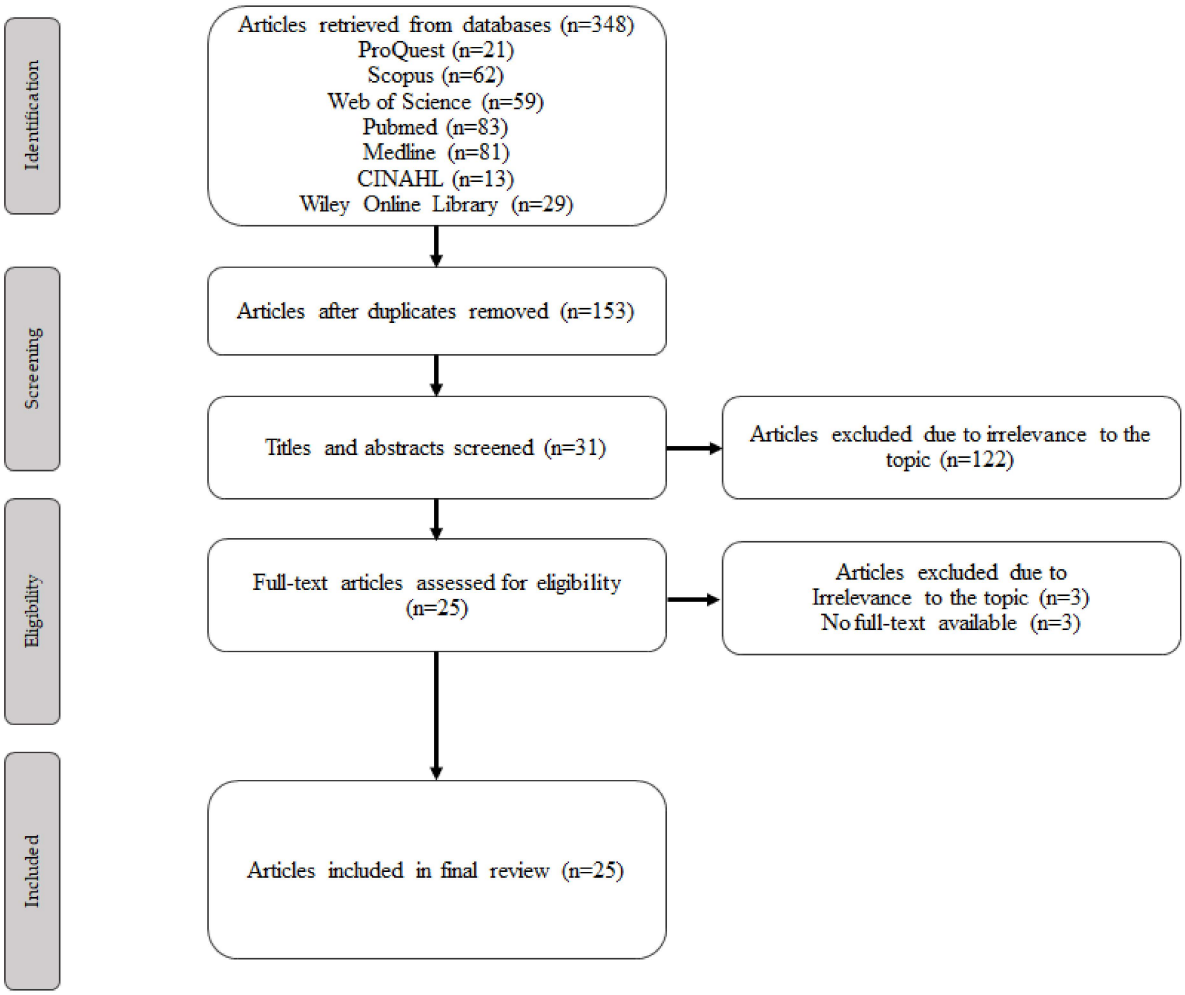

Figure 1. Flow chart showing search strategy to identify eligible studies.

The literature has been graded using the "Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields", by Kmet et al. [12] (Table 2). The use of this checklist focused on quantitative analysis, however, was adequately generic to address different research styles. Whilst most literature generated would be categorised as a cross-sectional observational style of study, there were exceptions such as McLaughlin et al. [13] Quen et al. [14] Kok et al. [15] and Kok et al. [16] which presented in either a non-randomised controlled trial or randomised controlled trial format. As such the varied research styles, we deemed it appropriate to utilise a generalised quantitative research review rather than checklists such as STROBE which is specified only towards observation trials [12,17]. The generic analysis style and presentation of the table by Kmet et al. [12] provided the best method of analysis. The articles have been scored individually and graded as to their strengths and limitations.

Most articles (88%) consisted of clear study designs and plans, with definite explanations of methodology, results and resolute conclusions. The majority of studies (92%) utilised some statistical analysis methods, which provides adequate understanding of proposed conclusions. Most of the articles (81%) outlined flaws and controlled variables, however some research failed to state their limitations [18–22]. Similarly, there appeared to be a lack of establishment of the study style across all studies which made determining the review approach difficult. Most appeared to be of an observational design, and few of interventional design (namely articles [13–16]). From in-depth analysis of the 25 articles selected, it appears most exist within the observational cross-sectional style, but some deviations occur with breaches into randomised controlled trials, non-randomised controlled trials and part-cohort studies (Table 3). This lack of explicit study design and combination of research styles provided reasoning for the use of a generic quantitative data assessment.

An additional characteristic of these studies is the sample size and variation of sample sizes. It was most common amongst the articles for authors to recruit an approximate 30 participants, however this varied from 6 to 136 participants. There is also consistent variation between participant groups within the studies, as not all articles focused on one educational experience. It is difficult, therefore, to compare statistical results across papers as the literature is discussing different levels of expertise in differing professions.

In the investigation of the articles listed in Table 3, it is noted that the majority of articles originate from Western Europe (52%), and 88% of articles originate from countries of Western influence. The remaining articles are from Hong Kong, Malaysia, Saudi Arabia and Japan. Focusing on location, there is also a consideration to be made as it is common for the same research teams to be conducting all of the eye-gaze metric analysis work for such region. For example, the two articles originating from the Netherlands by Kok et al. [15,16] were conducted by research teams with common researcher compositions. This is also found in two articles from the UK by McLaughlin et al. [13,23]. The repetition of the similar research teams is an important consideration when analysing the literature from a region and evaluating for bias, although they were all included in the review due to different study designs with inclusion of different categories of participants.

Table 2. Quality assessment of relevant articles via Kmet et al. [12].

