**qXR: an Evaluation and Recommendation**

**Prepared By**

Project Manager and UX Team Member: Cailin Feldman

UX Team Leader: Bronte Lee

Technical Team Leader: Kiran Papalkar

UX Team Member: Jingbang Men

Technical Team Member: Zichen Li

Technical Team Member: Zheyuan Zhang

Technical Team Member: Baoxi Liu

**Prepared For**

ABC Health Services Company (ABC)

Simon Poon

Neysa Petrina

October 2022

# Executive Summary

**Problem Articulation**

Tuberculosis (TB) approximately infects 10 million people a year and kills 1.5 million. If TB is diagnosed and treated early, most people are cured within six months of diagnosis. The WHO End TB Strategy aims to reduce TB deaths by 90% and to cut new cases by 80% between 2015 and 2030. To achieve this goal, missing cases must be identified and all cases diagnosed earlier. This can be achieved through more widespread and early screening of people at greater epidemiological risk than the general population (who are assumed to be disease-free). These early screenings can be done by computer-aided detection (CAD) commercial products which diagnose TB using artificial intelligence algorithms.

At the present moment, there is a large range of available certified CAD products that can be used for TB detection which can help make professional recommendations. At the present moment, due to the nature of the faced-past medical technology field, there is an abundance of technologies available that can detect TB, However, this wide variety of options causes clients, the ABC company to feel overwhelmed as they are unsure how to decipher between the products and pick the highest-performing one to purchase.

In addition, our clients are aware of the importance a positive user experience can have on the workflow when integrating new technology. Therefore another feeling of unease occurs amongst our ABC company clients as they do not know what are the most significant factors to focus on to ensure a positive integration experience.

**Overview**

The report involves an analysis of the various available certified AI technologies that are able to detect TB. In our analysis (as suggested by our client) our team used the CAD product, qXR produced by Qure.ai as the benchmark. Based on our findings, a final recommendation is reached and presented in the report.

In addition, to ensure the clients have a positive user experience, our team presents their findings and highlights the factors that they have found to be the most impactful. Therefore informing our clients that these aspects should be paid the most attention to when integrating new technology into the workflow.

**Technical Findings**

After performing a meta-analysis to compare and benchmark qXR against commercial AIs used for medical imaging in TB diagnoses, our team found that:

* That each product that was analysed has an accuracy equal to or greater than that of a radiologist.
* CAD4TB and qXR have similar performance and higher accuracy scores compared to the other AI products included in our study. Our team found only CAD4TB and qXR reached the minimum performance requirements of the WHO triage TPP (90% sensitivity and 70% specificity)

**UX Findings**

A systematic view was undertaken to collect articles of user experience and acceptance of people who work in radiology using AI products. The data was synthesised using thematic analysis, with the UTAUT model as the conceptual framework defining initial themes and codes. The themes, performance expectancy, effort expectancy, social influence, facilitating conditions, and personal attitude were found to explain the data through the coding process of data extracted from nine included studies. Overall, the majority of people expected that AI will be a useful tool, providing benefits such as workflow optimisation and improved patient care, however, many were also concerned about their job security, had a lack of trust in the AI, and were concerned about potential risks. When assessing the interaction between these themes, it was found that facilitating conditions directly impacted on the participant’s acceptance of AI in the workflow. In contrast, the other themes were interconnected, with social influence and personal attitude impacting on performance expectancy and effort expectancy. We took a deeper look at these and considered how someone would develop a positive personal attitude towards AI and become someone who socially influences others to adopt AI. To answer this, we propose that knowledge is the key factor for these two themes which is supported by various examples in the data where those who had more knowledge about AI were less concerned about their job security and risks, and generally less fearful of AI and had a more positive attitude towards it.

**Recommendation**

After considering the needs of the ABC Company and conducting a systematic review as well as a meta-analysis and thematic analysis, there are several recommendations our team has reached. Due to the high performance and suite of supporting products including qTrack, we recommend our client chooses the qXR product. In addition, our team suggests that the ABC should put a heavy amount of emphasis on the knowledge of the users to help improve the user’s acceptance towards the new technology. However, due to there not being any information about qXR’s user experience, we suggest our clients do a pilot of the qXR first. In this period our clients can identify the efficiency and effectiveness of the practical qXR performance and how the radiologists think and how they accept using qXR during their diagnosis. Therefore, ABC Company can use the results of the pilot to make a final decision on whether qXR is suitable for their company and whether to continue to invest in it.

**Table of Contents**

[Executive Summary](#_c2xkcg2dat9d)

[1 Introduction](#_a10batp60ygn)

[1.1 Problem Statement](#_sdqhvp8rcl95)

[1.2 Purpose](#_i6uy82h7d38h)

[1.3 Objectives](#_htfd59i1w956)

[1.4 Stakeholders](#_royjb4pqyifj)

[2 Relevance](#_6zi3jg3odgcd)

[2.1 What problem is the application intended to solve, and who is the application designed for?](#_vqn6m0hq7aj)

[2.2 What are the potential benefits, and for whom?](#_mrypuxh9hq3)

[2.3 What are the risks associated with the use of the AI system?](#_p1mm2hcxbggk)

[3 Performance and Validation](#_bnopjqv4a0ui)

[3.1 Are the algorithm’s design specifications clear?](#_ydckl0jetufe)

[3.2 How was the algorithm trained?](#_xk5607o36x8v)

[3.3 How has performance been evaluated?](#_5vnwwsnnc4ig)

[3.3.1 Abstract](#_zd5xmkc5bdf1)

[3.3.2 Method](#_v717ce56hmr9)

[3.3.3 Systematic Review](#_uavxmyajwsw6)

[3.3.4 Heterogeneity](#_ydyv68x8q6e0)

[3.3.5 Results](#_ws0zvlz1q3v4)

[3.3.6 Discussion](#_khtruixjvyjn)

[3.3.7 Conclusion and Limitations](#_j0nya65l1gnm)

[3.4 Have the developers identified and accounted for potential sources of bias in their algorithm?](#_4h32iyqfn13k)

[3.5 Is the algorithm fixed or adapting as new data comes in?](#_87gt5b8ojta0)

[4 Usability and Integration](#_hqerxdf34d5u)

[4.1 How can the product be integrated into your clinical workflow?](#_xng9gdqqtj69)

[4.2 How exactly does the product impact the workflow?](#_d6ue6fdugvfe)

[4.2.1 Abstract](#_eypom9buhafq)

[4.2.2 Introduction](#_k5ly0nnqmuut)

[4.2.3 Methods](#_mdhrx2sfmwwa)

[4.2.4 Systematic review of qXR](#_jq52zc4t6lxd)

[4.2.5 Systematic review of AI in radiology](#_54eqrf6awqbr)

[4.2.6 Study characteristics and risk of bias](#_1zkvexpblwnb)

[4.2.7 Results](#_zapcmo7hu3b)

[4.2.8 Discussion](#_1aora8llli2k)

[4.2.9 Conclusion](#_4ahup560devw)

[4.3 What are the requirements in terms of IT infrastructure?](#_aazsam1t5atj)

[4.4 Interoperability - How can the data be exported for research and other purposes?](#_5woi8v41g6s8)

[4.5 Will the data be accessible to non-radiologists?](#_yx25n24xr5mz)

[4.6 Are the AI model’s results interpretable?](#_r3wg0yqvto4g)

[5 Regulatory and Legal Aspects](#_1vg0msjeix86)

[5.1 Does the AI application comply with the local medical device regulations?](#_lj76shr1k7d4)

[5.1.1 Australian medical device regulation](#_l4hhdf88xt3j)

[5.1.2 Process to comply with local medical device regulations](#_tyrxirwfzcro)

[5.1.3 Applying the process to qXR](#_9u7bon4mn0hy)

[5.2 Does the AI application comply with the data protection regulations?](#_59m15i60rkxc)

[5.2.1 Application of Australian data protection regulation to ABC and its use of qXR](#_gjjx96eph584)

[5.2.2 qXR and the Australian Privacy Principles](#_uh88qrna2vf1)

[5.2.3 Action Required by ABC](#_3fz83lqzc56s)

[6 Financial and Support Services Considerations](#_nuq5era2xpj6)

[6.1 What is the licensing model?](#_em0hu54n594n)

[6.2 How are user training and follow-up handled?](#_syytgy6gd4gi)

[6.3 How is the maintenance of the product ensured?](#_5cv0fmxc13wa)

[6.4 How will potential malfunctions or erroneous results be handled?](#_c1mn568fgtrt)

[7 Conclusion](#_7dm58c6yljlu)

[7.1 Summary of Evaluation](#_mskrla5fquyt)

[7.1.1 Relevance](#_d2anqiyw52pp)

[7.1.2 Performance and Validation](#_lp8o557e9ta8)

[7.1.3 Usability and Integration](#_l15cgxi210tq)

[7.1.4 Regulatory and Legal Aspects](#_shjxmcn7xabg)

[7.1.5 Financial and Support Considerations](#_br98nmk6wmlc)

[7.2 Recommendation](#_lz7vgnsxcoln)

[8 Appendix](#_mvmmfoi6e7fe)

[8.1 Technical Systematic Review Protocol](#_e2e7gt4nw26e)

[8.2 Technical Articles](#_wocngpvg7wdj)

[8.3 Forest Plots for DTA](#_yiz4tlm02v3t)

[8.4 Technical Summary Characteristics](#_g3lz85hvz747)

[8.5 qXR Systematic Review Protocol](#_iuh2zs0h3wj)

[8.6 AI in Radiology Systematic Review Protocol](#_rbjmsmlzdguc)

[8.7 UX Codebook](#_oouyu389i3a6)

[8.8 UX Systematic Review Articles](#_lwgjrkgcojhw)

[8.9 UX Study Characteristics](#_nt12dgxoyomy)

[8.10 UX Data](#_qquig7vd88yv)

[9 References](#_m3euvf6oclqw)

# 1 Introduction

## 1.1 Problem Statement

At the present moment, there is a large range of certified CAD products that may be used for TB detection and to make professional recommendations. However, an issue arises: what product is most fit for purpose? Ideally, you want a product that is the most accurate at detection and offers the best user experience.

However, due to the amount of research and the difficulty of keeping up with the fast-paced medical technology field, finding all the needed information for analysis is difficult.

## 1.2 Purpose

The purpose of this evaluation is to benchmark across different certified AI products for detecting TB compared to qXR. We’ll focus on the performance of qXR and compare all the other certified AI products’ performance with qXR to draw the conclusion. We’ve divided into 2 teams which are technical team and ux team separately. The technical team will perform DTA evaluation to look for any evidence related to qXR and also find any comparison study between different certified AI products. The ux team will look for any user study related to qXR and probably also find other similar AI products competing with qXR.

## 1.3 Objectives

The primary objective of this evaluation is a successful benchmarking of certified AI products in TB against qXR, this will be done through a systematic review and meta-analysis to discern the relative diagnostic test accuracies. Further, this evaluation will discern the usability of qXR through a systematic review and thematic analysis, this will ensure that qXR can be adopted by the client and will outline the success the product will have in being integrated into workflow. This report will be used by the client to assess whether qXR is the ideal AI product for further investment and use in their business.

## 1.4 Stakeholders

There are many stakeholders of this evaluation. They are

* Information System researchers which is specially our client Simon Poon
* ABC Company
* Medical health care professionals
* Program administrator
* Software and technical managers
* Regulators
* Patients

Our stakeholders can use this evaluation to have a brief knowledge of different AI products for radiology, by reading the analysis provided in this evaluation, they will understand the strength and weakness of each product, know what are the differences between them, and make a decision or a suggestion based on the evaluation. The evaluation is based on the data researched and analysed through systematic review and meta analysis, this evaluation can be an academic basis for the stakeholders.

# 2 Relevance

## 2.1 What problem is the application intended to solve, and who is the application designed for?

​​The purpose of qXR is to automate the x-ray interpretation process in order to bridge the gap between the incidence of tuberculosis (TB) and the actual number of reported cases. With this application patients can be diagnosed at a much faster rate and can thus be put into treatment sooner with the process taking minutes rather than days, this can have a significant impact in mitigating the negative effects of TB and prevent further transmission within the community (*QXR- Artificial Intelligence Aided Tuberculosis Detection*, 2019). qXR scans for 30 different findings in the chest area and can locate abnormalities which are then processed into a report, further to this, qXR informs qTrack in order to track patients and monitor their results (*Qure.ai*, 2021). The end users of qXR are those in the healthcare field including physicians and x-ray technicians who will be able to efficiently diagnose and subsequently treat TB, further to this, as patients will be majorly impacted by choice of AI we consider them as users as well.

## 2.2 What are the potential benefits, and for whom?

The benefits for using AI algorithms in medical imaging include many aspects and will benefit different people, including radiologists and patients.

Some benefits of using qXR are (Qure.ai, 2021):

* **Time-Efficient:** qXR can read chest x-rays in less than one minute using deep learning algorithms which saves a lot of time. Traditionally, reading chest x-rays requires a radiologist’s eye and experience which will be time-cost.
* **Workflow optimisation:** qXR can cluster the x-rays into normal, abnormal, and to be reviewed, and prioritise cases that need immediate attention. This benefits patients because they can be correctly prioritised so the most serious are seen first, and overall, patient care is improved. Additionally, for radiologists, their workflow is streamlined and managed for them.
* **Automate Interpretation:** qXR offers an automated interpretation of chest x-rays which means radiologists don’t need to manually interpret the results from chest x-rays.
* **Reduce Misdiagnoses:** Since qXR only needs less 1 minute to read and triage chest x-rays, it will potentially reduce the chance of late diagnosis, under diagnosis and even misdiagnosis to improve the quality of patient care.
* **Identify multiple abnormalities:** in addition to TB, qXR has been trained to also detect 29 other abnormalities which are visible on chest x-rays. This will benefit patients because qXR will review the chest x-ray for many abnormalities, whereas a radiologist may miss something because they are only looking for what they suspect the patient’s illness could be.

Another benefit to using the qXR product from the company, Qure.ai is the suite of products that is offered from them that can enhance the radiological practice workflow. qTrack is an automated management platform that has been developed to specifically provide healthcare professionals with a singular, holistic data gathering and reporting system, that maintains and manages all patient information in one place. This platform makes the patient’s data collection, transmission and recall a seamless process between the administrative managers, healthcare professionals, lab technicians and patients. Healthcare professionals can collect all of the patient’s data through the qTrack app on their smartphone which then constructs a comprehensive patient profile that is automatically populated with data and integrations of the AI interpretations of chest X-rays, as well as details like results of tests that confirm the diagnosis. qTrack saves precious time which is otherwise lost locating and harmonising data manually.



This image shows how the combination of using qTrack and qXR would look in the workplace.

There are some general benefits of using AI in medical imaging which are not limited to qXR. It can increase efficiency and efficacy of the workflow, and reduce the workload of radiologists, particularly in terms of disease detection, characterisation, and monitoring (Hosny et al., 2018). Additionally, using AI to generate radiology reports can support the field to develop standardised terminology and structure for these reports to improve collaboration between medical professionals, support population sciences, and contribute to higher quality datasets for big data mining (Hosny et al., 2018).

Erroneous patient positioning is one of the biggest reasons for poor image quality, increased radiation dose and repeated examination (Malamateniou, 2021). However, using an automated AI system performed much better than radiologists in accurately positioning the patients. It will also potentially improve image quality and reduce patient dose. In addition, one common benefit for different AI applications in medical imaging is that AI algorithms promote the standardisation of practice and make the diagnosis processes more efficient (Malamateniou, 2021). By using automated AI systems in medical imaging, it will also change the way for diagnosing patients and improve the trust from users.

## 2.3 What are the risks associated with the use of the AI system?

No risks specific to qXR were identified through researching qXR, therefore, general risks that can apply to all AI in radiology were considered.

A major risk that can affect patient care is ‘automation bias’, where radiologists become reliant on the AI to make clinical decisions and neglect their own judgement. This raises legal liability issues regarding the standard of care radiologists owe their patients (Nair et al., 2022). For instance, if a radiologist simply relies on the AI for diagnosis, however, it is incorrect. Is it the radiologist’s or the AI provider’s fault that the patient was misdiagnosed? It remains unclear how this situation will be treated ethically and legally.

AI is reliant on data, however, data ownership and privacy laws risk limiting the data available for training and testing. A lack of comprehensive and high quality data can restrict AI-research and bias datasets and the AI algorithms (Nair et al., 2022). Data laws are important to protect patient’s’ rights, however, these laws differ between countries, and in some cases, within different states of the same country, which can make it difficult to collect the data required to ensure that the AI is highly accurate.

Weak cybersecurity risks cyberattacks, such as stealing health data or altering it to train the AI to become inaccurate (Nair et al., 2022). There are a range of access points in a workflow that uses AI which can be infiltrated by hackers, including scanner, PACS, server, and workstations (Chu et al., 2020). A scan of the World Wide Web identified 2,774 radiology or DICOM servers that are unprotected and highly vulnerable to attack (Chu et al., 2020). The radiology field lags behind in data security, however, this can be mitigated by implementing strong, multi-layered security controls (Nair et al., 2022).

# 3 Performance and Validation

## 3.1 Are the algorithm’s design specifications clear?

The algorithm used by qXR is clear, the development team used a deep learning approach to address usage with clinical tests. The algorithm first decomposes the task into different subtasks then uses an automatic segmentation model consisting of a deep neural network (DNN). The DNN segments the training samples (MRI images), outputs them, and learns the underlying features (Randhawa, 2016).

