**Clare Rainey**

**CoRIPS Research Grant 183**

**£10,000 awarded**

# Title: An observational study to investigate the impact of AI feedback on image comment, decision switching and trust perception in student and experienced radiographers

## Principle Aim

The team aims to discover how binary diagnosis and visual feedback from an AI algorithm, specifically built for this study, will affect decision switching of radiographers at both differing stages in their training and differing level of expertise when interpreting radiographic images of the upper extremities.

The principal aim is to quantify the decision switching and trust in an AI algorithm, following exposure to AI feedback when providing preliminary clinical evaluation on radiographs of the upper appendicular skeleton:

i. to determine the baseline diagnostic accuracy of radiographers of differing levels of expertise when providing binary diagnosis on a selection of radiographic images of the upper appendicular skeleton.
ii. to expose the participants to both binary and visual feedback from an AI algorithm.
iii. to investigate the effect of this feedback on decision switching.
iv. to investigate the perceptions of trust of participants on the AI system.

## Primary research question

Does experience level determine how likely radiographers are to switch from initial binary diagnostic decision following feedback from an AI algorithm?

## Secondary research questions

Does exposure to feedback from a poorly functioning AI affect diagnostic ability of interpreters of differing levels of experience when providing binary diagnosis?

What is the perception of trust in this system of users of differing levels of experience, following exposure to the feedback?

## Outcomes

Initial sensitivity and specificity of interpreters (participants) will be determined by comparison with a 'ground truth' diagnosis, obtained from consensus of 3-5 radiological reports, on each image. Weighted Cohen's kappa will be used to quantify inter-rater reliability of participants compared with ground truth diagnoses.

Following exposure to the AI binary and visual feedback, the participants will be asked if they wish to change their initial decision.

i. The likelihood of each group to change their decision following AI feedback will be calculated
ii. The effect on accuracy (expressed as sensitivity and specificity) of any change of decision will be determined.
iii. The resulting data from above will be grouped into level of experience of the participants.

A 6-point Likert scale will be used to quantify the level of trust in AI systems following exposure to the AI feedback.

## Review of literature and identification of current gap in knowledge

Artificial intelligence (AI) as a human adjunct in diagnosing pathology on radiographic images began in the 1960s with the development of a system of conversion of image to numerical data (Lodwick et al, 1963). The rapidly increasing computational power available has permitted the development of increasingly sophisticated detection systems.

Differing methods of analysis of medical, and other, images have been developed, with Deep Learning using Convolutional Neural Networks (deep convolutional neural networks (DCNN)) being the most recent and seemingly most promising for localisation and identification of disease on radiographic images. This system architecture is now so prevalent in the computer vision field that when the term AI is used, the DCNN system architecture is assumed.

Accuracies of deep learning algorithms of up to 97% for detection of specified pathology (Islam et al. 2017, Qin et al. 2018) and up to 87% accuracy for detection of unspecified pathology on chest radiographs (Guan et al, 2018) have been reported. However, studies to date have been conducted in the laboratory setting, without testing in the clinical environment.

Recent AI developments have been the focus of much media and professional attention and more recently AI has been targeted as an area of focus for modernising and future-proofing the National Health Service (DHSC, 2018). The chief executive of NHS England, Simon Stephens has prioritised development and implementation of 'new technologies' in the 2019 Long Term Plan. Among other goals, he aims for the NHS to be a world leader in AI and machine learning within 5 years (Stephens, 2019), however, despite reported accuracies, clinicians' trust in AI remains a barrier to implementation in the health care setting. This is particularly the case with the neural network architecture. DCNNs use multiple artificial neural layers to analyse and process image data but there are a number of these layers which are hidden to the user. It is not apparent, therefore, how the AI has reached the ultimate decision. Attempts are currently being made to make this process more transparent using 'heat' or 'saliency' maps (Kumar et al, 2018). These heat maps provide a visual indication of the area of the image the AI found most important in determining its decision.