Article	Year	Description of Question and Aim	Study Design	Method of Participant Selection	Participant Demographics	Randomisation	Investigator Blinding	Participant Blinding	Outcome Measurements	Sample Sizing	Analysis Methods	Variance Estimation	Confounding Variables	Report of Results	Conclusions	Quality Assessment
McLaughlin et al. [13]	2021	2	2	2	2	2	N/A	N/A	2	2	2	2	1	2	2	23/24 96%
Quen et al. [14]	2021	2	2	1	2	N/A	N/A	N/A	1	1	0	0	1	1	2	13/22 59%
Kok et al. [15]	2016	2	2	2	2	2	N/A	N/A	2	2	2	2	2	2	2	24/24 100%
Kok et al. [16]	2015	2	2	2	2	N/A	N/A	N/A	2	2	2	2	2	2	2	22/22 100%
Brams et al. [18]	2020	2	1	1	2	N/A	N/A	N/A	2	2	2	2	1	2	1	18/22 82%
Crowe et al. [19]	2018	2	2	2	2	N/A	N/A	N/A	2	2	2	2	1	2	2	21/22 95%
Lèvêque et al. [20]	2019	2	2	0	2	N/A	N/A	N/A	2	2	2	1	2	2	2	19/22 86%
Turgeon and Lam [21]	2015	2	1	2	2	N/A	N/A	N/A	2	2	2	2	1	2	2	20/22 91%
Wood et al. [22]	2013	2	1	2	2	N/A	N/A	N/A	2	2	2	2	2	2	2	21/22 95%
McLaughlin et al. [23]	2017	2	1	2	2	N/A	N/A	N/A	2	2	2	1	1	2	2	19/22 86%
Gnanasekaran et al. [24]	2022	2	1	1	1	N/A	N/A	N/A	2	2	2	0	2	2	2	17/22 77%
Giovinco et al. [25]	2015	2	1	1	1	N/A	N/A	N/A	2	1	2	2	0	2	2	16/22 73%
Hanley et al. [26]	2017	2	2	2	2	N/A	N/A	N/A	2	1	2	2	1	2	2	20/22 91%
Kelly et al. [27]	2016	2	2	2	2	N/A	N/A	N/A	2	1	2	2	1	2	2	20/22 91%
Vogel and Schulze [28]	2021	2	1	2	2	N/A	N/A	N/A	2	2	2	2	2	2	2	21/22 95%
Bahaziq et al. [29]	2019	2	1	2	2	N/A	N/A	N/A	1	0	1	0	1	1	2	13/22 59%
Bertram et al. [30]	2016	2	2	2	2	N/A	N/A	N/A	2	2	0–1	2	2	2	2	21/22 95%
Botelho et al. [31]	2019	2	1	2	2	N/A	N/A	N/A	2	1	1	2	2	2	2	19/22 86%
Matsumoto et al. [32]	2011	2	1	2	2	N/A	N/A	N/A	2	2	2	2	1	2	2	20/22 91%

Table 2. *Cont.*

Article	Year	Description of Question and Aim	Study Design	Method of Participant Selection	Participant Demographics	Randomisation	Investigator Blinding	Participant Blinding	Outcome Measurements	Sample Sizing	Analysis Methods	Variance Estimation	Confounding Variables	Report of Results	Conclusions	Quality Assessment
Kelahan et al. [33]	2019	1	1	0–1	2	N/A	N/A	N/A	2	1	1	0	2	2	2	15/22 68%
Hanna et al. [34]	2018	2	2	2	2	N/A	N/A	N/A	2	2	2	2	2	2	2	22/22 100%
Ba et al. [35]	2020	2	1	0	2	N/A	N/A	N/A	2	1	2	2	1	2	2	17/22 77%
Venjakob et al. [36]	2016	2	2	2	2	N/A	N/A	N/A	2	1–2	2	2	1	2	2	21/22 95%
Rubin et al. [37]	2015	2	2	2	2	N/A	N/A	N/A	1	1	2	2	2	2	2	20/22 91%
Krupinski [38]	2019	2	1	0	0	N/A	N/A	N/A	2	0	2	2	2	2	2	15/22 68%

For these quantitative studies, 14 items were scored depending on the degree to which the specific criteria were met: Yes = 2, partial = 1 and no = 0. Items not applicable to a particular study design were marked as N/A and were excluded from the calculation of the summed score.

Table 3. Study characteristics of eligible studies that were reviewed.

Author and Date	Purpose	Setting	Participants	Research Style	Relevant Findings
			COMPETENCY ASSESSMENT EVALUATION		
Brams et al. 2020 [18]	To explore main theories of radiologic search patterns across groups of different experience using eye-gaze metric analysis in the interpretation of chest x-rays	Belgium	***n* = 41** 15x medical students 2nd–4th years (MS) 13x medical residents (MR) 13x radiology residents (RR)	Observation; Cross-Sectional	The gaze metrics demonstrated that the RR were able to detect pathology at above chance level, in comparison to the MR and MS which at chance level. Additionally, RR and MR had faster response times and longer average fixation durations in comparison to MS.
Crowe et al. 2018 [19]	To differentiate the gaze/scanning patterns across different educational levels in the interpretation of brain tumour images (MRI)	UK	**Experiment 1 and 2: *n* = 35** 18x undergraduate students (excl. medicine, dentistry and veterinary) 10x medical students (3rd or 4th years) 7x experts (trainees and consultant neurologists and consultant neuroradiologists)	Observation; Cross-Sectional	There was a clear distinction between experience and accuracy and sensitivity. Additionally, the experts scanning patterns which were similar to each other whilst the medical students did not.

Table 3. *Cont.*

Author and Date	Purpose	Setting	Participants	Research Style	Relevant Findings
Lèvêque et al. 2019 [20]	To evaluate the difference of different levels of expertise in the interpretation of mammograms via the use of eye-tracking equipment	Belgium	***n* = 8** 3x expert radiologists 3x trainee radiologists 2x physicists	Observation; Cross-Sectional	There was no difference in the mean fixation duration amongst experts, however, trainees had a shorter mean fixation. Physicists' fixation duration was significantly longer than experts. Trainees and physicists deviated from the expert focus points, however trainees have a greater focus on area of interest than physicist participants.
Turgeon and Lam 2015 [21]	To compare the visual search strategies of oral and maxillofacial radiologists (OMR) and dental undergraduate students in the interpretation of panoramic dental x-rays via eye-tracking	Canada	***n* = 45** 30x 4th year dental students 15x OMRs	Observation; Cross-Sectional	The OMR covered more gaze distance than students for normal anatomical radiographs. For pathological images, the OMRs demonstrated faster analysis, fewer eye fixations, fewer saccades and less time to first fixation (within area of interest). The OMR group covered less distance than students for obvious pathologies.
Wood et al. 2012 [22]	To analyse the perpetual differences in radiographic interpretation of skeletal fractures between experts and novices	UK	***n* = 30** 10x undergraduate radiography students (novices) 10x pre-Fellowship radiology trainees (intermediates) 10x post-Fellowship radiologists (experts)	Observation; Cross-Sectional	The most experienced group was most accurate in diagnosis, confident and the fastest of the participants. They had a faster determination of the fracture site and a greater fixation duration in that area, and this was most pronounced in detecting subtle fractures.
McLaughlin et al. 2017 [23]	To investigate the general image interpretation of general radiographs between different expert groups of radiographers	UK	***n* = 58** 21x radiography students 19x qualified radiographers 18x reporting radiographers	Observation; Cross-Sectional	Reporting radiographers were 15% more accurate than radiography students and radiographers, and also had a longer interpretation time and greater confidence level.
Gnanasekaran et al. 2022 [24]	To evaluate the gaze patterns of dental undergraduates when analysing panoramic radiographs	Australia	***n* = 65** 65x dental undergraduates (5th year)	Observation; Cross-Sectional	Most participants failed to conclude a correct diagnosis, and the search patterns of the participants did not demonstrate sequential interpretation of the panoramic radiographs.
Giovinco et al. 2015 [25]	To evaluate the differences between experienced and novice surgeons in the interpretation of pre-surgical hallux valgus plain radiographs.	USA	***n* = 16** 7x advanced surgeons (AS) 9x novice surgeons (NS)	Observation; Cross-Sectional	The AS group demonstrated that they moved their attention faster through the radiograph and spent less examination time determining clinical diagnosis. NS spent most of their time in searching behaviour. There was no significance found for accuracy.