## 3.2 How was the algorithm trained?

The qXR development team used 24,384 CXRs for training, and the training samples underwent a 60-20-20 split pattern for train, validation and test sets. The development team trained an 18-layer deep residual convolutional neural network and predicted whether the chest x-ray was normal or abnormal and further tested for the presence of different indicators of illness such as cardiomegaly, opacity and pleural effusion which are further specific cases of abnormalities in CXR. In order to visualise and interpret the trained models of CXR that the algorithm was used on, prediction difference analysis was used, this is a method whereby each pixel is replaced with a pixel from a “normal” CXR and it is noted by what degree the prediction changed (Putha, 2017).

The final results on the test set show an area under the ROC curve of 0.89 for abnormal, 0.92 for detecting cardiomegaly, 0.84 for detecting opacity, and 0.91 for detecting pleural effusion (Putha, 2017).

## 3.3 How has performance been evaluated?

### 3.3.1 Abstract

**Background**

To analyse qXR’s DTA in comparison to other certified AI products, a systematic review and meta-analysis of DTA studies was conducted. The purpose was to benchmark qXR against these other certified products and determine which product had the superior DTA.

**Methods**

A systematic review was performed to identify all relevant articles, these were gathered by running search queries to find articles in which qXR was directly compared to each product. Data was then extracted and visualisations made to identify the accuracy of each product through meta-analysis.

**Results**

It was found that qXR and CAD4TB performed significantly better than the rest of the products, an SROC graph did indicate some outliers however it was still clear.

**Discussion**

We believe qXR is the most sound recommendation after completing our analysis. It was found to perform slightly better than CAD4TB and ABC already has interest in pursuing it as a viable alternative to human readers.

### 3.3.2 Method

We performed a systematic review to compare different AI products with qXR. First, we constructed a protocol which follows the PRISMA-P template to inform the ways in which our systematic review will be carried out, this process included defining the search queries and defining exclusion criteria for identified articles. The complete protocol is located in [Appendix 8.1](#_e2e7gt4nw26e). Following this we used the search queries in various databases and journals to find all the relevant comparison articles with qXR. Finally, we extracted the necessary data from each article as informed by our PRISMA writings and began the analysis. Our analysis involved creating forest plots of sensitivity and specificity for each article and creating visualisations of the data for more unique comparison.

### 

### 3.3.3 Systematic Review

In September of 2022 a systematic review was conducted across four databases and three journals, the queries used identified articles which discussed qXR and at least one other AI product in order to allow for benchmarking. Through this method we identified a total of 26 articles, upon screening for duplicates this was reduced to 18, a screening by name and abstract identified a further three unsuitable articles and our final screening of reading the full entry excluded a further nine reports, the major reason for exclusion in this area was a lack of appropriate data that was required for meta-analysis. Upon completion of the systematic review we were left with six articles that we would use in the completion of our meta-analysis. The complete list of articles is in Appendix 8.2. Further to this the systematic review necessitated the removal of five of our original certified products, it was found that products RADIFY, JLD-02K and AXIR did not have any relevant analysis because they are relatively new in comparison to the other CAD product. VUNO was also excluded due to the only analysis being biassed towards their particular product as the product owner had conducted the analysis. Finally, Chesteye only had one relevant article with information on DTA and therefore did not fit our needs for conducting meta-analysis. This left us with three products for benchmarking against qXR, that is, InferRead DR, CAD4TB and INSIGHT CXR.

### 3.3.4 Heterogeneity

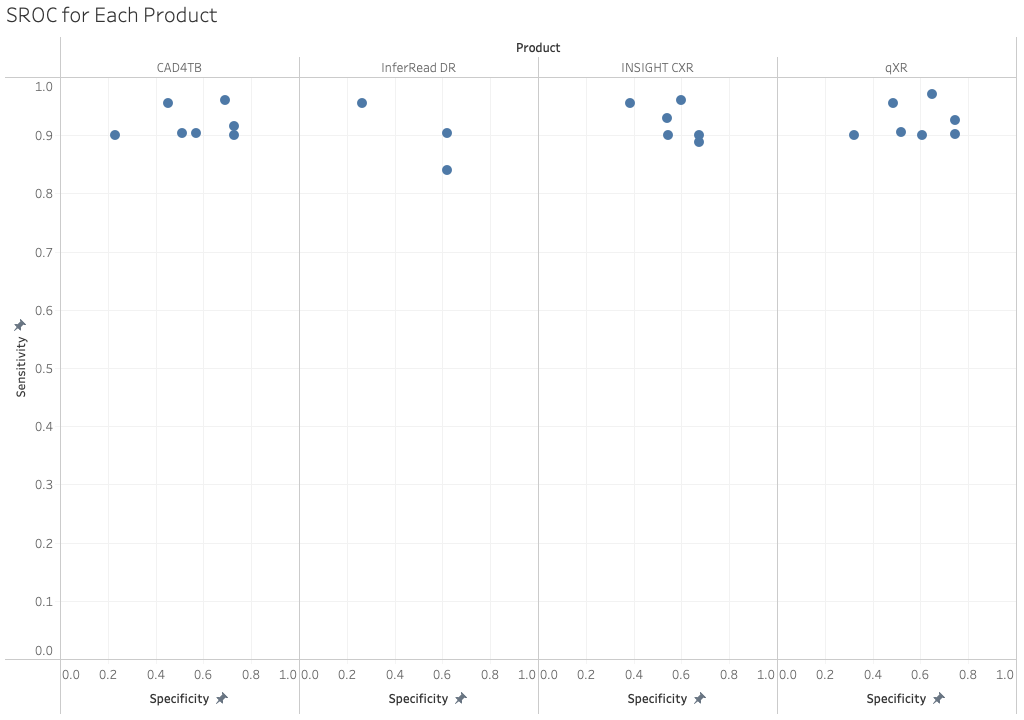
The next stage of our meta-analysis involved testing for heterogeneity of the data, in order to do this we created coupled forest plots for each product listed in each retained article. This allowed us to compare the results of these products and make a judgement on whether the results across studies were homogenous or heterogenous following the guidelines in Diagnostic test accuracy methods for systematic review and meta-analysis (Campbell et al., 2015). Although this article suggested using an SROC curve to inform judgements on heterogeneity it was found this could not be achieved with the data that was extracted from the selected articles, thus, we compiled the forest plots and visually concluded that the gathered results were consistent (homogenous) and therefore accepted the results of the studies. The forest plots that were generated can be found in [Appendix 8.3](#_yiz4tlm02v3t).

### 3.3.5 Results

This meta-analysis utilised seven articles and an overview of each of these articles can be found below:

| **Authors** | **Title** | **Description** |
| --- | --- | --- |
| Zhi Zhen Qin, MSc  Shahriar Ahmed, MHE  Mohammad Shahnewaz Sarker, BSc  Kishor Paul, MPH Ahammad Shafiq Sikder Adel, MPH  Tasneem Naheyan, BS  Rachael Barrett, BA Sayera Banu, PhD  Jacob Creswell, PhD | Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms | This article evaluates 5 commercial AI algorithms(qXR, CAD4TB, LUNIT, InferRead DR and JF CXR) for triaging TB using a large dataset that had not been used before. Every participant was verbally screened for symptoms and received a digital posterior-anterior chest x-ray and an Xpert MTB/RIF (Xpert) test. It compared the performance based on the WHO’s TPP principle. |
| Andrew J. Codlin  Thang Phuoc Dao  Luan Nguyen Quang Vo1  Rachel J. Forse  Vinh Van Truong  Ha Minh Dang  Lan Huu Nguyen  Hoa Binh Nguyen  NhungViet Nguyen  Kristi Sidney‑Annerstedt  Bertie Squire  Knut Lönnroth  Maxine Caws | Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis | This article developed a test library of chest X-ray (CXR) images which was blindly re-read by two TB clinicians with different levels of experience and then processed by 12 CAD software solutions. It uses cut-off thresholds which are selected to match the sensitivity of each human reader. It evaluates 5 certified AI products including qXR, CAD4TB, LUNIT, InferRead DR and ChestEye. |
| Zhi Zhen Qin  Melissa S. Sander  Bishwa Rai  Collins N. Titahong  Santat Sudrungrot  Sylvain N. Laah  Lal Mani Adhikari  E. Jane Carter  Lekha Puri  Andrew J. Codlin  Jacob Creswell | Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems | This article conducts a retrospective evaluation of three commercial AI systems for detecting TB in chest radiographs from patients in Nepal and Cameroon. 1196 individuals were screened by two sets of radiologists and each system and sensitivity, specificity and accuracy (as defined by AUC) were extracted |
| Mary Kagujje  Andrew D Kerkhoff  Mutinta Nteeni  Ian Dunn  Kondwelani Mateyo  Monde Muyoyeta | The Performance of Computer-Aided Detection Digital Chest X-ray Reading Technologies for Triage of Active Tuberculosis Among Persons With a History of Previous Tuberculosis | This article evaluates both CAD4TB and qXR among 1884 participants, 24% of which had a prior history of TB. This article found that CAD systems performance is limited within individuals with a prior treatment of TB. |
| Gamuchirai Tavaziva  Miriam Harris  Syed K Abidi  Coralie Geric  Marianne Breuninger  Keertan Dheda  Aliasgar Esmail  Monde Muyoyeta  Klaus Reither  Arman Majidulla  Aamir J Khan  Jonathon R Campbell  Pierre-Marie David  Claudia Denkinger  Cecily Miller  Ruvandhi Nathavitharana  Madhukar Pai  Andrea Benedettia  Faiz Ahmad Khan | Chest X-ray Analysis With Deep Learning-Based Software as a Triage Test for Pulmonary Tuberculosis: An Individual Patient Data Meta-Analysis of Diagnostic Accuracy | This article evaluates 3 CAD systems, CAD4TB, LUNIT and qXR, 3727 samples were taken, this article used CXR and individual patient data for study. 17% of samples were microbiologically confirmed with tuberculosis, and the result showed the accuracy was similar to human readers. |
| Sifrash Meseret Gelaw  Sandra V. Kik  Morten Ruhwald  Stefano Ongarello  Tesfa Semagne Egzertegegne  Olga Gorbacheva  Christopher Gilpin  Nina Marano  Scott Lee  Christina R. Phares  Victoria Medina  Bhaskar Amatya  Claudia M. Denkinger | Diagnostic accuracy of three computer-aided detection systems for detecting pulmonary tuberculosis on chest radiography when used for screening: analysis of an international, multicenter migrants screening study | This article evaluates three CAD systems, CAD4TB v6, Lunit INSIGHT v4.9.0, and qXR v2. The findings showed that all three systems had acceptable diagnostic accuracy for detecting TB on CXR when used for TB screening and performed similarly to radiologists. However, none of the CAD systems reached the minimum performance requirements of the WHO triage TPP (90% sensitivity and 70% specificity). |

Data was extracted from each of these articles and placed into a summary table of study characteristics including the first author, year published, geographic area of interest and participants involved in the study. We then found the AUC, sensitivity and specificity of each product analysed in each article that were plotted in the same table. This table can be found in [Appendix 8.4](#_g3lz85hvz747).



An SROC plot was also created in order to better illustrate the variability of sensitivity and specificity for each product across differing studies. As each identified study used a sensitivity threshold the main analysis is of the product’s specificity at each threshold. Due to this, clear standout performances by qXR and CAD4TB can be observed, with Lunit’s INSIGHT CXR remaining fairly competitive.

### 3.3.6 Discussion

Taking into account all relevant data qXR and CAD4TB appear to be similar with regard to DTA, although qXR was found to have slightly better performance in the majority of identified articles. Some identified articles had inconsistent results when compared in our analysis, for example, we found that the Gelaw results had LUNIT’s INSIGHT CXR perform particularly well above both qXR and CAD4TB although in other research the opposite was true. The most recent article, from this year, also showed lower specificity scores for both qXR and CAD4TB, this was explained as a large proportion of the participants (24%) had previously been diagnosed and treated for TB, an identified flaw of the algorithms is that they are less effective in diagnosing patients with a history of TB, and as more years pass and TB is more easily identifiable and treatable this issue of correct diagnosis will grow.

### 3.3.7 Conclusion and Limitations

In summation, our final recommendation for the product that should be pursued for implementation by the ABC based on DTA is qXR, we believe this product is most suited to the needs of ABC and will have the most positive contribution to their diagnosis work and treatment of patients. An identified limitation of our analysis is the diversity of versions of each product that was tested, as our results had to be gathered from many different years, nearly every product was a different iteration which led to some inconsistencies within the data. A further limitation lies in the time in which we conducted the systematic review and meta-analysis , as this was a shortened period of time as compared to others in the research sector we were limited on the extent to which we could gather and analyse data, in future projects we would suggest more time be allocated so that the final product can be more well-rounded and better researched.

## 3.4 Have the developers identified and accounted for potential sources of bias in their algorithm?

The developers have clearly indicated potential biases that exist within their algorithm in a blog post published by Qure Ai (Chilamkurthy, 2018), the primary source of alleviating bias was in their selection of datasets with which to validate the product. qXR avoided selection bias in this area by using a combination of CT scans and an algorithmically selected dataset. Another major bias that would be identified by the qXR developers is the underdiagnosis bias (Seyyed-Kalantari et al., 2021) which can be especially prevalent with patients who come from underserved populations, this is an issue as it can lead to patients who are ill not receiving the necessary care, qXR has not stated how this bias was accounted for in the development of the algorithm.

## 3.5 Is the algorithm fixed or adapting as new data comes in?

The clear adaptive nature of qXR was most recently observed in 2020, during the height of the COVID-19 pandemic Qure.ai was able to repurpose qXR to aid in the diagnosis of COVID-19 (Tarun et al., 2020). This process involved tuning the algorithm with 200 positive COVID results alongside 200 negative results and leveraging pre-existing abnormality identification to be in line with COVID abnormalities including opacities and consolidation. Further to this, qXR is continuously trained with further analysis of CXR, between 2020 and 2022, qXR trained using an additional one million CXRs (Carmichael et al., 2022). Through this evidence it is correct to say that qXR is capable of adaptation as new data is attained.

# 4 Usability and Integration

## 4.1 How can the product be integrated into your clinical workflow?

A starting point when integrating CAD technology is ensuring that the radiology practice is compliant with clinical guidelines and standards. This includes making sure that their software is always up to date. This medical tool is used for patient diagnosis and treatment therefore it is important to ensure that it meets all the required standards. Some of these standards include ISO27001 usability standards.

Once the initial compliance checks are made, our clients will then need to evaluate several alternative options available when integrating qXR into their practice. qXR is designed to be used as an AI diagnostic tool to be used in conjunction with other tools for TB detection. Therefore it is important to consider how this A.I. will fit into the current workflow and how it will be used by the practice’s clinicians.

Once the decision for the integration option has been decided the client then needs to consider practical management of the software. This includes decisions on:

I. How to keep the software up to date if hosted locally?

II. How to ensure that only authorised users have access to it?

III. How to manage user permissions?

Once implemented and decisions on managing the technology are completed, training the radiological practices' staff is the final step. The training will involve showing the scanners how to use the qXR software to improve the efficiency and accuracy of their work. It is to be noted that qXRs implementation and training is available from qure. AI. A large amount of time should be dedicated to improving the user’s knowledge as this improves the user’s attitudes towards qXR (this finding is further expanded on in section [4.2.7](#_zapcmo7hu3b)) . Training should not only include instructions on using the qXR system but also should inform users of the importance and accuracy of qXR.

Integration Options

qXR can be integrated into the clinical workflow of a radiology practice in the following ways:

1. API based: Dicom images can be uploaded to the qXR REST APIs through a token-based authentication. Results can be downloaded using a separate API endpoint. While the API documents are simple and easy to use, we still recommend training to ensure successful implementation.
2. PACS-based: Another option is to allow PACS integration based on DIMSE protocol to transfer raw scans and receive outputs. A gateway can be added as a Dicom node in the Radiology practice network which will receive images, anonymize them, upload them using the APIs, download results and send these results back to the PACS

The APIs provide a simple and intuitive method of accessing qXR and integrating the software into the existing radiology practice software workflow. However, finer-grained control exists when using the DIMSE protocol. This may be suited for a large radiology practice that has well-resourced in-house software engineering development capabilities. The advantage of having the raw scans means you can manipulate the images for example, using AI to upscale the size of the DICOM.

Alternative Simple Solution

A simple alternative is provided by Integration via AI marketplace or distribution platform. Companies such as Nuance, Incepto, Philips IntelliSpace, Sectra Amplifier Store, Blackford, GE Healthcare, and Siemens provide direct integration using the existing workflow.

Using an AI marketplace is our recommendation as it encourages greater use through familiarity and minimal technical requirements on the part of the Radiology practice.

The use of an AI marketplace vendor removes this complexity as the integration is managed by the distributor. The radiology practice benefits from a standard integration approach. A choice exists to use product manufacturer integrations such as Philips,Toshiba, Siemens Healthineers, GE healthcare or a vendor-neutral gateway such as Sectra or Blackford. The advantage of the latter is since they work with multiple vendors and do many integrations to connect multiple different algorithms they may work well in radiology practices that have not adopted single vendor strategies. Sanjay Parkh (2019) states that the AI marketplace integration requires ‘little or no support’ and ‘offering end-to-end solutions’ that solve the engineering issues (deployment, workflow integration, etc.) and commercial issues (contracting, billing, etc.) that arise when implementing AI.

## 4.2 How exactly does the product impact the workflow?

### 4.2.1 Abstract

Background

To investigate and evaluate qXR’s impact on the radiological clinic’s workflow, our team conducted a systematic review and thematic analysis. Our team examined users’ experience when AIs are implemented into the diagnostic radiology workflow. The aim of our research was to see what level of impact on the workflow CAD products have. In addition used their findings to highlight areas that our clients should focus on to help with ensuring a positive user experience when implementing the technology.