As discussed, numerous studies have demonstrated the accuracy of AI algorithms on detection of specific abnormality on other plain radiographs, in particular mammography and plain radiographs of the chest, (Islam et al. 2017, Lakhani and Sundaram, 2017, Qin et al. 2018) yet the use of AI in identification of abnormality on appendicular skeletal radiographs in the acute trauma setting remains a relatively unexplored area. This is despite plain radiographs being the most commonly performed radiographic examinations in the UK, with 23.5 million performed in NHS England per annum (NHS England, 2019) and 1.24 million MSK exams reported in Scotland 2018-19 financial year (personal communication: Reporting Radiographer Interest Group for Scottish Radiology Transformation Programme, 2019). This is assumed to be similar across the UK. Attendances at minor injuries units has increased from 28% of total Emergency Department attendances in 2008-9 to 33% in 2017-18 in England (NHS, 2018). Although official statistics are not available as to the nature of radiographic examinations undertaken in minor injuries units, it can be proposed that many of these are extremity examinations due to the specific scope of these facilities, where radiographers are in a prime position to expedite patient throughput by providing preliminary clinical evaluation (SCoR, 2013). The first publication of experimental results in the use of AI in plain skeletal radiography was in 2017 where a study compared accuracy of orthopaedic specialists and an AI on detection of fractures on selected, single projection skeletal extremity radiographs. Inter-observer reliability for fracture detection was determined and performance of the AI was found to be similar to two human experts, with an overall accuracy of 83% reported (Olczak et al, 2017). A study using AI to detect proximal humeral fractures reported accuracy of up to 95% (Chung et al, 2018), while studies on wrist radiographs also produced promising results, reporting accuracies similar to clinical experts (Buthgen et al, 2020; Lindsay et al, 2018). Many of these studies have been conducted on limited datasets of one anatomical region and usually single projection (Kim and McKinnon, 2017; Olczak, 2017;

Chung et al, 2018; Lindsay et al, 2018). No study to date has used radiographers as interpreters or to determine baseline diagnoses.

Other applications of deep learning algorithms on skeletal radiographs have been investigated and some studies suggest that AI systems also have a place in detection of osteoporosis, bone age, fracture detection and grading of osteoarthritis (Tecle et al, 2020; Badgeley et al. 2019; Brahim et al, 2019; Hirschman et al, 2019; Nehrer et al, 2019; Kim and McKinnon, 2018; Lindsay et al, 2018; Tiulpin et al, 2018; Lee et al, 2017).

As well as lack of transparency in the functioning of the AI (DCNNs), automation bias should be recognised as a potential issue when using computers to make decisions. Automation bias is defined as a human over-relying on computer information. The human believes the machine rather than their own cognitive conclusions (Goddard et al, 2012). This would not be a problem in a perfect system, but this is not reflective of real life, where myriad errors can occur in both humans and computer systems. This has been recognised as a problem in radiology and there have been studies conducted to quantify the impact of this as both omission and commission errors. In the field of computer aided detection in mammography, Philpotts (2009), recognised that the timing of the use of a computer-aided detection tool should be optimal and not be used to justify not performing additional procedure where the interpreter would otherwise recommend so. In a systematic review on the subject, Goddard et al (2011), attempts to clarify the extent of the problem and provide information on any potential mitigating factors. Hence, the rationale for this study is that that there is a paucity of information available on the effect of poorly functioning systems on the human user. This proposed study uses a poorly functioning AI and investigates the effect of this on interpreter performance.

## Methodology

An AI algorithm has been developed for use in this study which has been trained on a large, publicly available dataset of upper appendicular skeleton images. Promising results were obtained in initial testing (Cohen's Kappa 0.65) see [https://docs.google.com/document/d/1YY_U0paFvBri2puosFLCs0SYJYA7dB6ysX6TuY3GRFM/edit?usp=sharing](https://docs.google.com/document/d/1YY_U0paFvBri2puosFLCs0SYJYA7dB6ysX6TuY3GRFM/edit?usp=sharing) for examples, however, the AI performed poorly on a clinical dataset of radiographs of the same anatomical area from a different geographical region (Cohen's kappa 0.08), thus demonstrating a lack of generalisability.

<u>Clinical dataset</u>

The images used to test the AI model used in this study (the clinical dataset) were obtained on real patients presenting to a hospital in Australia and were used as part of another PhD study (McConnell, 2013). There are a total of 268 images in this dataset and have approximately a 3:7 split of pathology:no pathology. They have been used to determine diagnosis and 3-5 radiological reports from radiologists and reporting radiographers are available for each. Consensus diagnosis has been determined from this and the consensus is used as ground truth in this study.

This is an experimental study to investigate the impact AI feedback on decision switching and accuracy rates of radiographers of differing levels of experience. The data used will provide qualitative insight into how computer feedback impacts the eventual decision of the interpreter, measured as interrater reliability with a predetermined ground truth diagnosis (Cohen's kappa). Perceptions of trust in the AI system used will be investigated using a 6-point Likert scale for each participant and analysed based on their level of experience.

Method

A Qualtrics survey will be used for data collection.

1) Demographic information will be acquired from each participant to assess their level of experience as follows:

- Undergraduate student radiographers
- Qualified radiographers with less than 1-year experience,
- Qualified radiographers with greater than or equal to 1, to less than 6 years' experience,
- Qualified radiographers with greater than or equal to 6, to less than 11 years' experience,
- Qualified radiographers with greater than or equal to 11, to less than 20 years' experience,
- Qualified radiographers with greater than or equal to 20 years' experience.