Table 3. *Cont.*

Author and Date	Purpose	Setting	Participants	Research Style	Relevant Findings
Hanley et al. 2017 [26]	To quantitatively evaluate the differences between novice and expert orthopaedic trainees using eye-gaze metrics whilst analysing pelvic radiographs	USA	***n* = 23** 2x 4th year medical students 4x 1st year residents 4x 2nd year residents 3x 3rd year residents 6x 4th year residents 4x 5th year residents	Observation; Cross-Sectional	Whilst there was no relationship between identification of a fracture and experience, there was a relationship between the accurate identification of normal anatomy and expertise. Additionally, participants with more experience classified the fractures more effectively. Greater expertise correlated with a shorter interpretation time and fewer fixations.
Kelly et al. 2016 [27]	To evaluate the development of chest radiograph interpretation skill in medical training by comparing diagnostic accuracy and eye-gaze metrics.	Ireland	***n* = 21** 7x medical interns 5x senior house officers (i.e., 2nd year medical residents—not radiology) 4x radiology registrars 5x consultant radiologists	Observation; Cross-Sectional	There was a significant difference in accuracy between consultants and registrars. All the eye-gaze metrics and total reading time decreased with experience. Chest interpretation skill increased with experience.
Vogel and Schulze 2021 [28]	To evaluate the viewing patterns of dental students during different level of education in the analysis of panoramic radiographs	Germany	***n* = 48** 24x second clinical semester students (tested in both 1st [2a] and 2nd semester [2e]) 24x fifth clinical semester students (tested once [5a])	Observation; Cross-Sectional (Part Cohort)	More experience appears to correlate with an improvement in diagnostic capacity, and participants with greater expertise studied the radiograph more completely. The 2e cohort was the fastest viewing time, however the 5a cohort was more accurate. The time spent analysing the radiograph was not shown to correlate to diagnostic ability.
Bahaziq et al. 2019 [29]	To investigate differences between expert and novice orthodontists in the examination of panoramic radiographs via eye-gaze metrics	Saudi Arabia	***n* = 136** 72x novice orthodontists 64x expert orthodontists	Observation; Cross-Sectional.	No significance was found within the eye-gaze metrics. Expert orthodontists were found to spend a significantly longer time interpreting radiographs. There was no difference noted in the interpretation skills between participants.
Bertram et al. 2016 [30]	To investigate markers of expertise via visual markers in different levels of medical education of abdominal CT studies	Finland	***n* = 41** 15x early residents 14x advanced residents 12x specialists	Observation; Cross-Sectional	Specialists and advanced residents had longer fixation durations than early residents. Early residents detected a lower amount of low visual contrast lesions than specialists' counterparts.
Botelho et al. 2019 [31]	To differentiate the gaze patterns and identification ability between junior hospital dental officers and dental surgery assistants for radiographic (panoramic images) and non-radiographic images	Hong Kong	***n* = 18** 9x Junior Hospital Dental Officers (JHDO) 9x Dental surgery assistants (DSA)	Observation; Cross-Sectional	There were no significant differences pertaining to gaze metrics between the participants. The JHDOs had a higher percentage for area of interest identification and categorisation in the radiographic images.

Table 3. *Cont.*

Author and Date	Purpose	Setting	Participants	Research Style	Relevant Findings
Matsumoto et al. 2011 [32]	To investigate the neurologist search pattern of brain CT images and analyse the deployment of visual attention using eye-tracking saliency mop generation.	Japan	***n* = 30** 15x neurologists 15x controls (other medical professionals who do not have any education in interpreting brain CT)	Observation; Cross-Sectional	High salient areas were common fixations amongst both control and neurologist groups, however the neurologist groups also had high fixations on areas of low salience and high clinical importance.
Kelahan et al. 2019 [33]	To evaluate the scanning patterns of radiologists via eye-gaze tracking whilst analysing abdominopelvic CT	USA	***n* = 17** 9x attendings 8x trainees (radiologists)	Observation; Cross-Sectional	There were similarities concluded between trainees and attendings in most eye-gaze metrics. Attendings did have a lower fixation frequency, suggesting greater efficiency.
Hanna et al. 2018 [34]	To evaluate the impact of overnight shifts on fatigue, visual search and diagnostic performance of radiologists	USA	***n* = 12** 5x faculty radiologists 7x resident radiologists	Observation: Cross-Sectional	Overall statistics demonstrated that fatigued professionals interpreted at a slower rate with higher inaccuracies. Some eye-metrics were shown to increase in frequency during fatigued states.
			EDUCATIONAL TOOL EVALUATION		
McLaughlin et al. 2021 [13]	To evaluate an education tool for radiographic interpretation (for which eye-tracking assisted in programming) via eye-tracking data and performance	UK	***n* = 47** 12x reporting radiographers trained in chest image interpretation 35x reporting radiographers trained in MSK interpretation	Randomised Controlled Trial	The interventional group scored higher in diagnostic accuracy than the control group, with true positive diagnoses and true negative diagnoses increasing. False positive rates decreased for the interventional group. Interventional group was significantly more confident.
Quen et al. 2020 [14]	To evaluate whether the use of low-cost eye-tracking equipment provides adequate feedback for pedagogical development in the interpretation of chest x-rays	Malaysia	***n* = 8 (medical officers)** Split into two groups (Treatment and Control)	Non-Randomised Controlled Trial	There were no significant differences in accuracy, however the treatment group had faster decision speeds. Treatment groups also commented on the confidence boost. The tool was also rated qualitatively by the tutor and students as a positive tool for learning.
Kok et al. 2016 [15]	To investigate the relationship between systematic viewing, diagnostic accuracy and complete review (via education of participants) of an image in radiographic interpretation using eye-tracking	Nether-lands	**Experiment 1** ***n* = 30** 11x final year medical students 10x radiology residents 9x radiologists **Experiment 2** ***n* = 75** 75x 2nd year medical students	Non-Randomised Controlled Trial	**#1** The data suggests a lack of relationship between systematic viewing analysis techniques and coverage and diagnostic performance. Expert interpretation is more systematic that that of students ($p = 0.02$). **#2** There was a significant relationship demonstrated between coverage and systematic viewing techniques ($p < 0.01$), however this did not relate to specificity or sensitivity.