Methods

A systematic review protocol was designed using PRISMA-P. Four databases and a journal were searched to find studies around user experience and user acceptance specific to qXR. However, no suitable studies were identified so the search for user experience and acceptance studies was broadened to AI products used in radiology. Therefore, our group made the executive assumption that the acceptance factors and impact on workflow identified in this broader search are generic and therefore can be applied to qXR. After completing the PRISMA process, all the studies to be included in the thematic analysis were identified. The thematic analysis was primarily deductive, using the Unified Theory of Acceptance and Use of Technology model as the conceptual framework. Once the analysis was completed, our team then used the results to identify factors that contribute the most impact in the clinical workflow when implementing a CAD product to promote a positive integration experience.

Results

The themes our team decided on, based on the UTAUT model was, performance expectancy, effort expectancy, social influence, and facilitating conditions. However, during the analysis process, our team included a new theme, personal attitude, as several of our results (e.g. data concerned with user’s trust) were better suited to this new included theme. After the thematic analysis was complete, our team found that a majority of users thought that AI in radiology would provide a wide variety of benefits such as workflow optimisation, improved patient care, and improved effectiveness. However, our findings show that there is a high amount of concern about the lack of knowledge, support and trust towards AI.

Discussion

The lack of user experience and acceptance studies indicates a major research and evaluation gap for qXR.

### 

### 4.2.2 Introduction

Qure.ai’s product qXR uses artificial intelligence to interpret chest x-rays and is an example of the emerging use of AI in healthcare. There are many important considerations when evaluating whether such a product is suitable, for instance, its safety, legality, and accuracy. However, another important consideration is its impact on the existing workflow of radiologists. The previous ECLAIR guidelines question considered the workflow of radiologists in terms of how to integrate qXR into it. Whereas this question is about how integrating qXR impacts on the workflow of radiologists. It is answered by examining the impact on the workflow from the perspective of radiologists, using user experience and acceptance studies. The studies were chosen using a systematic review following the PRISMA process and synthesised using thematic analysis. The systematic review and thematic analysis was initially designed to be specific to qXR, however, no suitable studies were found, which meant the scope was broadened to AI in radiology. It is a research gap for qXR, however, the impact identified more broadly is assumed to likely be similar to qXR’s impact.

### 

### 4.2.3 Methods

Systematic Review Process

Two systematic protocols were developed using PRISMA-P, the first specific to qXR and the second broadened to AI in radiology. Hence, the protocols are very similar, and both are included in the Appendix, [8.5](#_iuh2zs0h3wj) and [8.6](#_rbjmsmlzdguc). The eligibility criteria was divided into seven items that the article must include. In brief these were journal articles from 2020 for qXR and 2018 for AI in radiology, the article includes primary data from a usability-related study undertaken in a standard clinical setting from the perspective of people who work or will work in the field of radiology, and the sample size is at least 120 for surveys and 20 for interviews and focus groups. The databases PubMed, ProQuest, Embase, and ScienceDirect were searched, in addition to the journal The Lancet and five articles were provided by the client. A range of search queries were defined with keywords including user experience, user acceptance, user satisfaction, and usability, alongside either qXR or AI products. Filters were used to restrict the results to journal articles published from 2018 onwards which are the first two sections of the eligibility criteria. The PRISMA process was followed by collecting the search results, excluding by title and abstract, and then excluding by full-text reading, with all remaining studies included for analysis.

Data Synthesis

To synthesise the data in the included studies, thematic analysis was chosen. It was primarily deductive, with the Unified Theory of Acceptance and Use of Technology used as the conceptual framework. The constructs from the model, performance expectancy, effort expectancy, social influence, and facilitating conditions, were used as themes, and through research and brainstorming, sub-themes were developed and defined as codes in a codebook which is included in the [Appendix 8.7](#_oouyu389i3a6). For analysts to become familiar with the data, the data from the articles were extracted. The articles’ study characteristics were also extracted and reviewed to consider the quality of the studies and the chance of bias.

The included studies were divided into two types, the first were those which contained the raw data of the survey questions and answers or direct quotes and the second were ones that evaluated hypotheses from the questionnaire results and the data available was the hypothesis testing. For the first type, each individual question was coded separately by all group members and then these were compared and discussed to reach a consensus. For the second type, the relationship between the elements of the hypothesis were coded. However, in many situations an element was too broad for a specific code and was labelled with a theme instead. Robust group discussion about the data and the codes and themes defined in the codebook resulted in the creation of an additional theme called personal attitude. The results section contains a description of each theme, their codes, and the data relating to those codes. The discussion section discusses the relationship between the themes and their impact on the workflow.

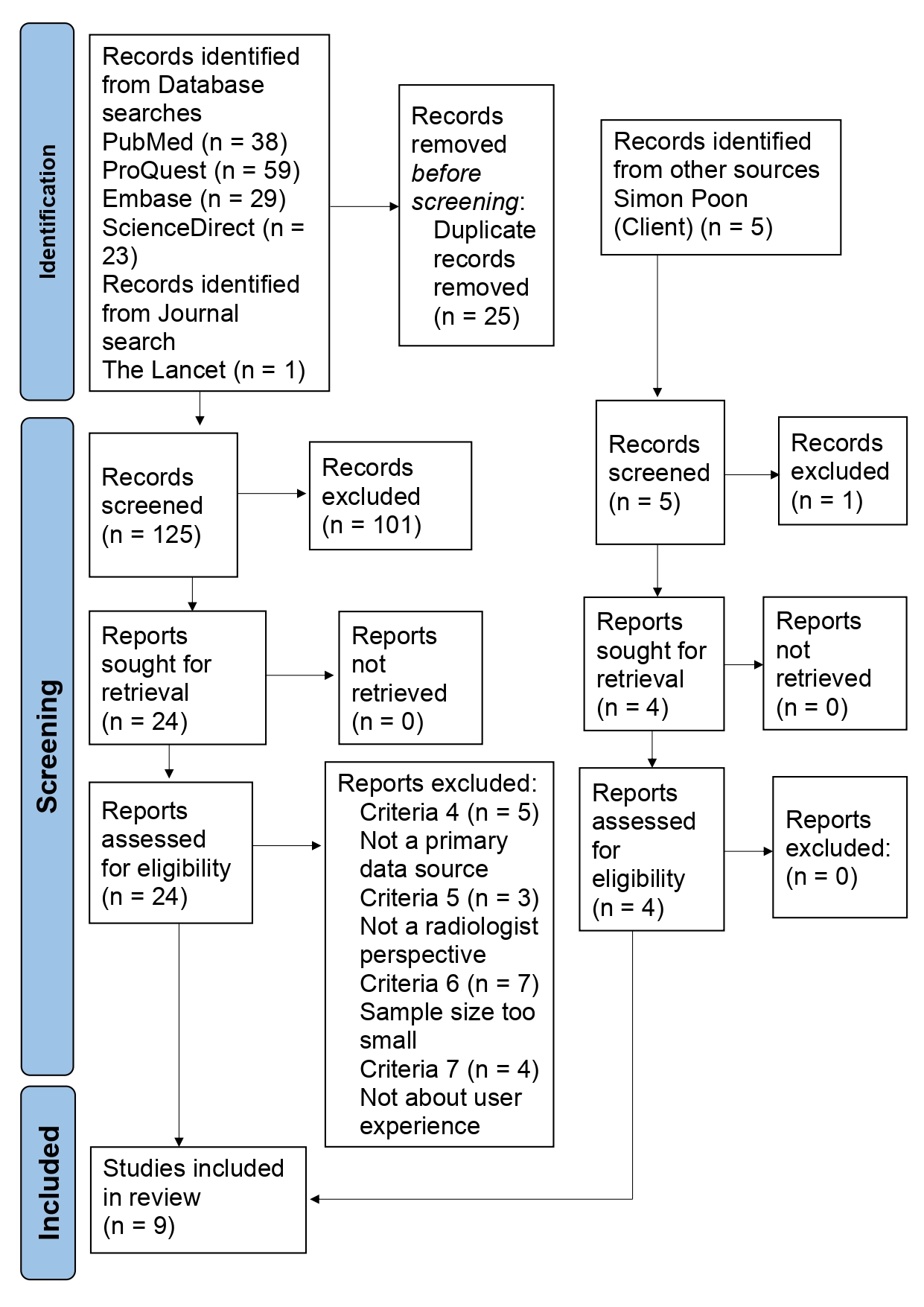
### 

### 4.2.4 Systematic review of qXR

The search was undertaken in September 2022. A total of 22 articles were identified and 18 remained after duplicates were removed.

The titles and abstracts of the articles were screened and 17 articles were excluded. The main criteria for inclusion was that the article referred to user experience of qXR in a standard clinical setting and was a primary data source. The majority of articles were excluded because they were studies of diagnostic test accuracy or reviews. The remaining article was excluded after a full-text read through of the article because it contained information about the use and implementation of qXR, however, did not contain a study on user experience. Complete details of all articles identified and their reason for exclusion are available in [Appendix 8.8](#_lwgjrkgcojhw). The lack of user experience studies on qXR posed a major limitation on the evaluation of qXR because its impact on the workflow cannot be directly assessed.

### 4.2.5 Systematic review of AI in radiology

Following the discovery of the lack of user experience studies of qXR, the purpose of the systematic review was expanded to user experience studies of AI in radiology. The same databases and journal were searched with amended search queries. This search was undertaken in September/October 2022 and identified 150 articles, 25 of which were duplicates. An additional five articles were provided by Simon Poon, one of our clients. From the two identification methods, 102 were excluded for not meeting the eligibility criteria when screening the titles and abstracts. The most common reason was that the article was not about user experience or acceptance. Many of these excluded articles were about performance, with only a brief mention of user experience or acceptance. The 28 remaining articles were assessed based on their full-text and 9 were suitable to be included for analysis. The most common reasons for exclusion were not having primary data or having a small sample size that was too small. Complete details of all articles identified are available in [Appendix 8.8](#_lwgjrkgcojhw).

### 4.2.6 Study characteristics and risk of bias

The nine included studies were published between 2018 and 2022 and took place in Africa, Portugal, China, India, and the Netherlands, two surveys were international, and one did not state where it took place. There is a risk that the perspectives of people who work in radiology vary across different countries due to cultural and local issues which is a limitation because no Australian studies were identified. However, combining the results from multiple countries and international studies would likely develop an overall world view of user experience and acceptance of AI in radiology that can broadly be applied to Australia. Five of the studies were surveys and four were focus groups or interviews. The surveys have a much larger sample size and sought broad opinions about AI in radiology, whereas the focus groups and interviews also required the participants to complete a task before providing responses. The complete study statistics and demographic information is included in the [Appendix 8.9](#_nt12dgxoyomy). Risks of bias and their mitigations in individual studies are discussed below.

| **Author, Year** | **Title** | **Risk of Bias/Limitation** | **Mitigation** |
| --- | --- | --- | --- |
| Botwe, et al. 2021 | The integration of AI in medical imaging practice | The survey was written in English, however, many people in Africa also speak other languages which may create an unrepresentative result because it only includes English speakers. | The survey was translated into French and Arabic. |
| Calisto, et al. 2022 | Modeling adoption of intelligent agents in medical imaging | Small sample size for specific nationalities. | Analysed the moderator effect of country development categories in the model. |
| Calisto, et al. 2021 | Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification | Not mentioned | Not mentioned |
| Fan, et al. 2018 | Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS) | Limited to China, did not consider moderators of the relationships, and a lower number of residents than what was desired. | Limitations to be addressed in future research. |
| Goel, et al. 2022 | The effect of machine learning explanations on user trust for automated diagnosis of COVID-19 | Small sample size. | Limitation to be addressed in future research. |
| Huisman, et al | An international survey on AI in radiology in 1,041 radiologists and radiology residents part 1: fear of replacement, knowledge, and attitude | Selection bias because there was only a 3.9% response rate. | Assume that people who did the survey are more interested in AI, therefore, in reality the open attitude is likely lower and fear is higher. |
| Huisman, et al | An international survey on AI in radiology in 1041 radiologists and radiology residents part 2: expectations, hurdles to implementation, and education | Selection bias because there was only a 3.9% response rate. | Assume that people who did the survey are more interested in AI, therefore, in reality the open attitude is likely lower and fear is higher. |
| Praskash, Das | Medical practitioner’s adoption of intelligent clinical diagnostic decision support systems: A mixed-methods study | Common method bias | Responses were anonymous and the independent and dependent variables were not in a linear sequence. |
| Strohm, et al | Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors | Risk of sample being biassed towards individuals who are very interested in AI and have an above-average positive attitude towards it. | Be aware of this and address it in future research. |

### 4.2.7 Results

The raw data and individual coding is available in [Appendix 8.10](#_qquig7vd88yv) and the results for each theme are discussed below.

Performance Expectancy (PE)

Performance Expectancy is defined as “the degree to which an individual believes that using the system will help him or her attain gains in job performance” (Venkatesh et al., 2003). As a construct of the UTAUT model the codes were pre-defined and no changes were required.

Code PE1 refers to the efficiency of the workflow being changed by integrating an AI product into the workflow. There were no direct questions about efficiency and only one data item labelled PE1. This was in relation to a study that looked at the impact of the AI providing explanations in addition to the highlighting of x-rays and a respondent said that explanations “save time for a clinician” (Goel et al, 2022).

PE2 is defined as the effectiveness of the work produced by radiologists when an AI product is introduced into the clinical workflow. For improving effectiveness, 82.8% agreed that AI can help reduce radiation dose levels while maintaining optimal image quality (Botwe et al, 2021) and 75% agreed that it would improve diagnostic practice (Strom et al, 2020). A majority also believed that it would improve radiology practice and patient care through providing an avenue for more research (90.6%) and improving the quality assurance for its efficient diagnosis (84.9%) (Botwe et al, 2021). A study (Goel et al, 2022) which used a NASA-TLX questionnaire to compare between using AI and not using AI, found that using the AI had a statistically significant improvement to the performance measure which is typically the question, “How successful do you think you were in accomplishing the goals of the task set by the experimenter?” (Hart, 1988). However, a concern amongst 42% of respondents was that technical performance was inconsistent (Strom et al, 2020), supported by quotes in another study that, “multiple regions are not highlighted properly”, “[need] more precise colour coding”, “all regions need to be highlighted for diagnosis”, "[highlight] patterns of opacities”, and “the system makes mistakes, in some cases made incorrect assessment” (Goel et al, 2022). Further, when AI explanations accompanied the highlighting of x-rays, weak explanations decreased perceived reliability of the AI and radiologists preferred either strong explanations or none at all (Goel et al, 2022).

PE3 is about the extrinsic motivators and demotivators for radiologists when integrating an AI product into the workflow. Data classified as PE3 focused on the role of radiologists and their job security. For the role of radiologists 76.4% believed that AI would change their role which would lead to extended practices (Botwe et al, 2021). However, in another study, only 33.33% thought that AI would require professional identity and responsibilities to be reframed. In one study, the radiographers were concerned that AI would replace most jobs and negatively affect the profession rather than just being an assistive tool (61.3%) and especially in the role of image interpretation (67.3%) (Botwe et al, 2021). Further, 57.8% were concerned that AI would reduce their salary and displace them from their job in the future (Botwe et al, 2021). However, in other studies 62% disagreed that radiologists’ jobs were in danger due to AI (Huisman et al, 2021) and 71% thought that AI provided operational benefits (Strom et al, 2020).

Effort Expectancy (EE)

Effort expectancy is defined as the degree of ease associated with the use of the system (Venkatesh et al., 2003) The three constructs and there results that makeup effort expectancy include:

Perceived ease of use (EE1) is the first construct in effort expectancy. The degree to which a person believes that using a system would be free of effort. Examples of this construct include ‘I thought this system was easy to use’. F.M. Calisto discovered that 69% of users thought that the system was easy to use.

The second construct that is a component of the effort expectancy theme is learnability (EE2 The degree to which a person believes that learning how to use the system will be easy or difficult. An example of this construct asking the user questions such as if they believe that the system is ‘easy to learn’. Calistio saw that 75% of users would imagine that most people would learn to use this system very quickly and in the same study 62% disagreed with the statement, that they needed to learn a lot of things before I could get going with this system. Lea Strohm discovered that only 42% of users believed that technical knowledge is necessary.

The last construct for the theme is complexity (EE3) Which is the degree to which the system is perceived as being difficult to understand and use. The questions that answer this construct include asking the user questions such as if they believe that ‘the system is easy to use’. Ann Oper Res saw that task complexity has a significant positive influence on effort expectancy. A majority of studies also showed the low scores for complexity and 82% of users did not find the system unnecessarily complex (F.M. Calisto, 2021). In the same study 80% of users also did not think there was too much inconsistency in this system nor did the same 80% of users find using the system very cumbersome to use.

Additional findings also show that the theme effort expectancy is heavily influenced with F.M. Calisto proving that facilitating conditions (e.g., knowledge to use the system) for using AI in the clinical workflow, has a significant direct positive impact on effort expectancy. Calisto also found significant evidence that shows social influence (e.g., recommendation from the medical community) for using an AI system positively predicts effort expectancy. Significant evidence was also found that showed that trust was also found positively affects effort expectancy (i.e., ease of use) to use

AI systems Ann Oper Res (2020).

It is interesting to note that in several studies, it reveals that effort expectancy has no significance in influencing behavioural intention This phenomenon is demonstrated Ann Oper Res (2020) study where they found effort expectancy has an insignificant score of (β = 0.078, p > 0.05 when measuring its influence on behaviour intention. This idea's conclusion is further illustrated from the A.V. Prakash and S. Das 2021 study where they had a (β = 0.004 , p > 0.05). where they also measured the impact effort expectancy has on behavioural intention.

Social Influence

Social influence is defined as the degree to which an individual perceives that important others believe he or she should use the new system (Venkatesh et al., 2003). This construct was adapted directly from the UTAUT model and all the codes were already defined by the UTAUT model. Hence no further changes are required.