2) The survey will contain 21 radiographic images from a clinical dataset of upper appendicular skeleton radiographs which have been exposed to the AI model discussed in the original application. Heat maps have been produced by the same model and represent the area the AI found most important in reaching a diagnosis. The participant will first be presented with the image without any AI diagnosis or overlay (figure 1)

3) The participants will be asked a series of questions based on the image:

    i. Do you believe there is a fracture evident on this radiographic image?

Following response to this question, the participant will then be presented with the binary diagnosis (i.e. fracture / no fracture) and the original image with a heat map (figure 2).

Further questions will then be presented:

ii. Now do you now believe there is a fracture evident on this image?

iii. Have you changed your mind from your original decision?

iv. How would you rate your trust in this AI decision from 0 to 5 based on the feedback for this image: 0 being no trust at all and 5 indicating absolute trust in the AI.

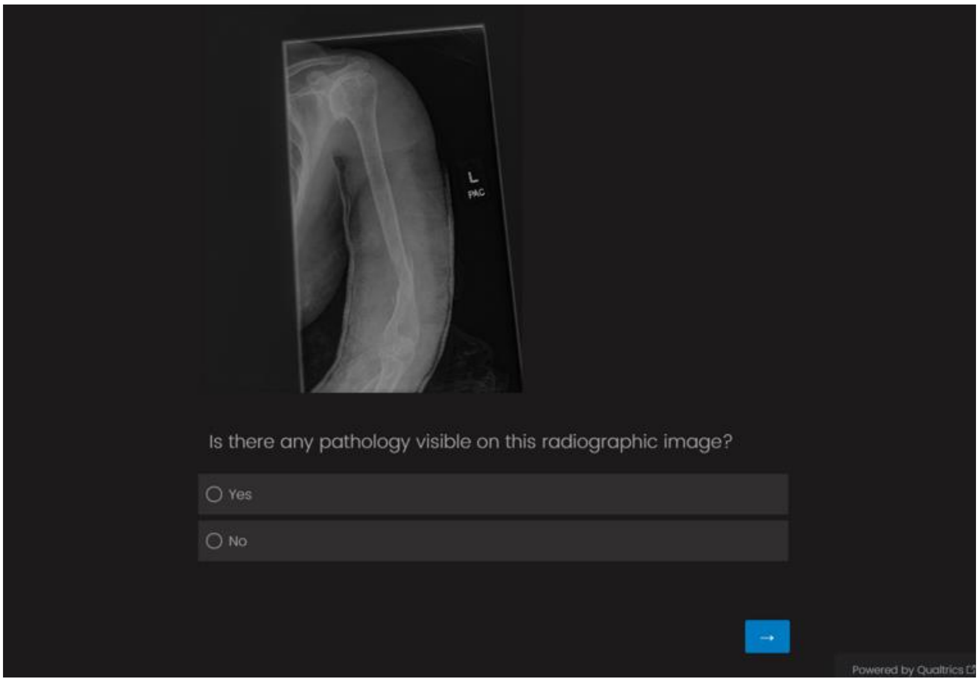


Figure 1

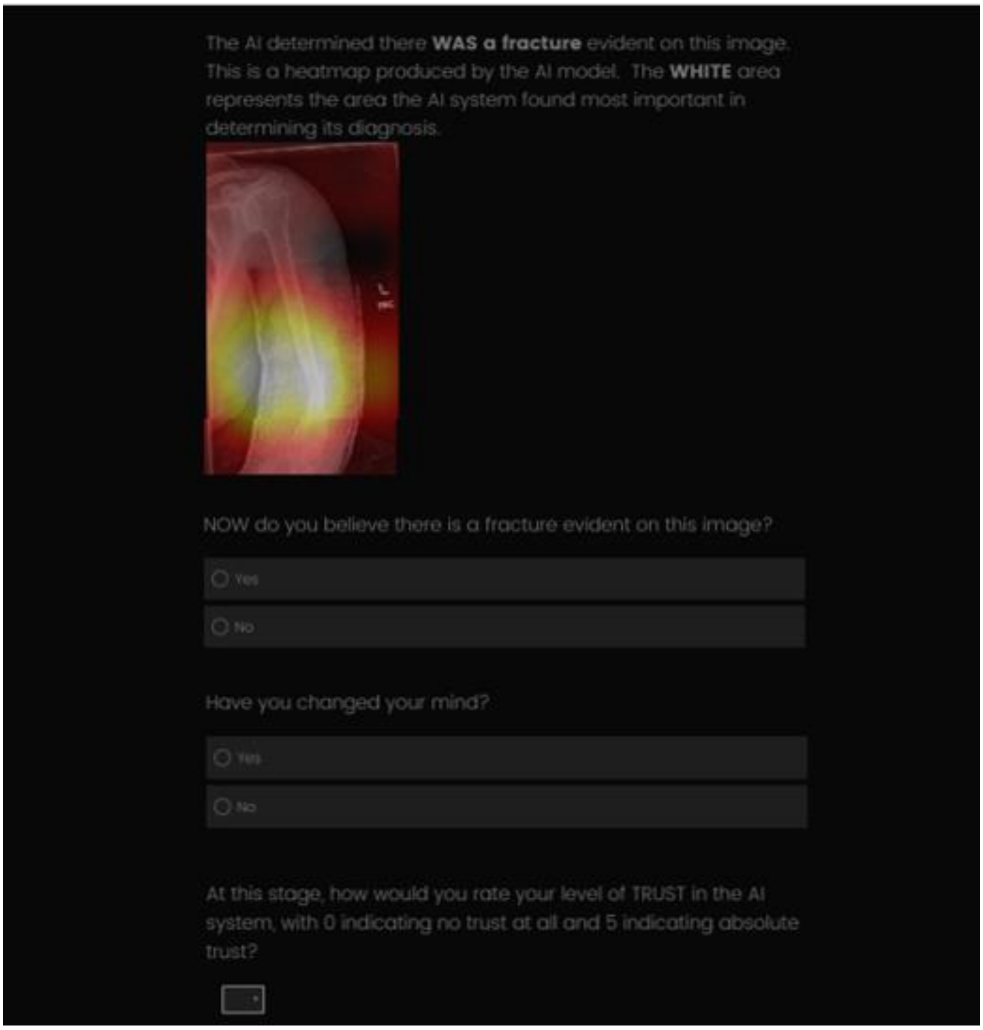


Figure 2

Sampling and Recruitment

Radiographers will be recruited via radiography-specific media, for example Synergy News, SoR online and affiliated social media platforms. An initial statement explaining implied consent will be available at the beginning of the survey. All students enrolled in BSc (Hons) Diagnostic Radiography and Imaging programmes in universities throughout the UK will be invited to participate in the study.

In order to provide adequate representation of the radiographer workforce in the UK a participant size of at least 10 per experience group (explained in full in Data Collection section) should be recruited to the study (Obuchowski, 2004). For this study, 3 participants from each geographical region, namely England, Scotland, Wales and Northern Ireland will be recruited from each of the 8 experience groups (described below) i.e. 12 per group. This equates to 96 participants in total.

Data collection

Data will be collected based on the responses from the Qualtrics survey:

i. Demographic information.
ii. Binary diagnosis for each image, before and after exposure to AI feedback will be statistically compared to the known diagnosis (as discussed in the original application) by calculation of interrater reliability using Cohen's Kappa.