Table 3. *Cont.*

Author and Date	Purpose	Setting	Participants	Research Style	Relevant Findings
			VISUAL SEARCH BEHAVIOUR EVALUATION		
Ba et al. 2020 [35]	To investigate the scrolling techniques of radiologists whilst analysing liver CT imaging using eye-gaze tracking to establish understanding of scrolling behaviour	Switzerland	***n* = 20** 1x undergraduate medical student 16x radiology residents 2x fellows 1x experienced radiologist	Observation; Cross-Sectional	The use of eye-gaze metrics was an inferior analysis tool. Radiologists who performed with a greater number of courses covered more volume at a greater rate, found a greater number of metastases and also made fewer search errors.
Venjakbob et al. 2016 [36]	To investigate the interpretation differences by radiologists of different stack modes (small and large) of cranial computed tomography slices	Germany	***n* = 21** 21x radiologists	Observation; Cross-Sectional	Small stack CT mode is better for overview and motion perception, however large stack CT mode is better for detailed analyses. There was no overall difference in performance between the two stack modes.
Rubin et al. 2014 [37]	To evaluate the search patterns, recognition and detection of lung nodules in CT images by radiologists	USA/ Canada	***n* = 13** radiologists with varying level of experience	Observation: Cross Sectional	Radiologists seem to search less than half of the lung parenchyma although encompassing 75% of nodules in their search volume. Significant inter-reader variations exist in radiologists' search and detection capabilities of lung nodules. Synchronized recording of eye tracking offers insight into development of consistently effective screening method in detection of lung nodules.
			ASSISTIVE AID EVALUATION		
Kok et al. 2015 [16]	To evaluate the usage of comparison films (of different or same diseases) in the image interpretation accuracy, via the use of eye-gaze metric analysis	Netherlands	***n* = 84** 84x 3rd year medical students	Randomised Controlled Trial	The highest level of efficiency was found between same-disease and different-disease comparisons for improvement of accuracy ($p < 0.05$). Eye tracking tool provides insight into the comparison process when students interpreted radiographs.
Krupinski 2019 [38]	To evaluate whether the use of patient photographs aids in the interpretation accuracy of radiographic images to determine the correct placement of tubes and lines	USA	***n* = 6** 6x radiology residents	Observation; Cross-Sectional	The addition of patient photographs improves the radiographic detection of tube placement. Data has also been shown on the extra time spent on interpreting when photograph is added. Decision confidence was significantly increased with the addition of photographs.

Key Findings

The gathered literature presented four key themes, being the use of eye-tracking technologies in the following area: competency assessment of health professionals, the improvement of educational repertoire, the analysis of visual search behaviour and the impact of assistive tools.

Within the spectrum of competency assessment, there was a significant link between more experienced professionals and higher levels of diagnostic accuracy and ratings of personal confidence. In comparison, there were varying conclusions in regard to interpretation speed, compounded by variation in the literature quality (i.e., articles with disputed themes had significantly larger sample sizes, robust study designs and strong results). Similar disparities were found pertaining to eye-metric conclusions, with variations in trends and statistical significance.

The theme of educational tool evaluation was addressed by few articles (13%) and provided generalisations which suggest that the use of eye-tracking technologies may improve or assist in the education of young professionals.

The final two themes (visual search behaviour assessment and assistive tool evaluation) may have presented with articles of varying purpose and methodology, yet provided insight into radiologist search behaviour in CT imagery and promoted the inclusion of assistive tools in diagnostic and educational capacities, respectively.

Individual article analysis is provided within Tables 2 and 3, with respective inquiry into study properties.

4. Discussion

4.1. Overarching Themes

The literature gathered can be presented in four main themes. The first conclusion, which was supported by the greatest amount of literature, was the use of eye-tracking in a form of competency assessment, demonstrating the differing fixation patterns and interpretation capacity of different participant expertise. The other themes are less extensively studied, however they are essential to the radiographic narrative, and include the assessment of eye-tracking as both an education tool and the evaluation of education tools. Other themes centre around the assessment of visual search behaviour during computed tomography and judgement of assistive aid use during radiographic interpretation.

4.2. Competency Assessment Evaluation

The method of competency assessment diverges depending on the research's outcome measures, however the seventeen articles which chose such aforementioned focus generally considered speed, accuracy, confidence and the gaze-metric values to be of significance. Comparison via eye-tracking devices was commonly conducted between two or more participant groupings, usually with quantifiable differences in expertise and experience. The singular study which was in contrast to this trend was the article by Gnanasekaran et al. [24] which continued to provide insight into visual patterns and competency, hence inclusion in this review.

4.2.1. Speed

Nine out of seventeen articles (53%) present statistics referencing the overall examination time of radiographic images. The overarching suggestion indicated a strong relationship between greater experience and faster interpretation times, with some literature suggesting alternative statistics.

The main articles which supported a decreased interpretation time with increased experience were namely Giovinco et al. [25] Hanley et al. [26] Kelly et al. [27] Turgeon and Lam, [21] Vogel and Schulze, [28] and Wood et al. [22]. Whilst comparing different professional disciplines (e.g., "advanced and novice surgeons" versus "oral and maxillofacial radiologists and dental students") there was commonality that the participant group at a higher level of experience interpreted the radiographs at an overall faster rate. The

quality of all articles was adequate, with appropriate sample sizes, participant recruitment techniques and demographics. Additionally, all articles utilised some form of statistical analysis to provide significance. There are some issues presented with error, such as for Vogel and Schulze [28] who compared different university dental students (2nd and 5th semesters), however also "re-tested" the 2nd semester students after structured interpretation training to provide greater participant comparisons. This could pose limitations in giving irregular exposure to students at such an early stage in education and failing to provide the same opportunities to the 5th semester students. Variation in results can also be found in the article by Kelly et al. [27] in which significance was found across participant groups, however stronger time statistics were found between registrars and senior house officers, and between senior house officers and interns specifically. These six articles present adequate arguments for the relationship between higher experience level and a decreased interpretation speed.

However, two studies by Bahaziq et al. [29] and McLaughlin et al. [23] reported that participants with "greater expertise" interpreted radiographic images at a slower rate than compared to inexperienced counterparts. One of these articles, McLaughlin et al. [23] would be considered high quality with a significant study design and suggests that reporting radiographers spent longer interpreting than both students and radiographers (which is statistically significant). Interestingly, it was noted that radiographers (diagnostic) were faster than both students and reporting radiographers, therefore not aligning to a linear pattern of experience. Such statistics were also statistically significant. Bahaziq et al. [29] also conducted a study with sound quality, however, did not document data sets as thoroughly as McLaughlin et al. [23]. Bahaziq et al. suggested that expert observers had a longer examination time of the panoramic radiographs and did provide significance, alongside a considerable sample size of 136 participants, a strong study design and statistical analysis via the Chi-square test and the Fisher's exact test. These results may be considered important in final conclusions.