Code SI1 is called subject norm which reflects the potential influences from the important people from the participant’s social network (Venkatesh et al., 2003). Based on a study targeting African radiologists regarding the issue of the attitude toward the use of AI in radiology. 69% of the participants believed that most of their patients would be excited about the use of AI technology in their standard care. 26.5% of the participants held a mutual attitude for this topic. Only an insignificant 4.4% of the participants disagree with this statement (Botwe et al., 2021).

Code SI2 represents the social factors of the participants, it is a construct that directly mirrors the individual's agreement with the general trend of the perceived trend of the society held toward the use of AI products in medical imaging practices (Venkatesh et al., 2003). For the statement, ‘I am aware of AI as an emerging trend in medical imaging in Africa?’ 78.2% of the participants agreed with this statement, 19.8% remained neutral, and only 1.9% disagreed. In another study, three of the seven hospitals investigated were rated with innovative managers’, and 42% of the participants identified their hospitals as having ‘local champions’ of AI (Strohm et al., 2020).

There were no questions related to SI3 which is the ‘image’ factor. This factor reflects the degree to which use of an innovation is perceived to enhance one's image or status in one's social System (Venkatesh et al., 2003).

Facilitating Conditions

Facilitating conditions is defined as the degree to which an individual believes that an organisational and technical infrastructure exists to support use of the system. (Venkatesh et al., 2003)

Perceived behavioural conditions (FC1) are the perceptions of internal and external constraints on behaviour and encompass self-efficacy, resource-facilitating conditions and technology-facilitating conditions. FC1 refers to the user's beliefs about the ease of use of the system and their ability to use it effectively. This includes factors such as self-efficacy, resource availability, and technology familiarity. The questions that answer this construct include asking the user questions such as if they were “Given the resources, opportunities and knowledge it takes to use the system, it would be easy for me to use the system”. B.O. Botwe asked his respondents what they see as the single greatest barrier when learning about AI. The answers to the impactful question was, Lack of dedicated courses and learning materials: 219 (21.5%); Lack of mentorship, guidance and support from “experts" 136 (13.3%); Lack of evidence-based material and proof of improved clinical outcomes 117 (11.5%); Lack of time to learn new technologies 33 (3.2%); Lack of funding/investment for new technologies 360 (35.3%); Lack of motivation for change and interest to learn 66 (6.5%); Fear of the unknown 89 (8.7%).

The second construct is, support, (FC2) (due to the confusing name we changed the original construct name, facilitating conditions to support) are the objective factors in the environment that observers agree to make an act easy to do including the provision of computer support conditions. FC2 refers to the objective factors in the environment that make it easy to use the system, such as dedicated support staff, training materials, and user-friendly interfaces. The questions that answer this construct include asking the user questions such as if ‘Guidance was available to me in the selection of the system?’. It is important to highlight that a majority of users believed they require further education and/or training to be able to embrace these emerging AI trends in medical imaging (B.O. Botwe et al, 2021). However it is interesting to note in a different paper conducted from F.M. Calisto had a majority of respondents who felt they would not need the support of a technical person to be able to use this system. Additional data findings for the support construct, they found that 42% of respondents agreed with their workplace having unstructured planning and monitoring when implementing AI. Lea also noted 33.3% of users agreeing that there was an absence of guidelines/best practices. Lastly, lea saw that 33% of users mentioned that there was no empirical evidence.

The last construct for the facilitating conditions theme is, compatibility (FC3). It is defined as the degree to which an innovation is perceived as being consistent with existing values, needs and experiences of potential adopters’ conditions. FC3 refers to the degree to which the system is perceived as being compatible with the values, needs, and experiences of users. This includes factors such as system flexibility, system interoperability, and data security. Questions that answer this construct include asking the user questions such as if they felt that ‘Using the system is compatible with all aspects of my work?’ Lea Strohm found that 33% of the users Limited communication between departments. They also found 62.5% of users found the process to be an easy integration in PACS and 58.5% of users noted minimal workflow changes, Lea also discovered 42 % of users felt there was smooth integration in PACS.

F.M. Calisto found significant evidence showing that The facilitating conditions for using AI in the clinical workflow, positively predict users’ intentions to use it. In the same study it was discovered that there is significant evidence showing that The facilitating conditions (e.g., knowledge to use the system) for using AI in the clinical workflow, has direct positive impact on effort expectancy; Additionally, Ann Oper Res (2020) found that Technology characteristics has a significant a positive influence on effort expectancy.

Personal Attitude

As the research advanced to the data coding process, the team recognised that the four initial factors adapted from the UTAUT model were not sufficient. The team discovered that a significant percentage of the data extracted points toward the subjective opinion of the participants because of the nature of these articles are evaluation user experience. Thus, a new factor ‘personal attitude’ is added after group discussion. This factor directly reflects the subjective attitude of the participants toward AI technology used in medical imaging from four aspects: acceptance, willingness, trust, and risk.

The first code, PA1, is ‘Acceptance’. It intentionally examines the personal level of acceptance of the participants. There are three evaluations from the survey from the Africa doctor perspective evaluation which relates to PA1. Overall , 82% of participants believe that ‘I am excited about the integration of AI tools into medical imaging practice worldwide’, 14.3% remain neutral and only less than 4% disagree. 67.5% were concerned about the integration of AI into medical imaging practice worldwide, 22.5% remained neutral on that topic, and around 10% disagreed. 76% of the participants agree that the use of AI will bring more benefit than harm to this industry in Africa, 20% of the participants stay neutral, less than 4% disagree. Also, the study demonstrates what kind of AI tools the participants are looking for. The majority, over 85%, of the participants are looking for ‘Tools that support image reporting by detecting or flagging common conditions like tuberculosis’ (Botwe et al., 2021). A study evaluating the opinions from world-wide doctors reveals that 79% of the participants think radiologists should take the lead on the development of AI technology for medical imaging. 17% marked this question as maybe, and only 4% disagreed. Also, in one study it was determined that 48% of the participants hold an open and proactive attitude towards AI (Huisman et al, 2021). According to part two of the same study, 89% of the participants marked ‘maybe’ as an answer for the question ‘Can AI help improve diagnostic radiology?’, only 10% were certain. 78% and 77% believed that AI can help their work by becoming a second reader and workflow optimizer respectively (Huisman et al, 2021). 79% held an opinion on AI should be incorporated in the residency programs, only four percent were against it.

PA2 is the factor ‘willingness’, it is the degree of the intention the participant holds on using AI in medical imaging. The article ‘Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification’ indicates that with a specific kind of assistant, 51% of the participants strongly agreed to use the system compared to only 4% without the assistant. The level of confidence in using the tool increased from 31% to 67% after receiving the special assistance (Calisto et al,2021). In the Huisman study mentioned earlier, 77% of the participants would have chosen to become a radiologist again with the current AI knowledge, 15% of them were uncertain, and only 8% opposed this statement. Apart from that, 85% of the participants were willing to use AI software in the clinical setting, 14% thought maybe and only 1% believed no. 70% of the participants were interested in collaborating with computer scientists or data scientists to develop AI algorithms, 75% were considering learning AI as a topic even if it is not a requirement. In user experience evaluation using a modified UTAUT model, the construct ‘Behaviour Intention’ is measured using a five-point scale. The mean for this construct is 3.088, and the standard deviation is 1.038 (Praskash and Das, 2021).

PA3 is ‘Trust’, it is the degree to which the participant has faith in the innovation. For the study conducted by Fan et al, the participants found the AI interpretations with strong explanations the most trustworthy, the mean on a five-point scale is 2.12, the results with no explanations 2.10, and the results with weak explanations only 1.78 (Fan el al, 2021). The study ‘Medical practitioner’s adoption of intelligent clinical diagnostic decision support systems: A mixed-methods study’ pointed out that, on a five-point scale, the trust index for AI used in medical imaging is at the mean of 2.68, and the standard deviation is 0.983. The initial trust index mean is 2.797, standard deviation is 0.840 (Praskash and Das, 2021). For a study done in Netherlands hospitals, 58.5% of the participants revealed inconsistency in accepting and trusting the use of AI products. Only 25% showed acceptance and trust of referring clinicians (Strohm et al, 2021). Based on the Botwe study, 45.4% of the participants agree that ‘The use of AI tools could lead to unethical utilisation of patient data for unwarranted commercial purposes.’, 34.5% remains neutral and 20% disagreed (Botwe et al, 2021).

The last factor in the personal attitude section is ‘Risk’, labelled as PA4. It demonstrates different hurdles AI will bring to the participant itself and the radiology community in general. The aspects evaluated include ethical issues, legal issues, funding, and safety. In Botwe, et al's study, 64% of the participants agreed that ‘There is a possibility of errors associated with AI technologies integrated into my clinical radiography practice’, 10% of them disagreed, and 26% remained neutral, 64% agreed (Botwe et al, 2021). For the Huisman study part two, 35% of doctors identified ‘costs of development’ and anticipated hurdles to implementation, 38% identified ‘costs of software itself’ and anticipated hurdles to implementation. The major failings for implementation are ‘Knowledge of stakeholders’(56%) and ‘Generalizability of the software’ (39%) (Huisman et al, 2021). For the Praskash and Das study, on a five-point scale, the participants rated the medico-legal risk for using AI with a mean of 3.763, and performance risk with a mean of 3.624 (Praskash and Das, 2021). Lastly, the Strohm study highlighted that 21% of the participants for that study believed that the development of AI in the medical imaging industry pressures the healthcare budget. 42% of the participants doubted the quality and safety of the application. 58% of the participants the funding is uncertain. 33.33% of them thought there were legal and regulatory issues. 21% of them believed that the use of AI should be ‘Reference to post-market surveillance MDR’ and have ‘Legal responsibility for mistakes’ (Strohm et al, 2021).

### 4.2.8 Discussion

Synthesising the data from the nine included studies resulted in the emergence of five themes, performance expectancy, effort expectancy, social influence, facilitating conditions, and personal attitude. Four of these themes were from the conceptual framework UTAUT, and the fifth theme, personal attitude, became apparent from the data.

Performance Expectancy

It was anticipated that efficiency would be an important factor as it is a commonly described benefit of integrating an AI into the clinical workflow (Malamateniou et al, 2021). For instance, if the AI improves the efficiency of the workflow radiologists would more likely agree with integrating the AI. Whereas if it slows down the workflow and increases the workload, radiologists would not want to use it. However, in the included studies there were no direct questions about the efficiency of the workflow when using AI. Although in one survey, 77% of respondents agreed that AI can be used to help diagnostic radiology by workflow optimisation (Huisman et al, 2021). This highlights that radiologists do expect efficiency to be one of the benefits of integrating AI into their workflow and is an important factor to be considered. Research into efficiency may be present in studies that were not identified through this systematic review, and a potential reason is that it may have been approached through an objective comparison, rather than user experience. It would be beneficial to conduct further research into this factor to determine whether improved efficiency has been actually reported or observed when using AI, or simply remains an assumed benefit.

Effectiveness is another important consideration for determining whether to integrate an AI into the workflow. Increasing the effectiveness of work output can be observed as improved patient care, less incorrect diagnoses, earlier detection, and reduced radiation doses (van Leeuwen et al, 2022). A focus of improved effectiveness is improving diagnostic test accuracy, which refers to how well a test can discriminate between different conditions, for instance, a healthy person and a person with a disease (Šimundić, 2009). This was investigated in section 3.3 of this report. The data from the included studies indicated that the vast majority of people interacting with the field of radiology support the view that AI will increase effectiveness by improving patient care through increased research, quality assurance, and improved practices. However, 42% also reported concerns about the technical performance of the AI. Clinicians in another study reported concerns about the precision of x-ray highlighting and that in some cases the AI made a mistake. Overall, this suggests that while radiologists anticipate that increased effectiveness will be a benefit of integrating AI into their clinical workflow, there are current concerns about the performance of AI.That this concern will likely be addressed as AI algorithms continue to be improved through training and evidence of excellent diagnostic test accuracy emerges.

Extrinsic motivators and demotivators can have significant influence over whether radiologists support the integration of AI into their workflow. The included studies focused on the role of radiologists and their job security, with the responses varying greatly across different studies. When investigating the reason for the significant variance, the level of knowledge of AI emerged as an important factor. In the study where concerns of reduced salary, job loss, and negative impact on the profession were reported, 61.9% of radiographers had a basic knowledge of AI and 31.9% had no knowledge. However, in the study where these concerns were minor, 45% of radiologists had an intermediate or higher knowledge of AI. These findings are supported by the second study which found that a basic knowledge of AI increased the fear of replacement by AI, whereas, intermediate or advanced knowledge of AI decreased the fear of replacement by AI. This is a similar case for questions about the role and responsibilities of radiologists where the majority with no knowledge or a basic knowledge of AI thought that AI would change their role leading to extended practices. Whereas, in another study which did not report knowledge of AI, however, commented that participants were skewed towards those with a particular interest in AI, only a third thought that the role of radiologists would need to be reframed. Therefore, this suggests that assisting radiologists to learn more about AI to increase their knowledge and understanding will decrease their concerns of job security and changing roles, which will in turn increase their support of AI in their workflow.

In considering performance expectancy’s relationship with other factors, across multiple studies, it was found that performance expectancy is affected by social influence and effort expectancy. This suggests that others' opinions of AI, combined with how much effort they expect it requires impacts on how they expect it to perform. While performance expectancy affects intention to use the system and to a mixed extent, trust of the AI. This supports the earlier discussion that if people who work in radiology view the integration of AI into the workflow as something that positively impacts efficiency and effectiveness while not replacing their jobs, they would be more likely to accept its use.

Effort Expectancy (EE)

It is interesting to note that in several studies, it reveals that effort expectancy has no significance in influencing behavioural intention

A.V. Prakash and S. Das explain the reason for the insignificance of the effort expectancy metric is that ‘radiologists are used to highly complex machines in their routine clinical practice, and the ease of use associated with the system may not factor in as an influential criterion in the adoption-related decision-making. Another reason may perhaps be that radiologists did not foresee any difficulty in using ICDDSS, and they perceived the use of ICDDSS to be as easy as using their current systems/machines.’

Radiologists may simply assume that the software will have been designed with appropriate UX and that appropriate training will be provided. The product may also have been demonstrated to show the ease of use and radiologists may have had discussions or tested the system before undertaking the effort expectancy study.

Putting in despite there not being proof, effort expectancy is still important as shown in several studies from Venkatesh, where it is demonstrated that effort expectancy and performance expectancy are direct determinants of behavioural intention.

After deliberating the insignificant effort expectancy results, our team has concluded that this theme should not need to be high priority when figuring out what areas of focus when integrating CAD technology into the radiology practises. However, our team has concluded that the effort expectancy theme can still be an important aspect when working on improving the technology itself, for example, the data can be used by companies who are building CAD products for medical diagnosis. So these companies can take effort expectancy findings into account when designing their technology to make sure they improve their customer’s user experience. These findings can be done by getting their users to conduct effort expectancy focused surveys, questions could include “list the all the reasons that made you believe the system seemed easy to use” or “list the all the reasons that made you believe the system seemed difficult to use”, “list all reasons why you felt using the system was complex” and the last open question could involve “list all the reasons why you felt it was easy to use the system”.

Social Influence

Social influence was also an important factor in the UTAUT structure (Ursavaş, 2022), it was hypothesised to have an impact on the participants’ intention of use. For this modified UTAUT model in general, most of the studies related to social impact were able to prove that impact. More than that, some of the articles discover the factors impacting social influence, the most significant one is the factor trust.

It was expected that the use of AI is the current trend in the medical imaging industry because of the publication on how fast and beneficial AI development can be. In other words, even though there are still some doubts about the use of AI, society is embracing the benefits of AI and adapting to the changes AI development brings. As for SI1, ‘subjective norm’, it collects the attitudes from the people who are important to the participants including co-workers, patients, supervisors, and so on. It has been proven in the investigation on the perspective from Africa doctors, that around 70% of the patients were excited about the use of AI in diagnosis, this directly reflects the patients’ attitudes in Africa, and the patients’ attitudes is critical in medical work.

SI2 is the code for social factors which represents the perspective of the participants on the current trend of using AI in radiology, whether the participants believes that there is a trend or not, the opinion on the potential trend, were all evaluated within this factor. Still from the African doctor perspective study, it shows that approximately eighty percent of the participants believed that there is an emerging trend in Africa in the use of AI products in radiology and only two percent disbelieved in the trend. This is a critical evaluation because it successfully distinguishes that the majority of the doctors in Africa are aware that the use of AI is becoming more and more popular. In the other study conducted in the Netherlands, 3 of the 7 hospitals the participants are working are rated to have ‘innovative managers’ by the participants themselves. This could critically impact on the willingness the organisation holds when it comes to the acceptance of the innovative AI technology. The organisation is expected to hold a more open attitude in the use of AI products if there are innovative managers. However, three of seven hospitals is not a significant enough ratio to demonstrate the trend. In the same study, 42% of the participants identified their organisation to have a ‘local champion’ which refers to an individual who is willing to strongly advise the organisation to implement AI in medical work, in other words, an AI promoter. Considering the level of willingness of such promotion, 42% is considered to be reflective of the publication effort of AI products.

More importantly, if the construct ‘social influence’ is considered as a whole, many correlations with other factors in the UTAUT model could be identified. There are also some impacts being shown not only under the traditional UTAUT model but also the reflection of the personal attitude of the participants. In a study conducted internationally, the construct ‘Social Influence’ is proven to be a direct impact of ‘Effort Expectancy’. The theory behind this is that the social network of the participants could provide the necessary support on accepting changes due to innovation. If the participant could find assistance within his or her social circle, the effort of adapting new technology can be reduced. The study also proves that social influence has a positive impact on the willingness, trust, and risk regarding AI. It is indicated that if the medical community recommends the use of AI, the participant is less worried about the risks related to such technology and more willing to trust AI, hence gaining more willingness to use AI. This is common because the community is able to offer use cases as evidence, and this could easily affect the perspective of the participants. More importantly, the study also proves that social influence and performance expectancy are positively correlated. This result demonstrates the importance of social influence, and how it could alter the user perspective of AI products (Botwe et al., 2021).