Gnanasekaran et al. [24] additionally commented on interpretation speed of undergraduate dental students analysing panoramic radiographs, however this does not provide a comparison to another experienced demographic. The average speed statistics were provided at 245.58 $\pm$ 106.7 s, however no further significant results were determined.

Whilst articles with alternative themes are important to consider when drawing overall conclusions (i.e., McLaughlin et al. [23] and Bahaziq et al. [29]), the volume and strength of the literature (i.e., [21,22,25–28]) provide adequate evidence to generalise that there is a relationship between a "greater experience level" and a shorter interpretation time for health professionals when interpreting radiographic images.

4.2.2. Accuracy

Eleven of the seventeen articles (65%) directly addressed accuracy or performance of radiographic interpretations. The majority posed suggestions of a positive relationship between participants with greater experience and a greater level of accuracy.

Nine articles [18,19,22,23,27,28,30–32] argued that there is a relationship between greater expertise and greater accuracy levels. The researchers expressed their suggestions differently, such as "detection rate increased with working hours ... [and that early residents] detected less of the low visual contrast lesions than did specialists" [30]. Many articles provided evidence in the form of statistical significance, and the consistent thematic conclusions across such a large volume of literature formulates adequate generalisations.

Two articles which provided an alternative to the main theme were reported by Giovinco et al. [25] and Hanley et al. [26]. Giovinco et al. [25] stated that there was no differentiation between advanced and early surgeons in accuracy over the whole radiograph data set, although there did appear to be a relationship between accuracy and experience in regard to moderate bunion radiographs, suggesting an advanced surgeon's acuity in subtle cases. This is similar to Hanley et al. [26] where significance was not found between experienced and novice orthopaedic trainees in the interpretation of fractured

pelvic bones, however, they did prove significance for the interpretation of non-fractured images. Additionally, Hanley et al. [26] demonstrated advanced use of fracture classification systems in the expert group which influences accuracy and diagnostic capability.

In addition to variations within articles, both Giovinco et al. [25] and Hanley et al. [26] provided smaller samples with minimal demographic variation. Whilst presenting with strong study design and statistically significant results, considering the inconclusive themes it would be prudent to conclude that there is a relationship between a greater expertise level and greater diagnostic accuracy.

4.2.3. Confidence

Confidence was not a focus many articles have undertaken in combination with eye-tracking analysis, yet three articles considered a relationship between professional confidence and level of expertise [19,22,23]. All three articles surmised that a greater level of experience related to a greater level of confidence in interpreting radiographic materials.

In regard to the study design, both Wood et al. [22] and McLaughlin et al. [23] used a scale from zero (not confident) to ten (very confident) as a measure after each radiograph diagnosis. Crowe et al. [19] used a similar method, with a four-point Likert scale including the options "Guess, Maybe, Probably and Definitely" to grade their own confidence in their diagnosis. Both of these studies used an oral recording to capture responses, whilst Crowe et al. [19] did not explicitly mention the method of collection. Whilst there may be differences in data collection, there appears to be no history of gathering methods influencing confidence results [39]. All articles utilised strong study designs, large samples, appropriate analysis and reflective conclusions.

A disadvantage to the comparative nature of these studies is that they all utilised different methods of statistical analysis. Crowe et al. [19] used a standard signal detection statistical method, whilst McLaughlin et al. [23] and Wood et al. [22] used Kruskal–Wallis and the proportional odds model, respectively. Whilst these are all adequate statistical analysis methods, it does pose challenges when comparing the significant figures of the articles.

Despite differences in statistical analysis and data collection methods, the literature points to a common theme which suggests a relationship between a higher level of education and participants' greater confidence level in radiographic interpretation.

4.2.4. Eye-Metrics

The majority of literature (17/25, 68%) referenced eye-gaze metrics in the analysis of interpretation competency. This gaze-monitoring technique is generally referenced as a manner to understand search pattern behaviour [1]. For example, conclusions may arise that a faster time to first fixation (within an area of interest) may indicate a participant has more experience or is more familiar with a radiographic presentation and can identify abnormalities quicker than inexperienced counterparts. The goal of many studies is to observe any relationships between experience factors and eye-gaze metrics.

There was no clear overarching positive or negative trend amongst the studies collected. Whilst individual articles present significant results, an adequate synthesis is not possible due to variable findings across these studies. For example, the studies by Bertram et al. [30], Brams et al. [18], Turgeon and Lam [21] and McLaughlin et al. [23] provided significant evidence that an increase in experience; increases fixation duration, increases distances between fixations, provides greater fixation count and mean visit count. These studies are extremely strong in methodology, participant numbers and variance in demographics and analysis of significance. Other studies provided support in the form of generalised trends, such as Bahaziq et al. [29] which reflect similar statements.

Studies by Wood et al. [22], Kelahan et al. [33], Hanley et al. [26] and Lèvêque et al. [20] provided results which reflected relationships between expert counterparts and a lower fixation frequency, shorter mean fixation durations, shorter times to first fixation and fewer fixations.

Additionally, there were also studies which found a lack of significance between all participant groups and eye-gaze metrics, and some articles which found significance only between particular sub-categories of participants [27]. Conflicting information to an apparent main theme was also found within singular articles, such as in Hanley et al. [26] which suggested that there was no correlation of fixation duration and experience (in comparison to significance comparing greater experience to a lower fixation count), and Kelahan et al. [33] which suggested that whilst time to first fixation statistics were shorter in attendings that produced accurate diagnoses (in comparison to trainees), the time to first fixation statistics were insignificant when considering overall results.

There are four articles which did not explicitly utilise statistical significance in regard to eye-gaze metric analysis provide no additional persuasion or strong trends to change interpretation of the synthesis [19,25,28,32]. The inconsistent findings from the current literature make it difficult to discern a main theme. Further analysis of the literature with additional information available could be been conducted to provide sufficient base for argumentation.

4.2.5. Fatigue and Competency

Another article that analysed radiologist competency by Hanna et al. [34] focused on the impact of fatigue on accuracy, speed and eye-metric data. Results of their study suggest that a fatigued state decreases overall diagnostic accuracy and speed, and increased time to first fixation (on the area of interest) and number of fixations. There were statistical differences between radiology residents and fellows, however this study utilised fatigue as the independent variable rather than degree of expertise. Whilst this article provides interesting observation into the influence of professional fatigue and radiology interpretation, generalisations are limited due to small sample sizes (n = 12) and lack of comparison with other articles. This study provided a thorough methodology and significant results which suggest opportunities for further research and development.

4.3. Educational Tool Evaluation

The use of eye-tracking in an educational role is a theme which was addressed in the literature. It must be noted that whilst synthesis may be difficult considering differing purpose and methodology, all articles provide important discourse to radiological discussion.

The first article pertaining to training for radiographic interpretation via eye-tracking was that of McLaughlin et al. [13] in which eye-tracking was used as a tool to both inform production and assessment of the program. The program appeared to significantly improve diagnostic accuracy of the intervention group (with an increase in true positive and decrease in false positive diagnoses) as well as a significant increase in confidence. Gaze metrics were analysed, yet were more varied in results, as there was a decrease in fixation count, visit duration and interpretation time for the control group, the chest x-ray (CXR) reporting radiographers presented with a decrease in fixation duration and fixation count for both groups, but the intervention CXR reporting radiographer group made diagnostic decisions in significantly more time. Whilst these results may reflect themes from the competency assessment section of this review (Section 4.2), one must consider that a similar research team also produced the article by McLaughlin et al. [23] who posed similar questions about speed and gaze-metrics. This is important to consider in the overall review for sources of bias in result reporting. In other aspects, both articles were presented well with strong methodologies, samples, analysis, discussions and conclusions.

A similar study by Quen et al. [14] presented eye-tracking as a form of self-reflection and assessed performance post-intervention. Whilst there was no significance in regard to improved accuracy for the intervention group, there was an identification that interventional trainees shortened decision time [14]. There was also a subjective rating of the self-study system by the trainees and a senior radiologist mentor, who cohesively rated the program positively as it allowed for review into educational improvement [14]. This study had many limitations which impact quality, such as the lack of significance analysis,

very small sample size (eight trainees only, which introduced error) and a lower-cost, lower efficiency system which can affect the feedback quality. Despite these study restraints, there is a benefit to the utilisation of such systems for educational purposes.

The article by Kok et al. [15] focused on the manner in which interpretation is taught to professionals who interpret radiographs. Authors used interventions based on systematic training, non-systematic training or full coverage training. The results reflected that that systematic groups presented the most homogenous viewing technique, non-systematic and systematic training was best for sensitivity and that non-systematic training was the fastest for abnormality detection. The overall conclusion stated that students did not benefit from systematic training. Whilst this study conducted dual experiments, main focus remained on the second experiment which referenced educational styles. There was an adequate sample size of 64, with appropriate procedures and equipment for data collection. Limitations resided in the length of training, as longer training may provide further benefits in interpretation strategy.

Overall thematic conclusions for these studies suggest that eye-tracking technologies may have a considerable role in the education of radiological professionals and the improvement of diagnostic skills.

4.4. Visual Search Behaviour Evaluation

Three articles focus on the visual search behaviour of radiologists' reviewing of CT imagery. Ba et al. [35] provided context for the analysis of scrolling behaviour of radiologists, and compared eye-tracking data to other methods of interpretation measurement. On an alternative discussion, Venjakob et al. [36] analysed the search behaviour differences between radiologists when reviewing CT images of different sizes, and the eye-metric data which originated from such. Lastly, Rubin et al. [37] reviewed radiologists' fixations and review of lung parenchyma and nodule detection.

Ba et al. [35] suggests that the singular use of eye-gaze metrics is not appropriate for considering radiologist search behaviour when compared to the number of courses (defined as the "plotting of the image slices [in the z direction] versus time for each trial and each reader"). They suggest that radiologists with a higher course number covered greater CT volume in a faster time, were more detailed in metastases detection and also had reduced search error in comparison to those with lower course numbers. The study continues to discuss how a combination of eye-metric data and "course number" analysis is important to classify radiologists under titles relevant to search behaviour styles.

In the review of CT image sizes, Venjakob et al. [36] suggests that smaller images (i.e., size of 14 × 14 cm) produced a reduced number of fixations, yet the fixation duration for this size was longer. Additionally, these fixations were across a greater number of slices. Conclusions suggest that smaller images may provide better overview and motion perception, whilst larger images aid in detailed analysis. It was shown that there was no difference in performance between these two sizes.

A separate article by Rubin et al. [37] focuses on the search recognition and fixation patterns of nodule detection by radiologists' review of lung parenchyma on CT imagery. Not focusing on the scrolling behaviour, authors studied gaze behaviour and volume and the correlation of nodule sensitivity by radiologists. They found that nodule sensitivity throughout entire volume sets ranges from 0.3–0.73, however, there did appear to be a slight correlation between increased CT volume search and sensitivity. Interestingly, the authors noted that on average, 26.7% of the lung parenchyma was searched of the CT volume. Limitations of this study were broad, including a limited sample size and demographic, lack of nodule variation in simulated CT volumes (including nodule size), generalised error of equipment and gaze radius restrictions. This study is also minimally comparable to other studies assessing competency due to differences in methodology and outcome measures.

It is difficult to make an appropriate synthesis of these three articles due to the variation in discussion material and focus questions, yet all provide insight into the radiological visual search behaviour of CT imagery. It is important to consider for future projects,

however no definite conclusions can be drawn from the literature by Ba et al. [35], Venjakob et al. [36] and Rubin et al. [37].

4.5. Assistive Aid Evaluation

Another minor theme presented is the use of eye-tracking in the assessment of assistive tools. Two articles by Kok et al. [16] and Krupinski [38] demonstrate the use of eye-tracking in this new area.

Krupinski [38] analysed the influence of patient photographs on the accuracy of radiographic interpretation concerning the placement of tubes or lines (i.e., central lines, orogastric, nasogastric or endotracheal tubes) through analysis of radiographic dataset of 37 images. There was an adequate variation of tubes and lines presented to the participants, with 23 central lines, 1 orogastric, 12 endotracheal tubes and 17 nasogastric tubes, all reviewed for gold standard by an abdominal and cardiothoracic radiologist with high levels of experience. The analysis of the data was conducted via either ANOVA or paired *t*-tests depending on the number of variables compared. However, this study suffered from several limitations, such as the small sample size of six radiology residents. There was no demographic information of participant experience, nor an outlined sampling method. Other limitations included the lack of literature in regard to patient photograph placements for clinical optimisation, the lack of "no true gold standard" for tube placement analysis and the lack of generalisability. In regard to results, there were no statistically significant results for accuracy measurements. Trends were present which suggest that the use of patient photographs improved accuracy in the detection of nasogastric and orogastric tubes and central lines, however such results could not be significantly included. There was, however, a significant increase in the confidence of diagnoses.

The article by Kok et al. [16] has a similar concept in which comparison films (with either the same disease, different disease or normal films) have an impact on the analysis of a radiographic image. This article presented a large sample size of 84 medical students with demographic information and randomisation into participant groups. There was also a large data set of radiographic images available, improving statistical power. The results found that the most efficient study method was between same disease and different disease counterparts, whilst the control group spent 30% more time in analysis. It is also noted there were no significant differences marked between the different image pairs. Limitations of the study by Kok et al. [16] focused on the observational nature of the experiment without instruction on how to utilise comparison films—therefore reducing the effect of the performance being dependent on the comparison method. This decision, however, allowed for the participant's natural technique of observation to be explored. There are issues with generalisability due to the specific population of inexperienced medical students, whereas application to population may require greater awareness of medical students at different levels of education.