Facilitating Conditions

Facilitating Conditions (FCs) are important because they provide the infrastructure and support needed to use a system effectively. FCs link to results because it impacts attitude and intention to use AI systems, as well as a positive impact on belief and usage of these systems. Increasing knowledge about FC can improve trust because it can help to overcome common barriers to effective implementation, such as lack of key knowledge. FCs play an important role in the successful adoption and implementation of clinical decision support systems (CDSSs), as they help to ensure that users have the necessary resources and support to use the system effectively.

In our research, we saw that a majority of users believed that the greatest barrier when working with A.I. is the lack of dedicated courses and learning materials (B.O. Botwe, 2021). However, none of the included studies in our thematic analysis specifies what specific courses and learning materials would be the most beneficial to provide the users to help ensure a smooth integration process. There is a wide variety of resources that could be distributed including, qXR instruction manuals, evidence proving the AIs accuracy, providing users a hands on experience with using a similar technology before they have qXR integrated into their own work practises or even basic information technology lessons. Therefore our team suggests further research to be done where various radiological work practices have different resources provided to their staff before the technology is integrated. These studies can then help identify which specific materials are crucial to provide to guarantee a positive workflow integration.

The focus on training should not just stop at the start of the integration process. B.O. Botwe also raises attention towards the need of having a large amount of support and training throughout the entire life span of the A.I. technology in the radiological process. This extension of support is due to a wide variety of reasons but a large one being a large proportion of users have no experience using technology (Merel Huisman, 2021) and therefore there is a likely chance that some users may need a longer amount of time with support than practices may have originally intended for.

In Calisto’s studies, they also pointed out that the large proportion of respondents agreed that their workplace had unstructured planning and monitoring when implementing. This high value causes great concern and it should be made clear to our client about the importance of having a structured implementation plan as well as ongoing monitoring when the technology has been integrated, The importance for a structured process is that it can create a sense of order to the workplace which is essential as there is a high chance of users are worried with this unfamiliar technology that will change their original workflow. Additionally by having users continuously monitor the new technology and the user’s experience towards the technology, it can be used as feedback for our stakeholders to see if the A.I. is working well and is being used and if not then they are able to then have the users re-group analysing reasons why which can allow for re-strategisation to ensure the AI is used and creating efficiency in the workplace.

Personal Attitude

It is fair to say, without the acceptance from the users, any further developments of AI as an innovation is needless, therefore it is critical to evaluate the general personal attitude of the users and identify potential hurdles they perceive. Only by then, the development and implementation of AI find the correct direction of development and grow healthily. The studies found according to this project’s specific PRISMA protocol highlights many personal attitude evaluations, more than that, there are some valuable relationships between personal attitude factors and other constructs from the UTAUT model identified by the articles.

PA1 reflects whether or not the participant welcomes AI as an innovative factor to the medical imaging industry as well as personal career development. In general, the majority of the radiologists are willing to accept the use of AI based on the result from PA1. Furthermore, the radiologists already figured out the designated roles for AI, including abnormality detector, second reader, and workflow optimizer. However, less than 50% of the participants do not hold an open mind regarding this matter, and around 80% of them do not believe AI will make a difference in their work quality, the main benefit AI brings are indicated as efficiency. They also showed both excitement and concern for the use of AI, 82% and 67.5% respectively, which demonstrated the mixed feelings they hold in the use of AI. It is also proven by the Fan study that there is a significant positive correlation between PA1 and effort expectancy in general. This is direct proof that if the user is willing to accept the product at a higher level, he or she will find the product easier to use. This phenomena has been shown in many other innovative products as well, for example smartphones. Another correlation is that inertia has a positive relationship with resistance to change, which is also not surprising because inertia and resistance to change are closely related in not wanting to adopt AI.

PA2 measures the perceived intention to trust the algorithm, the intention to learn about AI technology, and the intention to cooperate. It represents the personal inner willingness to use this technology. The interesting finding of this factor is that radiologists are more willing to trust the AI with assistance, and are more confident using the system with assistance. It is understandable that the radiologists require some level of assistance when using an innovative system. Therefore it is important for further developers to consider not only to develop with more functionalities but also offer proper assistance. The result of willingness is optimistic, almost all of the evaluations show more than 70% of the radiologists are willing to accept, trust, and use AI in their workflow. The important factors impacting PA2 are quite obvious, they are facilitating conditions in general, performance expectancy in general, perceived security and effectiveness, initial trust, extrinsic motivation, and social influence in general. Hence, these factors should all be considered when developing and promoting AI products in medical imaging in order to make the development and promotion process more comprehensive. On the other hand, willingness is also proven to impact factors including trust and risk. After discussion, it is to the team's belief that if the participants have the will to trust and use that certain product could be transformed into the actual trust for an individual.

PA3 mainly reflects the level of trust on the accuracy of AI in diagnoses. Some minor aspects related to implementation and job security. The result of the evaluation against trust reveals some neglectable issues. Multiple studies show that, on a five point scale, the mean index for trust evaluated is less than 3. Based on one of the studies, 58% of the participants showed inconsistency in trusting the product, and only 25% trust and accept a certain type of AI. These statistics are direct indications that even the radiologists are willing to embrace new technology, they do not trust the technology wholeheartedly. Another intriguing finding is that radiologists trust AI results with strong explanations rather than weak explanations, they even prefer no explanations more than weak explanations. This indicates that weak results can be confusing to radiologists which is a point the developers should hold accountable when developing AI products. It was assumed that risk will have a major impact on trust, and it was proven by multiple studies. The existing risks are undermining the trust of the users, so it is important for the AI community to tackle those risks. More importantly, trust is found to have an impact on effort expectancy and willingness. The reason behind this is because doctors will need to do more investigation into the technology in order for them to trust the technology which is effortsome. More than that, there are four factors influencing trust which are effort expectancy, performance expectancy, social influence, and willingness. None of these correlations shown are surprising.

PA4 is the factor reflecting the risks in implementing AI products based on the radiologists’ perspective. Almost all of the studies related to this topic are able to point out risks this industry is currently having. Around half of the participants are worried that AI products will produce both legal risks and performance risks. The participants are worried that the current legal system is not being able to hold anyone or hold the wrong person accountable if an AI related product makes a mistake medically. Another risk that the users are concerning is lack of funding. The ratios of the doctors identified risks were significant because risk evaluation is unlike performance evaluation, there is not an absolute safety barrier for risks, the best statistics for risk is always zero especially in the field of medical, it is related to peoples’ lives. It is proven that social influence and willingness are the impactors for risks. It is important that the studies highlighting the more willing a radiologist is, the less he identifies the risks. The insight on social influence impacting risks is also a critical piece of advice when it comes to tackling those risks. Furthermore, risk is considered as a factor impacting willingness, trust, and effectiveness which are not surprising.

In general, the evaluation of personal attitude shows a level of contradiction, even though high willingness and high acceptance are shown, the level of trust is low and the risks are obvious. It is encouraging to see the performance of AI being recognized and accepted, but the risks and trust issues are not being paid enough attention to by the developers or the government. The analysis toward trust and risk reveals there are still significant efforts to be made to eliminate those risks because the current results are not satisfactory.

### 4.2.9 Conclusion

A systematic review was undertaken to identify studies of people who work in radiology’s user experience and acceptance of AI. This search identified nine studies to be included for analysis, and while not as specific as was intended, it revealed broad insights that can be applied to qXR. In investigating the themes identified through the thematic analysis, they were found to be interconnected and explained impacts of AI on the workflow alongside the user experience and acceptance of people who work in radiology. Overall, this evaluation took the view that by considering the perceived impact of AI on the workflow and how people who work in radiology become accepting of AI in the workflow, this would reveal valuable information about workflow impact concerns and how to address them to support the adoption of AI.

Facilitating conditions had the least inter-connectedness as it was only found to impact on effort expectancy and intention to use AI, although both of these had very strong correlations. This indicates if an organisation provides good infrastructure that is compatible with the AI, and provides resources and support to its staff, it has significantly decreased the expected amount of effort required and encourages them to use the system. It was found that performance expectancy, social influence, and personal attitude directly impacted on a person’s intention to use AI in radiology. Additionally, social influence and personal attitude influenced performance expectancy and effort expectancy.

Based on the earlier discussion of the social influence and personal attitude themes and in seeking to further understand how someone develops a positive attitude towards AI and would influence others to adopt AI, we propose that it fundamentally comes down to knowledge. When a person gains knowledge of AI, such as an understanding of how it works, its benefits, and its risks, they are more likely to become more accepting of AI in their profession, willing to use it, and trust it. This was supported by various studies included for analysis which found that people with a higher level of knowledge about AI were more accepting of AI and those without knowledge were more fearful. By gaining knowledge about AI, this fear generally dissipates and AI becomes less intimidating. Then those with knowledge positively influence others to seek out knowledge of AI, creating a ripple effect of more people understanding AI, resulting in more people accepting it. Additionally, the facilitating conditions relate to the resources, instructions, and support provided, which is ultimately the organisation giving their medical professionals knowledge to confidently use the AI.

Additionally our team has concluded that there is a need to increase user’s knowledge as it can specifically help to improve trust in the users, as it can help to ensure that users understand the importance of these factors in the successful use of the system. In addition, increasing knowledge can help to identify potential barriers to adoption and implementation, and help to develop strategies to overcome these barriers. This idea is reinforced by studies conducted by Andreas B. Eisingerichl and M.D. Santoro found that an increase in knowledge had a positive and strong impact on trust. Hence why they should focus on providing knowledge to the stakeholders, including the information documentation/training to increase their trust with the system and the benefits it provides, the effectiveness and efficiency that comes when integrating a CAD system into a radiological practice? In addition, by providing users with experience in working with technology so users that have had limited experience using technology are not scared of using technology in their work practices and can get excited for the advantages that adopting technology can provide.

However, due to the limited research showing the impact that FC has on trust, our advice is for the client to track what resources made the biggest impact on users’ trust towards AI in their radiological practices. So future companies can then ensure their users have been provided the needed resources to ensure they have a smooth integration process. We propose further research into the importance of AI knowledge as an important factor impacting on the acceptance of AI in radiology, rather than merely a demographic or moderating factor. This could include more information about what participants know about AI to determine what specific knowledge and what level of knowledge is broadly needed for people to become more accepting of AI.

## 

## 4.3 What are the requirements in terms of IT infrastructure?

Our team reached out to Pushpendra Rawat, who is leading Qure AI for the APAC region. Rawat’s answer to integrating qXR is: ‘This can be deployed in both settings on-premise & Cloud. For cloud we do not require any additional hardware and existing systems can be used for installing our gateway, which anonymises the patient personal information before sending it to our cloud. For on-premise all data stays within your network.’

Through our research we found the common hosting requirements are:

* Cloud-Based - Qure.ai maintains cloud hosting predominantly through AWS and Microsoft Azure. These are highly secure reputable cloud hosting options and do not require any specific requirements for the Radiology practice
* API Based-Private Cloud - Maintaining a private cloud moves the security of the network and patches and upgrades into the responsibility of the Radiology practice.
* API Based-Local IT Infrastructure - No specific requirements are listed on provided from the Qure.ai representative. However, common hardware requirements are High-performance servers, AI enabled GPU’s and large storage systems with backup drives in case of failure.

With the advent of cloud-based hardware solutions such as virtual machines, having a good broadband internet connection means that infrastructure can be scaled up as required without needing to purchase any additional hardware. The cloud-based network security is maintained by ensuring data is encrypted and network access is limited to authorised and authenticated users.

Our advice, if you decide to not go the AI Marketplace route, is to use Qure.AI’s cloud-based services. Supporting our decision, several research papers have analysed the benefits in the adoption of cloud computing concluding that using SaaS provides great a wide amount of benefits compared to on-prem. Lewis (2022) found that by using SaaS advantages includes: the system being ‘readily affordable’ as it is a pay-what-you-use service. Lewis and the team also found using cloud computing there is an ‘efficiency improvement’ as the cloud providers, AWS and Microsoft Azure do all the maintenance for the program. By having the computer server, data storage, firewall, load balancer managed externally, it means radiology practices can focus on important tasks instead of doing mundane tasks such as patching and updates. With cloud public computing there comes great hesitation with data security. However, as stated on the qure.ai website they are HIPAA which ensures data is de-identified when data is being processed in the qXR’s software. The data is protected as qXR encrypts the data at rest and in transit. The main issues with using cloud computing is the heavy reliance on the internet and the decreased control. However, if the radiology practice does have poor connection they can consider getting additional internet hotspots, purchasing more internet connections to connect to the cloud server is significantly cheaper than having on-premise hardware.

## 4.4 Interoperability - How can the data be exported for research and other purposes?

## 

The product qXR distinguishes between normal and abnormal chest X-rays, classifying them into three categories, normal, abnormal, undetermined (Incepto Medical, 2022). For the undetermined results, a human expert might need to be involved for further diagnosis. For the abnormality detection, qXR can help diagnose approximately 24 diseases and a report is generated for further examination. The deployment of qXR along with other products of the qure.ai company are all based on two integration models, API and PACS (Qure AI, 2022). Those features are mentioned in the integration into workflow section previously in this report but it is equally important when regarding the matter of data exportation. Also, another important technology being utilised for exporting the results is cloud service, the client has the option of uploading to the appointed cloud platform provided by the company Qure.ai itself, or storing the scans and reports into a private cloud based on the client’s preference. Not to mention, downloading the scans and reports to local hard drives is also an option, but this method can be time-consuming and require a large amount of storage. For further research and other purposes, the researchers can simply export the report generated by qXR and conduct the research. For example, if the research aims to investigate the correlation of smoking and lung cancer, the researcher can extract the reports of a group of patients randomly who did chest x-rays and identified smoking as a preliminary condition. The researcher can then compare a matched group study for smokers’ and non-smoker’s chest x-rays. qXR results will be the factor determining if the patient is diagnosed with lung cancer or not based on the abnormality that appears. This is comparatively a simple and convenient process for any parties planning on extracting the designated data from qXR.

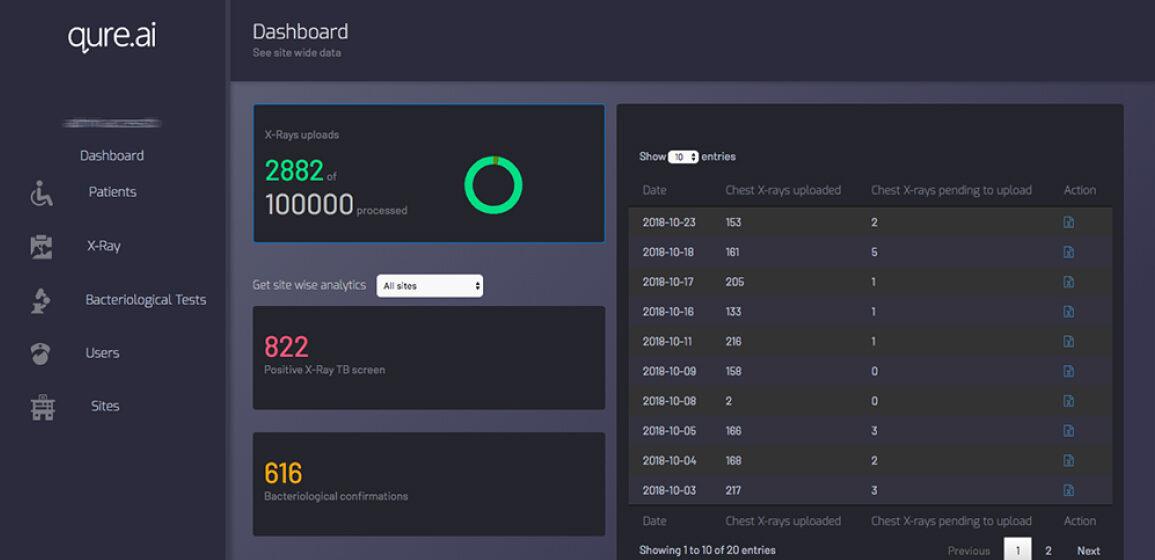


Figure 1: This is the patient management dashboard for all the products qXR company is using.

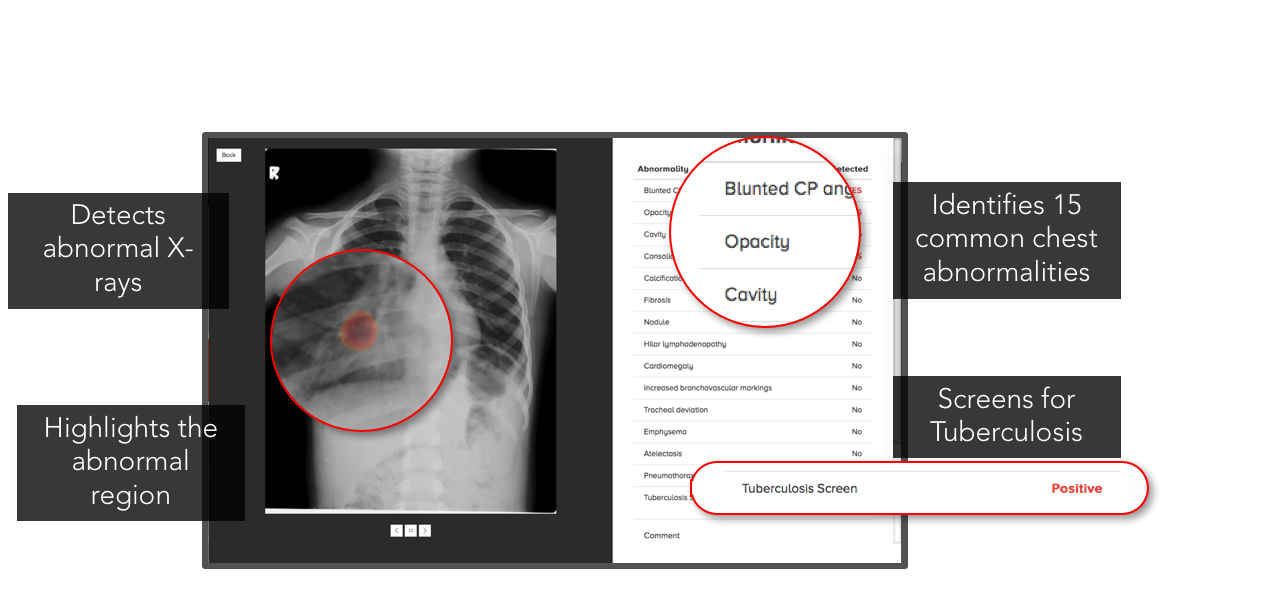


Figure 2: This is a sample qXR generated result. All the features above can be used for further research. The main purpose is to list out the abnormal part of the patient’s scan, point out the location of such abnormality and the type of abnormality, such as opacity, cavity, nodule…

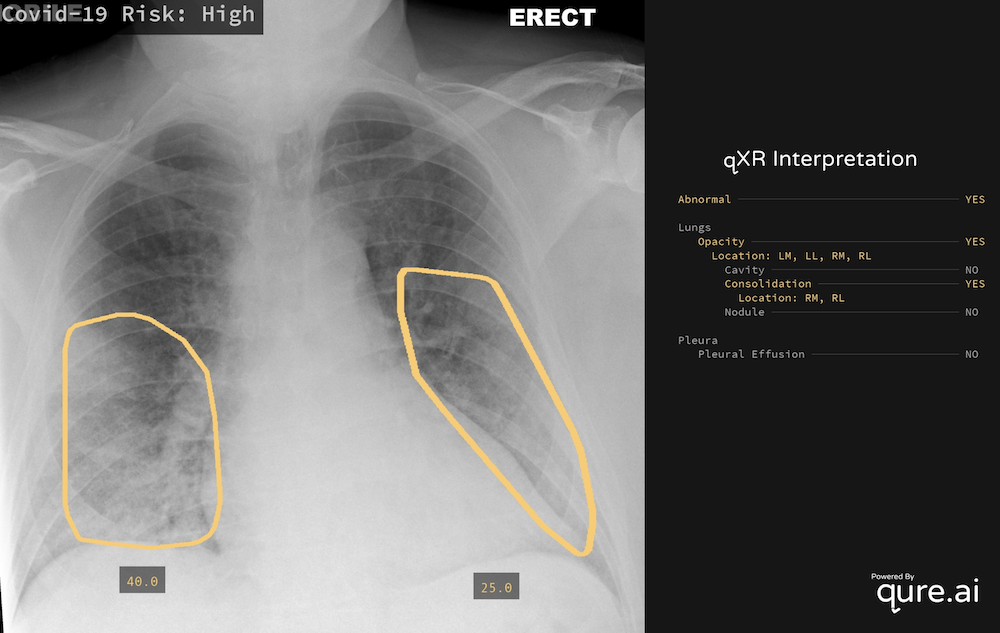


Figure 3: Another sample output from the qXR software .

Such results are solidly based on radiology and medical practices, they are easy to read and further analysed by people with medical expertise. Assuming qXR can produce results with high accuracy, such results can be widely accepted by the medical imaging community.

## 4.5 Will the data be accessible to non-radiologists?

Pushpendra Rawat was again contacted here and revealed: ‘For cloud based & on-premise (intranet) deployments all images can be made available to all physicians with their logins. These can be shared with external physicians as well with just a web-link as well. An original image is also given to remove any biases for the readers.’

The raw data from the X-Ray and qXR’s output is not sent to anyone other than the reporting radiologists. A report generated from the qXR software will be sent to the practitioner who requested the patient’s scan. This report will state Yes or No to the thirty abnormalities/diagnoses that are tested. Once the report is received, the doctor will discuss these results with the patient. This standardised protocol is due to compliance procedures. The practitioner is the one who is responsible for the patient's care. The practitioner will use the information in the report to make decisions about the patient's treatment.

## 4.6 Are the AI model’s results interpretable?

Rawat again informed our response to integrating qXR: ‘Secondary capture has AI annotations on top of the original image. A DICOM viewer also comes along with it with tools such as measurements, levels, zoom etc.’

The generated report is easily interpretable for all medical healthcare professionals, clearly stating if there is or is not a detection of a certain abnormality. This simple reporting structure decreases the time for a practitioner to interpret the results, hence improving efficiency in the workplace.

The structured report involves:

* Probability score as well as dichotomous output indicating whether each abnormality is present or absent
* Probability score for TB as well as dichotomous output indicating whether TB is likely present or likely absent
* A box indicating the location of the abnormalities
* Abnormalities detected by the product for which a separate abnormality score is given include: abnormal, TB, opacities (atelectasis, cavities, calcification, consolidation, fibrosis, nodules, reticulonodular pattern), emphysema/hyperinflation, pleural effusion, blunted costophrenic angle, pneumothorax, cardiomegaly, tracheal shift, degenerative spine changes, scoliosis, hilar prominence, rib fractures, COVID-19, pneumoperitoneum, mediastinal widening, elevated hemidiaphragm, abnormal diaphragm shape, lines and tubes (4 types) and lung nodule malignancy risk for nodules

# 5 Regulatory and Legal Aspects

## 5.1 Does the AI application comply with the local medical device regulations?

### 5.1.1 Australian medical device regulation

In Australia, qXR is a medical device as defined in s41BD of the *Therapeutic Goods Act 1989* (Cth) because it is a software used for human beings for the purpose of diagnosing a disease. Therefore, it must be registered with the TGA and added to the Australian Register of Therapeutic Goods (ARTG) to be supplied in Australia. qXR is not included on the ARTG, which means that it currently cannot be used in Australia.

### 5.1.2 Process to comply with local medical device regulations

There are two applications that must be submitted and approved in sequential order for the inclusion of a medical device on the ARTG. These applications must be submitted by a sponsor, who is the person that imports the goods to Australia (s3(1) *Therapeutic Goods Act 1989* (Cth)). The sponsor must be an Australian resident and can be a person or company.

The first application is a Manufacturer’s Evidence application containing a TGA conformity assessment or evidence from a comparable overseas regulator (Therapeutic Goods Administration (TGA), 2022).

The second application is the inclusion on the ARTG application and is submitted through the TGA Business Services portal. The application includes the sponsor certifying that the device is a medical device, it is intended for a specific purpose, it has the correct medical device classification, and it complies with the essential principles (s41FC *Therapeutic Goods Act 1989* (Cth)). The Therapeutic Goods Regulations contain six general essential principles. It also contains specific principles and the design and construction of the medical device. The sponsor also must show conformity assessment procedures have been applied to the device and there is sufficient information available to substantiate the application of them.

### 5.1.3 Applying the process to qXR

For the Manufacturer’s Evidence application, qXR has evidence from a comparable overseas regulator. Qure.ai has obtained a class IIa CE-mark certification for qXR. This means qXR is a medical device that has been approved for use in the European Economic Area. It is under the European Union Directives which is a comparable overseas regulator recognised by Australia.

For the inclusion on the ARTG application, as stated above, qXR is a medical device under Australian law and Qure.ai states that its intended purpose is to aid in the detection of abnormal findings on a chest x-ray. Medical devices are classified based on the risk they pose and Australia typically provides the same classification as the European Union. Hence, it is likely that qXR will also receive a class IIa classification. This is supported by a TGA example of a class IIa medical device as one that provides information to a relevant health professional to support diagnostic decision-making for a serious disease. It avoids a higher classification because the results are intended to be read by a radiologist, not an untrained person, and it is intended to support decision-making, rather than replace the decision-maker. It is expected that qXR would satisfy the essential principles, as outlined below:

1. Use of medical devices not to compromise health and safety: as it is software, it cannot physically harm a patient or other and it is intended to be used by radiologists in the workflow of reviewing patient x-rays, therefore, they have the appropriate technical knowledge to check whether qXR is correct. The main risk that qXR poses is that it fails to identify an abnormality on a chest x-ray. This risk is mitigated because it is intended to be used by a radiologist who is expected to also review the x-ray themselves.
2. Design and construction of medical devices to conform with safety principles: qXR’s algorithm has been thoroughly trained and tested with a range of chest x-rays and found to have a high accuracy.
3. Medical devices to be suitable for intended purpose: there has been extensive studies on the performance of qXR which all show that it achieves its intended purpose and is highly accurate.
4. Long-term safety: qXR is only briefly used for analysing a chest x-ray, therefore, there are no long-term safety concerns.
5. Medical devices not to be adversely affected by transport or storage: qXR does not have any components that can be damaged or affected by transport or storage.
6. Benefits of medical devices to outweigh any undesirable effects: there is a risk that radiologists may become over-reliant on qXR’s results and qXR may miss something that can lead to a mis-diagnosis. However, this risk is outweighed by the benefits that it provides as an additional tool that radiologists can use in their workflow of diagnosing tuberculosis.

Qure.ai stated that qXR received a conformity assessment by a European Notified Body and an audit of their quality management system. In addition, Qure.ai would need to provide additional information about qXR to support the application.

In conclusion, qXR is not currently approved for use in Australia because it is not included on the ARTG. This process can be easily started because it is certified in Europe. The application for the ARTG will need to be discussed with Qure.ai to ensure they have sufficient information to support it and ensure that it can be approved for use in Australia.

## 5.2 Does the AI application comply with the data protection regulations?

### 5.2.1 Application of Australian data protection regulation to ABC and its use of qXR

Australia’s data protection regulations are contained within the Privacy Act 1988 (Cth). ABC is a health services company and using qXR is part of an activity to diagnose an illness and requires storing health information about patients to provide that health service (ss 6FB, 6FA Privacy Act 1988 (Cth)). This means ABC is required to comply with the Australian Privacy Principles regardless of their annual turnover (ss 15, 6(1), 6C, 6D(4)(b) Privacy Act 1988 (Cth)).

Therefore, it is expected that ABC would already be complying with Australia’s data protection regulations, however, it must be determined whether qXR complies with the Australian Privacy Principles.

### 5.2.2 qXR and the Australian Privacy Principles

The Australian Privacy Principles (APPs) are set out in Schedule 1 of the Privacy Act 1988 (Cth) and contain 13 principles which relate to data privacy and managing it in an open way, data collection, use and disclosure of information, integrity of information, and access and correction of personal information. Qure.ai does not state whether qXR complies with the APPs, however, it is likely to do so because its products comply with the United States Health Insurance Portability and Accountability Act and the European Union’s General Data Protection Regulations.

1. Data must be managed in an open and transparent way. This includes having an APP Privacy Policy. ABC would need to update their existing policy to include information such as data collection and use specific to qXR. For example, collecting chest x-rays of patients to be used by qXR to analyse it which results in the creation and collection of more patient health data.
2. Individuals must have the opportunity to be anonymous or use a pseudonym, unless it is impractical for ABC to deal with anonymous or pseudonymous individuals. ABC is providing a health service and by its nature, it is impractical for someone to remain anonymous or use a pseudonym when seeking medical assistance. However, in terms of qXR itself, if ABC uses cloud processing, the data is de-identified before it leaves its premises. Whereas if processing occurs on ABC owned servers, the data may not be de-identified.
3. The collection of solicited personal information. All information collected by ABC is classified as sensitive information because it is information about the health of individuals and their personal information collected to provide a health service. Therefore, the collection of the information must be reasonably necessary and the individual consents to it. Therefore, when using qXR, the patient’s consent must be sought first and only the necessary information collected.
4. Dealing with unsolicited information. It is unlikely that ABC will receive unsolicited information in the course of using qXR.
5. Notifying an individual of the collection of their personal information. Since all information collected when using qXR is sensitive information, individuals are aware that their data has been collected and they are notified within the process of giving consent. Therefore, there are only limited circumstances when this would be relevant, for example, if their data was provided by someone else.
6. Personal information cannot be used or disclosed for a secondary purpose unless that purpose is directly related to the primary purpose. The primary purpose of collecting data to use qXR is to provide a health service for diagnosing a lung abnormality. Related purposes
7. Sensitive information cannot be used or disclosed for direct marketing unless the individual consents to it. ABC is using qXR to provide a health service, so it is expected that no data will be used for direct marketing, however, if ABC wishes to, they must ask for the individual’s consent first.
8. Before disclosing information to an overseas recipient, reasonable steps must be taken to ensure that they comply with the APPs. This may be relevant if ABC chooses a cloud-based solution for qXR because Qure.ai is an Indian-based company. However, as discussed through these principles and its certification of complying with US and European Union data protection regulations, it is expected that Qure.ai would comply with the APPs.
9. Government-related identifiers cannot be used as individual identifiers by an organisation. ABC would need to use alternative identifiers for individuals for storing data collected for the use of and from qXR as they would already with the data they collect.
10. Reasonable steps must be taken to ensure that personal information is up-to-date and accurate. During the workflow which includes using qXR, the individual must be asked if their details are correct.
11. Reasonable steps must be taken to ensure the security of personal information. In addition to ABC’s existing controls, the security of qXR itself is important. Qure.ai states that its products are certified under two International Organization for Standardization standards. ISO/IEC 27001 is an information security standard and ISO 13485:2016 which is specific to medical devices and includes the protection of confidential data. In addition it has a cybersecurity team and submits cybersecurity audit reports to US and European regulatory bodies.
12. Provide an individual with access to their personal information. ABC’s existing procedures would outline the process for gaining access to information and whether there are any grounds upon which the access can be refused.
13. Reasonable steps must be taken to correct personal information that is incorrect. In line with ABC’s existing procedures, in personal information related to the use of qXR, the individual should be contacted so the information can be corrected.

### 

### 5.2.3 Action Required by ABC

From the above consideration of the APPs, it is anticipated that there will be no issue with ABC using qXR and complying with their obligations under the Privacy Act 1988 (Cth). Although ABC will need to take some steps to ensure their compliance such as updating their APP Privacy Policy, this will be minimal because they already comply with the Act in relation to the other services they provide.

# 6 Financial and Support Services Considerations

## 6.1 What is the licensing model?

The exact costs of implementing and using qXR is not elaborated upon in the company’s official website or its vendors’ website. The pricing model for qXR is pay-per-use and licensing is offered in year periods determined by the volume of scans required, there is also a further deployment and set up cost that is charged according to how many centres it is to be deployed in (Carmichael et al., 2022). AI marketplace was contacted directly as they are a seller of qXR, they relayed that the exact price differs according to different geographical locations as the levels of implementation vary for different healthcare systems. AI marketplace was not able to confirm the exact pricing for any one region due to a confidential agreement of pricing, however, the platform mentions that the pricing of qXR is considered ‘competitive’.

Our team leader reached out to the Citadel Health general manager, Dave Crocket, to get guidance on pricing for CAD software, Citadel Health is one of the leading radiology AI marketplaces and is therefore expected to have more specific knowledge on qXR’s price structure. Crocket stated that there are several different factors that contribute to the quote for the software implementation and subscription. These factors include:

* Number of annual studies
* Number of concurrent users
* Number of sites
* Further specific requirements

These factors play a significant part in the cost as the level of scalability and the amount of people using the same server would affect the amount of resources used from qure.ai and the AI marketplace.

## 6.2 How are user training and follow-up handled?

There is a specific four-step workflow that the company qure.ai follows regarding user training. The first step is orientation, this is a general introduction of the product and its functionality, during this process the clients will gain a preliminary understanding of AI itself and how it is used in the medical imaging industry. The second step is technical readiness which guides clients on the specific technical operation of qXR, after this stage clients are expected to fully comprehend how to operate the product and interpret the result at the clinical level the clients are also expected to familiarise themselves with the interface of the qXR application. The service level agreement will be demonstrated and comprehended by the clients after the first two steps, the company will then assist the clients in deploying the system to their own devices including computers and CT scanner. After these steps are completed qure.ai offers extra training and support throughout the client’s use of the system. The training offered at this stage is based on the client's specific needs, the client is able to contact Qure and ask for further training regarding the use of this product (*How We Deploy*, n.d.).

As for the follow-up service, there are two teams responsible for communication with the clients, the client engagement team and project delivery team and their technical lead. The clients can contact the client engagement team directly if they encounter any problem and be assisted in diagnosing and fixing said problem. The project delivery team, on the other hand, can assist the clients in designing the best technical approach for a specific task according to the nature of the task. Hence, the follow-up service is ensured for any client who is willing to purchase qXR.

## 6.3 How is the maintenance of the product ensured?

Our team reached out to Pushpendra Rawat who said “We have a dedicated team which provides support for all our clients in running the services. There is no AMC for cloud deployments.”

The service team from qure.ai constantly monitors the performance of the qXR to ensure that it is running smoothly and efficiently and in the case of any issues, the team is always available to provide support and assistance. The need for maintenance is to prevent potential malfunctions and erroneous results, the consequences of this are addressed in question 6.4.

## 6.4 How will potential malfunctions or erroneous results be handled?

The process for identifying potential malfunctions and erroneous results from qXR scans may vary depending on the particular implementation. However, some possible methods for identifying such errors could include comparing the results of multiple qXR scans of the same patient to look for discrepancies, reviewing the images generated by the qXR scan for signs of artefacts or other abnormalities or consulting with a radiologist or other expert to interpret the scan results, reviewing error logs, monitoring user feedback and testing the qXR system regularly.