Overall, these studies advocate for the inclusion of such assistive tools in a diagnostic environment to improve recognition and detection. There were differences in the methodology and purpose of these articles, which makes generalisations difficult, and some conclusions may improve with statistical significance (i.e., Krupinski [38]). This research field in still within relative infancy, and further discussion and investigation may provide insight into further radiological opportunities.

5. Limitations

There were some limitations to this review, especially with regard to the narrow research field and few articles available for analysis (assumed due to the required development of new technologies). Additionally, the articles varied significantly in methodology which made literature comparison difficult. Differences in study design, sample sizes, populations, data collection methods, data analysis and other factors across studies made quality and data comparison limited, as demonstrated by broad assessment factors in Kmet et al. [12]. Thematic limitations for Sections 4.3–4.5 exist due to the lack of investigation in

such fields, which restricts the development of robust conclusions. Further research and technological advancement may improve the conclusions generated.

Whilst there was an effort to diversify literature collection, we recognize that the inclusion of additional databases and investigation into non-English studies may provide more insight. We are aware of the influence of bias in this study, and have attempted to reduce such wherever viable.

6. Conclusions

The purpose of this review was to understand the manner in which eye-tracking has been used for the analysis of radiological practice. Four main investigation pathways were identified, namely competency assessment, education tool evaluation, analysis of visual search behaviours and the impact of assistive tools.

The literature primarily focused on competency assessment as the most common research study in which speed, accuracy, confidence and eye-metrics were individually discussed. Whilst there was variation in some outcome measures, initial conclusions can be assumed due to the link between high levels of expertise and accuracy and confidence rates. Other tentative conclusions may also be drawn between shorter examination times and greater expertise, however further research is recommended. The analysis tool of eye-gaze metrics poses new dimensions of competency measurement, and may be a great method of assessment for future study designs.

Other studies focusing on educational and assistive aids provided promising results for both the positive influence on education and interpretation performance, respectively. Generalisation is limited due to the lack of literature, however further development of new literature and applied examples may provide significant impacts in the realm of radiology.

Similarly, the analysis of visual search behaviour via eye-tracking is still within infancy. Further literature may provide insight into the methodology of interpretation, therefore facilitating understanding of the radiological space.

This literature review provides a snapshot of the current climate of eye-tracking in radiology. The opportunities for the development of evaluation, educational, research and diagnostic spaces in radiology may be aided and advanced by eye-tracking technologies into the future.

Author Contributions: Conceptualization, E.A. and Z.S.; Data analysis: E.A. and Z.S.; Writing—original draft preparation: E.A.; Writing—review and editing: E.A. and Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Tom Lee, Welber Marinovic, Melissa Black and An Nguyen for their support in this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hammoud, R.I.; Mulligan, J.B. Introduction to eye monitoring. In *Passive Eye Monitoring: Algorithms, Applications and Experiments*; Hammoud, R.I., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–19.
2. Wu, C.-C.; Wolfe, J.M. Eye movements in medical image perception: A selective review of past, present and future. *Vision* **2019**, *3*, 32. [CrossRef]
3. Harezlak, K.; Kasprowski, P. Application of eye tracking in medicine: A survey, research issues and challenges. *Comput. Med. Imaging Graph.* **2018**, *65*, 176–190. [CrossRef] [PubMed]
4. Davies, A.; Brown, G.; Vigo, M.; Harper, S.; Horseman, L.; Splendiani, B.; Hill, E.; Jay, C. Exploring the relationship between eye movements and electrocardiogram interpretation accuracy. *Sci. Rep.* **2016**, *6*, 38227. [CrossRef] [PubMed]