If a potential malfunction is detected from qXR, the system will automatically generate a notification to the user. The qure.ai team will then investigate and determine the root cause. If the problem is identified as a software issue, the qure.Ai engineers patch the issue and write a regression test to make sure it is resolved and the fix can then be deployed. If the root cause is a hardware issue however, the hardware will simply be replaced.

There are a few potential consequences of erroneous results:

1. Inaccurate results – If qXR malfunctions, it could provide inaccurate results. This could lead to patients not getting the proper treatment or diagnosis.
   1. False positives – This could lead to patients being unnecessarily sent for expensive and invasive molecular testing for conditions they do not have (Clin Med (Lond) 2021).
   2. False negatives – This could lead to patients not getting the treatment or diagnosis they need and can then go on to infect others (Ying Lui 2020).
2. Delayed results – Another potential consequence of a qXR malfunction is delayed results. This could cause a delay in treatment or diagnosis, which could be detrimental to the patient. (Mei-Sing et Al 2018)

Medical healthcare readers must be aware of the consequences of potential malfunctions and erroneous results from qXR by qure.ai. Having the clinic’s physicians consistently check the software can prevent incorrect diagnoses and time delays, as they can temporarily stop the upcoming chest x-rays that need to be processed. This will allow time for the software to be fixed and in the meantime getting a radiologist to do the entire CXR reading .

# 7 Conclusion

## 7.1 Summary of Evaluation

### 7.1.1 Relevance

The aim for using qXR is to use AI to reduce the gap between the TB incidence and the actual number of reported cases in order to streamline and make diagnosis more convenient. qXR can make the TB diagnosis more time-efficient, as the processing time for qXR is normally less than 1 minute, therefore, it can save a great deal of time and allow patients to receive care faster. As qXR will interpret the radiology result automatically, it can make diagnosis more convenient and prevent misdiagnosis from human error.

Using AI in radiology still has risks, it may bring ethical issues and legal liability problems, the radiologists may over rely on the AI and not use it as an assistive tool which can in turn lead to misdiagnosis, furthermore, some countries may have not enough laws enacted to regulate the use of AI in diagnosis. The AI system may also be attacked, in this case, private patient data will be vulnerable and could be leaked to the hacker and used for malicious purposes.

| Recommendation One | qXR should be used by the ABC in order to optimise workflow |
| --- | --- |
| Recommendation Two | qXR can be used to prevent misdiagnosis of patients |
| Recommendation Three | Maintain a risk register to monitor and manage the potential risks of using qXR |

### 7.1.2 Performance and Validation

The key evaluation that took place in this section was with regard to the DTA of qXR when compared to similarly certified AI products, it was found that CAD4TB and qXR both had similar performances that were superior to other certified products, however, qXR slightly outperformed CAD4TB.

This section also evaluated the ways in which qXR was trained as well as its design specifications, this indicated it was trained through deep learning and initial training of the algorithm utilised 24,384 CXR samples.

Biases of the algorithm as identified by developers were also analysed, it was found they had accounted for selection bias by choosing their scan datasets which were used to validate the product in a way which did not unfairly bias the process. Another bias in AI detection was found to be underdiagnosis bias as this can lead to underserved populations not receiving necessary care for their conditions although there was no clear mitigation of this bias from qure.ai.

Finally, this section concluded qXR was adapting as new data was attained by identifying that over one million further CXRs had been utilised in additional training for the algorithm since 2020, this shows that the algorithm is getting more accurate and can be more trusted by clients, further to this, we found the qXR product has been previously adapted to meet other diagnosis needs such as repurposing the product in 2020 to diagnose COVID-19.

| Recommendation One | qXR should be pursued as the superior DTA choice in AI-assisted TB detection |
| --- | --- |
| Recommendation Two | ABC should keep in mind issues of underdiagnosis as this could lead to members of underserved communities being misdiagnosed |
| Recommendation Three | Patients with a known history of TB should have extra screening when utilising AI tools |

### 7.1.3 Usability and Integration

From our research our team has concluded that the stake holder’s level of knowledge and facilitating conditions are most important when integrating CAD technologies, therefore, focus needs to be placed on providing the evidence supporting these technologies. Further investigation is suggested into experimenting with various information that is provided to the users to identify what resources have the biggest impact on user trust towards AI in their radiological practices. Future companies can then ensure their users have been provided the necessary resources to ensure they have a smooth integration process.

The use of an AI marketplace vendor removes this complexity as the integration is managed by the distributor. Since distributors communicate with multiple vendors and are familiar with the integration process for multiple different algorithms they may work well in radiology practices that have not adopted single vendor strategies.

However, if you decide against the AI marketplace our suggestion is to use the qure.ai cloud-based solution. As stated on the qure.ai website they are HIPAA which ensures data is de-identified when data is being processed in the qXR’s software. The data is protected as qXR encrypts the data at rest and in transit. The main issues with using cloud computing is the heavy reliance on the internet and the decreased control.

As our client’s end goal is to increase efficiency and effectiveness, qure.ai has a suite of products that can enhance the radiological practice workflow even more compared to using qXR as a standalone product. Using a combination of products such as the qTrack in addition to qXR will increase workflow efficiency and accuracy. The qTrack’s management platform provides a centralised database, automates monotonous tasks and due to the automation can improve data quality all while being significantly faster than having manual input from admin staff, allowing admin staff to focus on more important tasks instead.

| Recommendation One | Use of an AI marketplace vendor removes this complexity as the integration is managed by the distributor. |
| --- | --- |
| If you choose to go the API route - Our advice is to use the qure.AI’s cloud-based services. |
| Recommendation Two | Take advantage of the supporting products made from qure.Ai such as qTrack. |
| Recommendation Three | Focus on the level of knowledge every stakeholders and spend time experimenting what are the key pieces of knowledge make the biggest impact |
| Recommendation Four | Ensure that the qXR is compatible with existing systems and staff are given resources and support to facilitate the integration of qXR into the clinical workflow. |

### 7.1.4 Regulatory and Legal Aspects

The legal status of qXR in Australia and whether it complies with data regulations was evaluated. It was found that qXR would likely be a class IIa medical device, which means that to be used in Australia it must be included on the Australian Register of Therapeutic Goods, which it currently is not. However, it is certified in the European Economic Zone which is one of the first steps for its inclusion on the register. Therefore, it is recommended that to use qXR, ABC seeks a sponsor for it to submit an application to the Therapeutic Goods Administration. If ABC chooses to purchase qXR through an AI Marketplace, they would likely become its sponsor. Alternatively, ABC could choose to be qXR’s sponsor, however, this would have significant upfront and moderate ongoing compliance requirements. In terms of data protection, it is expected that ABC would already be complying with the Australian Privacy Principles (APPs) contained in the Privacy Act and by considering the APPs, qXR would comply with them. Although, it is recommended that ABC updates their privacy policies.

| Recommendation One | ABC seeks a sponsor for qXR to be included in the ARTG so it can be legally used as a medical device in Australia. |
| --- | --- |
| Recommendation Two | ABC updates their privacy policy and any related documentation to ensure continued compliance with the Australian Privacy Principles. |

### 7.1.5 Financial and Support Considerations

Even Though the exact pricing can not be accurately disclosed, the product qXR has been informed to have a comparatively ‘competitive’ pricing, according to multiple vendors. The price for qXR differs regionally and is charged by the method of pay-per-use, however volume of scans, number of concurrent users and number of sites are all determining factors regarding the charge of the product qXR.

It was found that the best way to purchase qXR is through AI marketplace. The most obvious reason is that it provides a personalised service on implementation and installation. The targeted clients are hospitals and it is highly unlikely for them to have the technical ability to install the product by themselves. AI marketplace in this case, is able to provide a personalised service, adjust according to different circumstances, and help integrate the system into a hospital’s unique workflow.

| Recommendation One | Purchase through an AI marketplace because it helps with installation. |
| --- | --- |

## 7.2 Recommendation

Based on the evaluation of qXR and specific recommendations above, we recommend that ABC proceeds with a pilot of qXR. As AI technology is an emerging technology in medical imaging, ABC should consider using AI products to help aid with diagnosis. In addition, according to the previous technical analysis, qXR performed the best in all certified AI products in terms of sensitivity and sensitivity in DTA studies. Using qXR products such as qTrack will potentially be beneficial for ABC in efficiently diagnosing and managing databases. Moreover, since there aren't many studies related to UX of qXR, having a pilot study for qXR can help ABC identify the performance of qXR in their unique workflow, a pilot program will also assist in changing the attitude and thinking of using qXR for radiologists.

# 

# 8 Appendix

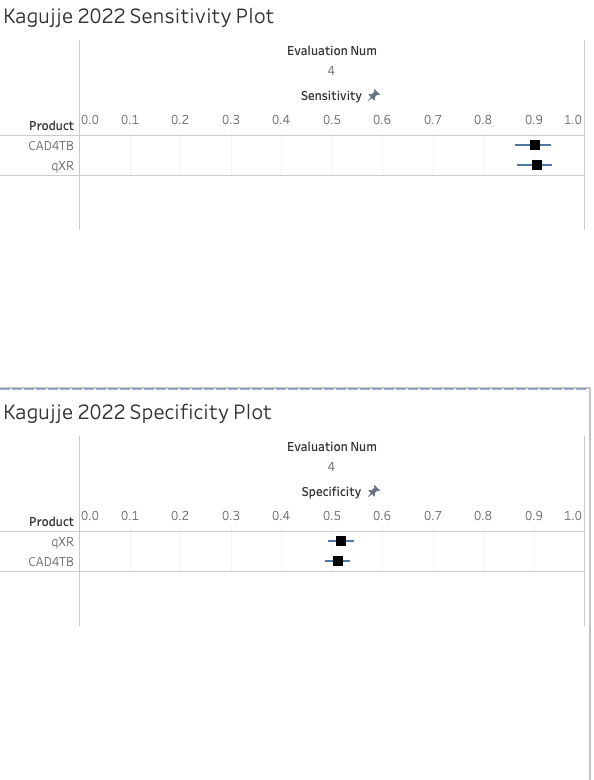
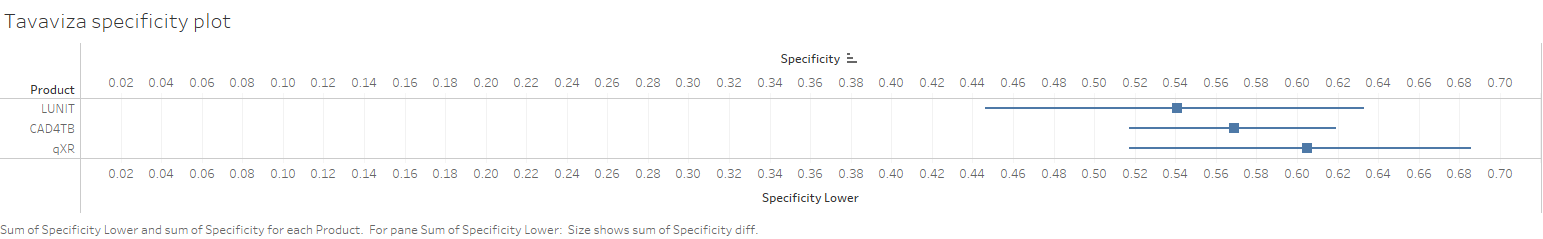
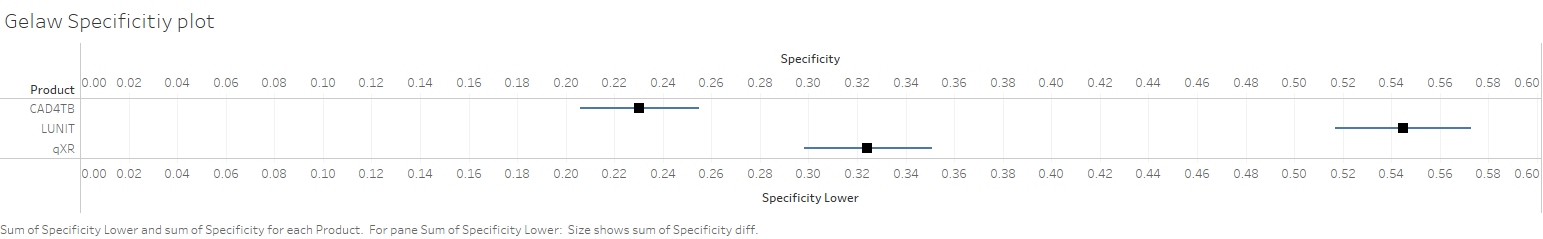
## 8.1 Technical Systematic Review Protocol

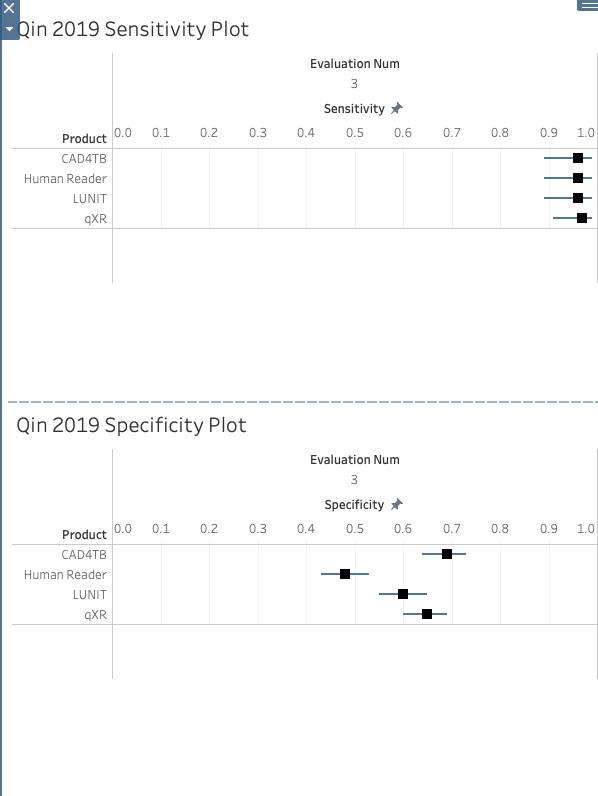
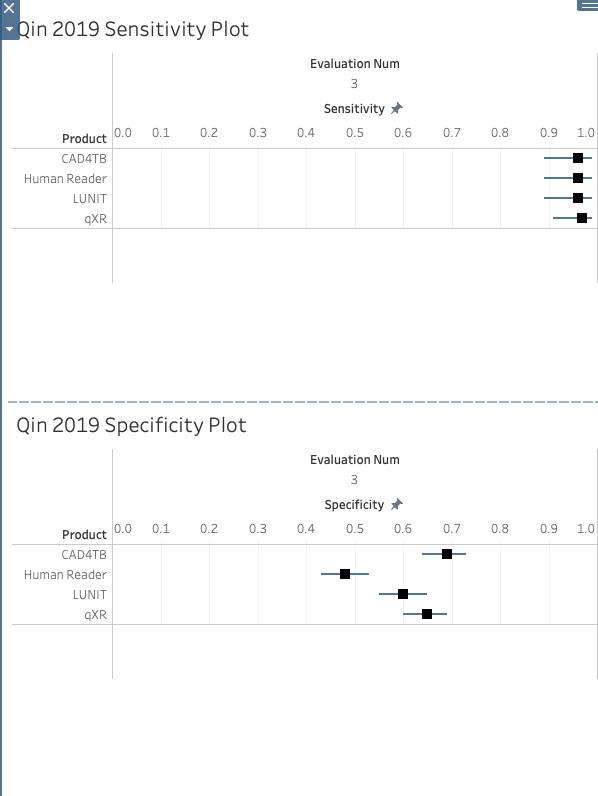
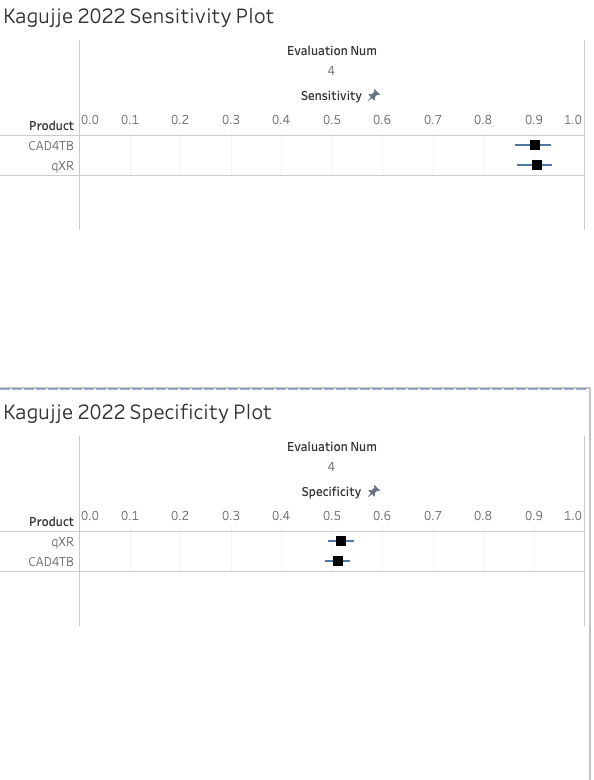
This document is available as supplementary material in the Appendix folder and labelled as ‘Appendix 1’.