5. Bruno, M.; Walker, E.A.; Abujudeg, H.H. Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction. *Radiographics* **2015**, *35*, 1668–1676. [CrossRef]
6. Bushby, L.P.; Courtier, J.L.; Glastonbury, C.M. Bias in radiology: The how and why od misses and misinterpretations. *Radiographics* **2018**, *38*, 236–247. [CrossRef]
7. Itri, J.N.; Tappouni, R.R.; McEachern, R.O.; Pesch, A.J.; Patel, S.H. Fundamentals of diagnostic error in imaging. *Radiographics* **2018**, *38*, 1846–1865. [CrossRef] [PubMed]
8. Brunyè, T.T.; Drew, T.; Weaver, D.L.; Elmore, J.G. A review of eye tracking for understanding and improving diagnostic interpretation. *Cogn. Res. Princ. Implic.* **2019**, *4*, 1–16. [CrossRef] [PubMed]
9. Ganesan, A.; Alakhras, M.; Brennan, P.C.; Mello-Thomas, C. A review of factors influencing radiologists' visual search behaviour. *J. Med. Imaging Radiat. Oncol.* **2018**, *62*, 747–757. [CrossRef] [PubMed]
10. Moher, D.; Shamseer, L.; Clarke, M.; Ghersi, D.; Liberati, A.; Petticrew, M.; Shekelle, P.; Stewart, L.A.; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **2015**, *4*, 1–9. [CrossRef] [PubMed]
11. Stewart, L.A.; Clarke, A.; Rovers, M.; Riley, R.D.; Simmonds, M.; Stewart, G.; Tierney, J.F. Preferred reporting items for a systematic review and meta-analysis of individual participant data: The PRISMA-IPD Statement. *JAMA* **2015**, *313*, 1657–1665. [CrossRef]
12. Kmet, L.M.; Lee, R.C.; Cook, L.S. *Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields*; Alberta Heritage Foundation for Medical Research: Edmonton, AB, Canada, 2004; Available online: https://era.library.ualberta.ca/items/48b9b989-c221-4df6-9e35-af782082280e/view/a1cffdde-243e-41c3-be98-885f6d4dcb29/standard_quality_assessment_criteria_for_evaluating_primary_research_papers_from_a_variety_of_fields.pdf (accessed on 22 April 2022).
13. McLaughlin, L.; Hughes, C.M.; Bond, R.; McConnell, J.; McFadden, S.L. The effect of a digital training tool to aid chest image interpretation: Hybridising eye tracking technology and a decision support tool. *Radiography* **2021**, *27*, 505–511. [CrossRef]
14. Quen, M.T.Z.; Mountstephens, J.; The, Y.G.; Teo, J. Medical image interpretation training with a low-cost eye tracking and feedback system: A preliminary study. *Healthc. Technol. Lett.* **2021**, *8*, 97–103. [CrossRef]
15. Kok, E.M.; Jarodzka, H.; de Bruin, A.B.H.; BinAmir, H.A.N.; Robben, S.G.F.; van Merriënboer, J.J.G. Systematic viewing in radiology: Seeing more, missing less? *Adv. Health Sci. Educ. Theory Pract.* **2016**, *21*, 189–205. [CrossRef] [PubMed]
16. Kok, E.M.; de Bruin, A.B.H.; Leppink, J.; van Merriënboer, J.J.G.; Robben, S.G.F. Case comparisons: An efficient way of learning radiology. *Acad. Radiol.* **2015**, *22*, 1226–1235. [CrossRef] [PubMed]
17. STROBE. STROBE Checklists. Available online: https://www.strobe-statement.org/checklists/ (accessed on 1 May 2022).
18. Brams, S.; Ziv, G.; Hooge, I.T.C.; Levin, O.; De Brouwere, T.; Verschakelen, J.; Dauwe, S.; Williams, A.M.; Wagemans, J.; Helsen, W.F. Focal lung pathology detection in radiology: Is there an effect of experience on visual search behavior? *Atten. Percept. Pyschophys.* **2020**, *82*, 2837–2850. [CrossRef]
19. Crowe, E.M.; Gilchrist, I.D.; Kent, C. New approaches to the analysis of eye movement behaviour across expertise while viewing brain MRIs. *Cogn. Res. Princ. Implic.* **2018**, *3*, 1–14. [CrossRef]
20. Lèvêque, L.; Berg, B.V.; Bosmans, H.; Cockmartin, L.; Keupers, M.; Ongeval, C.V.; Liu, H. A statistical evaluation of eye-tracking data of screening mammography: Effects of expertise and experience on image reading. *Signal. Process. Image Commun.* **2019**, *78*, 86–93. [CrossRef]
21. Turgeon, D.P.; Lam, E.W.N. Influence of experience and training on dental students' examination performance regarding panoramic images. *J. Dent. Educ.* **2016**, *80*, 156–164. [CrossRef]
22. Wood, G.; Knapp, K.M.; Rock, B.; Cousens, C.; Roobottom, C.; Wilson, M.R. Visual expertise in detecting and diagnosing skeletal fractures. *Skelet. Radiol.* **2013**, *42*, 165–172. [CrossRef]
23. McLaughlin, L.; Bond, R.; Hughes, C.; McConnell, J.; McFadden, S. Computing eye gaze metrics for the automatic assessment of radiographer performance during X-ray image interpretation. *Int. J. Med. Inform.* **2017**, *105*, 11–21. [CrossRef] [PubMed]
24. Gnanasekaran, F.P.; Nirmal, L.; Sujitha, P.; Bhayyaa, R.; Muthu, M.S.; Cho, V.Y.; King, N.M.; Anthonappa, R.P. Visual interpretation of panoramic radiographs in dental students using eye-tracking technology. *J. Dent. Educ.* **2022**, *86*, 887–892. [CrossRef] [PubMed]
25. Giovinco, N.A.; Sutton, S.M.; Miller, J.D.; Rankin, T.M.; Gonzalez, G.W.; Najafi, B.; Armstrong, D.G. A passing glance? Differences in eye tracking and gaze patterns between trainees and experts reading plain film bunion radiographs. *J. Foot Ankle Surg.* **2015**, *54*, 382–391. [CrossRef] [PubMed]
26. Hanley, J.; Warren, D.; Glass, N.; Tranel, D.; Karam, M.; Buckwalter, J. Visual interpretation of plain radiographs in orthopaedics using eye-tracking technology. *Iowa Orthop. J.* **2017**, *37*, 225–231. [PubMed]
27. Kelly, B.S.; Rainford, L.A.; Darcy, S.P.; Kavanagh, E.C.; Toomey, R.J. The development of expertise in radiology: In chest radiograph interpretation, "Expert" Search Pattern May Predate "Expert" levels of diagnostic accuracy for pneumothorax identification. *Radiology* **2016**, *280*, 252–260. [CrossRef]
28. Vogel, D.; Schulze, R. Viewing patterns regarding panoramic radiographs with different pathological lesions: An eye-tracking study. *Dentomaxillofac. Radiol.* **2021**, *50*, 20210019. [CrossRef] [PubMed]
29. Bahaziq, A.; Jadu, F.M.; Jan, A.M.; Baghdady, M.; Feteih, R.M. A comparative study of the examination pattern of panoramic radiographs using eye-tracking software. *J. Contemp. Dent. Pract.* **2019**, *20*, 1436–1441. [PubMed]
30. Bertram, R.; Kaakinen, J.; Bensch, F.; Helle, L.; Lantto, E.; Niemi, P.; Lundbom, N. Eye movements of radiologists reflect expertise in CT study interpretation: A potential tool to measure resident development. *Radiology* **2016**, *281*, 805–815. [CrossRef] [PubMed]

31. Botelho, M.G.; Ekambaram, M.; Bhuyan, S.Y.; Kan Yeung, A.W.; Tanaka, R.; Bornstein, M.M.; Kar, Y.L. A comparison of visual identification of dental radiographic and nonradiographic images using eye tracking technology. *Clin. Exp. Dent. Res.* **2020**, *6*, 59–69. [CrossRef]
32. Matsumoto, H.; Terao, Y.; Yugeta, A.; Fukuda, H.; Emoto, M.; Furubayashi, T.; Okano, T.; Hanajima, R.; Ugawa, Y. Where do neurologists look when viewing brain CT images? an eye-tracking study involving stroke cases. *PLoS ONE* **2011**, *6*, e28928.
33. Kelahan, L.C.; Fong, A.; Blumenthal, J.; Kandaswamy, S.; Ratwani, R.M.; Filice, R.W. The radiologist's gaze: Mapping three-dimensional visual search in computed tomography of the abdomen and pelvis. *J. Digit. Imaging* **2018**, *32*, 234–240. [CrossRef]
34. Hanna, T.N.; Zygmont, M.E.; Peterson, R.; Theriot, D.; Shekhani, H.; Johnson, J.-O.; Krupinski, E.A. The Effects of Fatigue from Overnight Shifts on Radiology Search Patterns and Diagnostic Performance. *J. Am. Coll. Radiol.* **2018**, *15*, 1709–1716. [CrossRef]
35. Ba, A.; Shams, M.; Schmidt, S.; Eckstein, M.P.; Verdun, F.R.; Bochud, F.O. Search of low-contrast liver lesions in abdominal CT: The importance of scrolling behavior. *J. Med. Imaging* **2020**, *7*, 045501. [CrossRef] [PubMed]
36. Venjakob, A.C.; Marnitz, T.; Phillips, P.; Mello-Thoms, C.R. Image size influences visual search and perception of hemorrhages when reading cranial CT: An eye-tracking study. *Hum. Factors* **2016**, *58*, 441–451. [CrossRef]
37. Rubin, G.D.; Roos, J.E.; Tall, M.; Harrawood, B.; Bag, S.; Ly, D.L.; Seaman, D.M.; Hurwitz, L.M.; Napel, S.; Choudhury, K.R. Characterizing Search, Recognition, and Decision in the Detection of Lung Nodules on CT Scans: Elucidation with Eye Tracking. *Radiology* **2014**, *271*, 276–286. [CrossRef] [PubMed]
38. Krupinski, E.A. Impact of patient photos on detection accuracy, decision confidence and eye-tracking parameters in chest and abdomen images with tubes and lines. *J. Digit. Imaging* **2019**, *32*, 827–831. [CrossRef] [PubMed]
39. Tekin, E.; Roediger, H.L. III. The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cogn. Res. Princ. Implic.* **2017**, *2*, 1–13.