## 8.2 Technical Articles

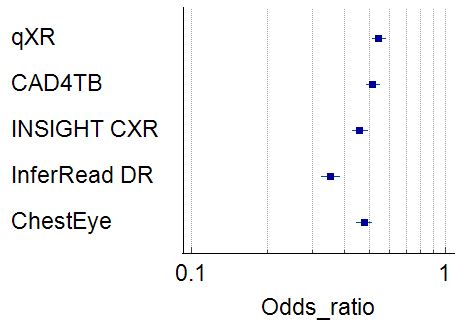
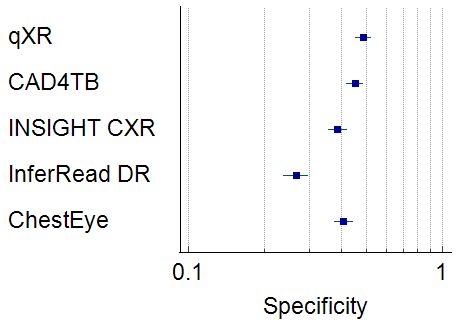
This document is available as supplementary material in the Appendix folder and labelled as ‘Appendix 2’.

## 8.3 Forest Plots for DTA

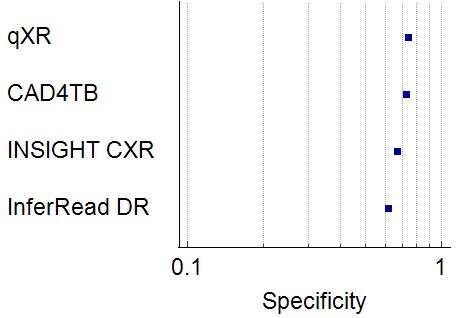




Zhi Zhen Qin 2021

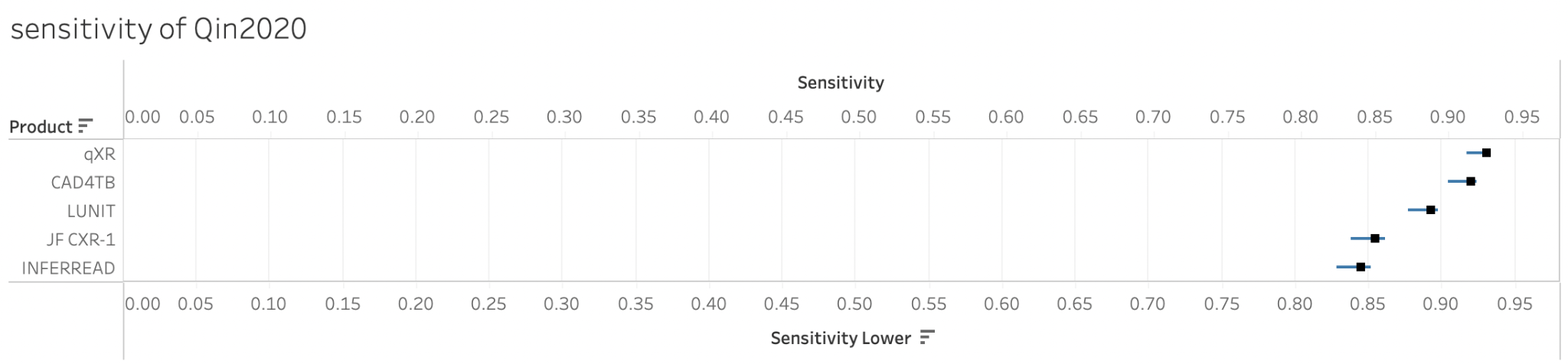


Andrew J

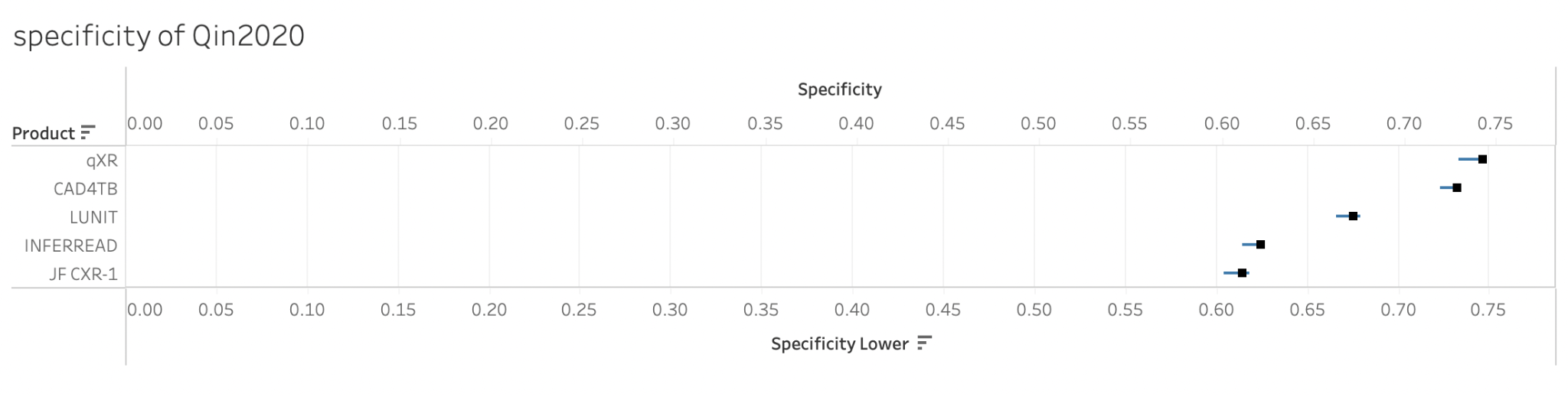


**Qin, 2020**

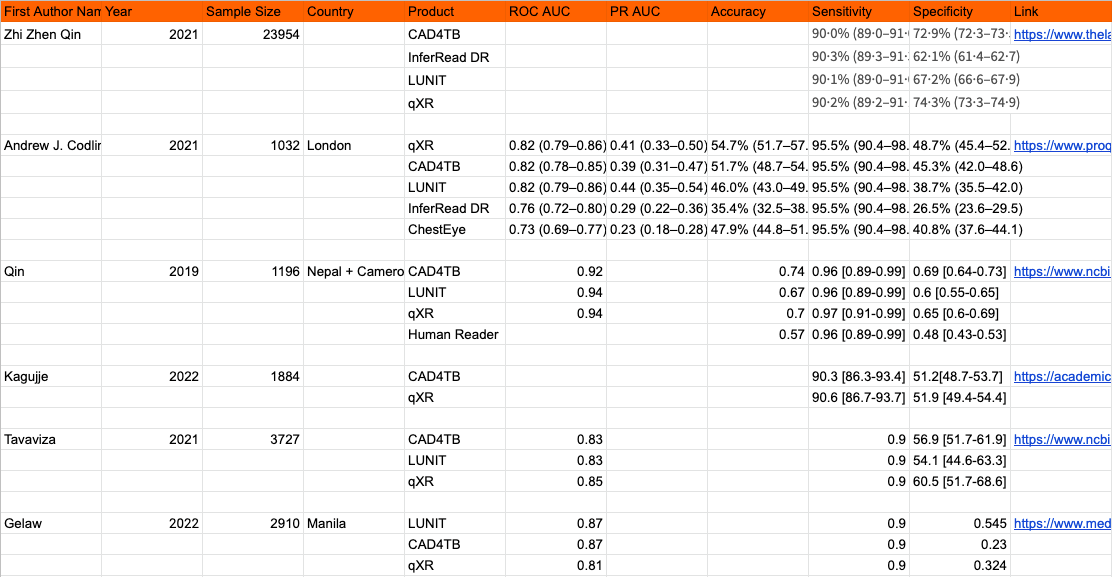
**Sensitivity with 70% specificity**

****

**Specificity with 90% sensitivity**

****

## 8.4 Technical Summary Characteristics



### 

## 8.5 qXR Systematic Review Protocol

This document is available as supplementary material in the Appendix folder and labelled as ‘Appendix Five’.

## 8.6 AI in Radiology Systematic Review Protocol

This document is available as supplementary material in the Appendix folder and labelled as ‘Appendix 6’.

## 8.7 UX Codebook

This document is available as supplementary material in the Appendix folder and labelled as ‘Appendix 7’.

## 8.8 UX Systematic Review Articles

This document is available as supplementary material in the Appendix folder and labelled as ‘Appendix 8’.

## 8.9 UX Study Characteristics

This document is available as supplementary material in the Appendix folder and labelled as ‘Appendix 9’.

## 8.10 UX Data

This document is available as supplementary material in the Appendix folder and labelled as ‘Appendix 10’.

# 9 References

Arrow. (2018, July 31). *Cloud vs In-House Servers – Which is right for your business?* Arrow Voice & Data; Arrow Voice and Data. https://arrowvoice.com.au/blog/cloud-vs-inhouse-servers/

Campbell, J., Klugar, M., Ding, S., Carmody, D., Hakonsen, S., Jadotte, Y., White, S., & Munn, Z. (2015). Diagnostic test accuracy methods for systematic review and meta-analysis. *International Journal of Evidence-Based Healthcare*, *13*(3), 154-162. 10.1097

Carmichael, J., Hollings, N., Irion, K., Reynolds, J., & van Beek, E. (2022, March 29). *Artificial intelligence for analysing chest X-ray images*. NICE. Retrieved October 23, 2022, from https://www.nice.org.uk/advice/mib292/resources/artificial-intelligence-for-analysing-chest-xray-images-pdf-2285965931918533

Chilamkurthy, S. (2018, October 29). *Challenges of Development & Validation of Deep Learning for Radiology*. Qure.ai Tech Blog. Retrieved October 4, 2022, from https://blog.qure.ai/notes/deep-learning-radiology-challenges

*Challenges and Implications of False Negative COVID-19 Testing*. (2020, August 17). UC Davis Health. Retrieved October 23, 2022, from https://health.ucdavis.edu/blog/lab-best-practice/challenges-and-implications-of-false-negative-covid-19-testing/2020/08

Chu, L. C., Anandkumar, A., Shin, H. C., &amp; Fishman, E. K. (2020). The potential dangers of Artificial Intelligence for Radiology and Radiologists. *Journal of the American College of Radiology*, 17(10), 1309–1311. https://doi.org/10.1016/j.jacr.2020.04.010

Codlin, A. J., Phuoc, D. T., Vo Luan, N. Q., Forse, R. J., Van, T. V., Dang, H. M., Nguyen, L. H., Nguyen, H. B., Nguyen, N. V., Sidney-Annerstedt Kristi, Bertie, S., Lönnroth Knut, & Maxine, C. (2021). Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis. *Scientific Reports (Nature Publisher Group), 11*(1)https://doi.org/10.1038/s41598-021-03265-0

Gelaw, S. M., Kik, S. V., Ruhwald, M., Ongarello, S., Egzertegegne, T. S., Gorbacheva, O., Gilpin, C., Marano, N., Lee, S., Phares, C. R., Medina, V., Amatya, B., & Denkinger, C. M. (2022). Diagnostic accuracy of three computer-aided detection systems for detecting pulmonary tuberculosis on chest radiography when used for screening: analysis of an international, multicenter migrants screening study. *medRxiv*. https://doi.org/10.1101/2022.03.30.22273191

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9

Healy, B., Khan, A., Metezai, H., Blyth, I., & Asad, H. (2021). The impact of false positive COVID-19 results in an area of low prevalence. *Clinical Medicine (London, England)*, *21*(1), e54–e56. https://doi.org/10.7861/clinmed.2020-0839

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial Intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500–510. https://doi.org/10.1038/s41568-018-0016-5

*How we deploy*. (n.d.). Qure AI. Retrieved October 23, 2022, from https://qure.ai/how-we-deploy/

Incepto Medical. (2022, March 24). *qXR*. Retrieved October 13, 2022, from https://incepto-medical.com/en/solutions/qxr

Kagujje, M., Kerkhoff, A. D., Nteeni, M., Dunn, I., Mateyo, K., & Muyoyeta, M. (2022). The performance of computer-aided detection digital chest X-ray reading technologies for triage of active tuberculosis among persons with a history of previous tuberculosis. *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciac679

Malamateniou, C., Knapp, K. M., Pergola, M., Woznitza, N., & Hardy, M. (2021). Artificial Intelligence in radiography: Where are we now and what does the future hold? *Radiography*, *27*. https://doi.org/10.1016/j.radi.2021.07.015

Nair, A. V., Ramanathan, S., Sathiadoss, P., Jajodia, A., & Blair Macdonald, D. (2022). Barriers to artificial intelligence implementation in radiology practice: What the radiologist needs to know. *Radiología (English Edition)*, *64*(4), 324–332. https://doi.org/10.1016/j.rxeng.2022.04.001

Office of the Australian Information Commissioner (OAIC). (2019). *Guide to Health Privacy.* Retrieved from https://www.oaic.gov.au/\_\_data/assets/pdf\_file/0011/2090/guide-to-health-privacy.pdf.

Ong, M.-S., Magrabi, F., & Coiera, E. (2018). Delay in reviewing test results prolongs hospital length of stay: a retrospective cohort study. *BMC Health Services Research*, *18*(1). https://doi.org/10.1186/s12913-018-3181-z

Privacy Act 1988 (Cth) ss 6(1), 6C, 6D(4)(b), 6FA, 6FB, 15.

*Private Cloud vs. Public Cloud: What is the difference?* (2022, October 12). VMware. https://www.vmware.com/topics/glossary/content/private-cloud-vs-public-cloud.html

Putha, P. (2017, November 25). *Generating Heatmaps to Visualize the Evidence of Deep Learning Based Diagnosis of Chest X-Rays*. Qure AI. Retrieved October 23, 2022, from https://qure.ai/evidence/generating-heatmaps-to-visualize-the-evidence-of-deep-learning-based-diagnosis-of-chest-x-rays/

*Qure.ai*. (2021, August 9). AI4HLTH. Retrieved September 30, 2022, from https://www.ai4hlth.org/product-profiles/Qure.ai

Qure.ai. (2021). *Improving Lung Health Outcomes* [Brochure]. Retrieved October 18, 2022, from https://www.qure.ai/wp-content/uploads/2021/10/Qure\_qXR\_brochure-1.pdf

*Qure AI | We ensure that our products are safe & effective*. Qure AI | AI to enable accessible, affordable & timely care across the globe. (2022). Retrieved 30 September 2022, from https://qure.ai/regulatory-and-privacy/.

*qXR- Artificial Intelligence aided Tuberculosis detection*. (2019, August 19). YouTube. Retrieved September 30, 2022, from https://www.youtube.com/watch?v=keDCOLsn1hw

Randhawa, R. (2016, October 16). *Deep Learning for Ischemic Stroke Lesion Outcome Prediction*. Qure AI. Retrieved October 23, 2022, from https://qure.ai/evidence/a-deep-learning-based-approach-for-ischemic-stroke-lesion-outcome-prediction/

*Selecting an AI marketplace for radiology: Key considerations for healthcare providers*. (2019, October 18). Imaging Technology News. https://www.itnonline.com/article/selecting-ai-marketplace-radiology-key-considerations-healthcare-providers

Seyyed-Kalantari, L., Zhang, H., McDermott, M., Chen, I., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, *27*. https://www.nature.com/articles/s41591-021-01595-0

Šimundić A. M. (2009). Measures of Diagnostic Accuracy: Basic Definitions. *EJIFCC*, *19*(4), 203–211.

Tarun, Bhargava, Manoj, Pooja, & Preetham. (2020, April 10). *Re-purposing qXR, a chest X-ray AI tool for COVID-19*. Qure AI. Retrieved October 23, 2022, from https://qure.ai/re-purposing-qxr-for-covid-19/

Tavaziva, G., Harris, M., Abidi, S. K., Geric, C., Breuninger, M., Dheda, K., Esmail, A., Muyoyeta, M., Reither, K., Majidulla, A., Khan, A. J., Campbell, J. R., David, P. M., Denkinger, C., Miller, C., Nathavitharana, R., Pai, M., Benedetti, A., & Ahmad Khan, F. (2022). Chest X-ray Analysis With Deep Learning-Based Software as a Triage Test for Pulmonary Tuberculosis: An Individual Patient Data Meta-Analysis of Diagnostic Accuracy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, *74*(8), 1390–1400. https://doi.org/10.1093/cid/ciab639

TGA. (2021). *Classification of active medical devices (including software-based medical devices)*. Retrieved from https://www.tga.gov.au/sites/default/files/classification-of-active-medical-devices-including-software-based-medical-devices.pdf

*Therapeutic Goods Act 1989* (Cth) ss 3, 9A, 41BD, 41BIB, 41CA, 41FD.

Therapeutic Goods Administration (TGA). (2022, August 22). *Manufacturer evidence for medical devices and IVD Medical Devices*. Therapeutic Goods Administration (TGA). Retrieved October 23, 2022, from https://www.tga.gov.au/how-we-regulate/supply-therapeutic-good-0/supply-medical-device/medical-device-inclusion-process/supporting-documentation-inclusion-medical-device/manufacturer-evidence-medical-devices-and-ivd-medical-devices

*Therapeutic Goods (Medical Devices) Regulations 2002* (Cth) sch1.

van Leeuwen, K. G., de Rooij, M., Schalekamp, S., van Ginneken, B., & Rutten, M. J. (2022). How does artificial intelligence in radiology improve efficiency and health outcomes? *Pediatric Radiology*, *52*, 2087–2093. https://doi.org/10.1007/s00247-021-05114-8

Qin, Z. Z., Ahmed, S., Sarker, M. S., Paul, K., Adel, A. S., Naheyan, T., Barrett, R., Banu, S., & Creswell, J. (2021). Tuberculosis detection from chest X-rays for triaging in a high tuberculosis-burden setting: An evaluation of five artificial intelligence algorithms. *The Lancet Digital Health*, *3*(9). https://doi.org/10.1016/s2589-7500(21)00116-3

Qin, Z. Z., Sander, M. S., Rai, B., Titahong, C. N., Sudrungrot, S., Laah, S. N., Adhikari, L. M., Carter, E. J., Puri, L., Codlin, A. J., & Creswell, J. (2019). Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Scientific reports*, *9*(1), 15000. https://doi.org/10.1038/s41598-019-51503-3