iii. Effect of AI feedback (diagnosis and heatmap) on decision switch – did the AI have a positive effect (i.e. participant corrected an incorrect initial diagnosis), negative effect (i.e. participant changed initial diagnosis to an incorrect one), or neither (i.e. remained correct, or remained incorrect). A z-test will determine the statistical significance of any change for each image.
iv. The tendency of decision switch for each participant will also be calculated. The percentage decision switch for each participant across all 21 images in the survey will be calculated. This will allow assessment of the likelihood of decision switching for each experience group.
v. Perception of trust in AI system, quantifiably measured by a Likert scale of 0 – 5, described above and presented visually.

Data will be grouped according to participants' experience and analysed as discussed in the application and represented visually

Data analysis

Interrater reliability with ground truth diagnosis will be calculated at two stages: before and after exposure to the AI feedback. Ground truth is taken from consensus diagnosis provided by 3-5 experienced radiographers and radiologists. A weighted Cohen's kappa will be calculated both before and after the AI feedback to determine any change in interrater reliability as a result of the AI. A z-test will be conducted to determine if any difference in diagnosis before and after AI exposure is statistically significant. The participant response will be analysed to determine if the AI feedback made the participant opinion more or less in agreement with the ground truth diagnosis by a before and after AI feedback Cohen's kappa calculation. The % decision switch will be calculated for each participant as a mean across all images in the survey. This will be expressed as the number of switches (x) per dataset of images in the survey (i.e. 21): 100 (x/21) = % switch

Reliability/validity/credibility and trustworthiness of data

Data will be collected throughout the UK. It is anticipated that the participant selection process will allow representation from throughout the UK radiographer workforce and student population. The experience level of participants is grouped into small boundaries, thus ensuring that results are specific to experience and minimising the influence of other factors.

The AI feedback used for this study is taken from an AI developed specifically for this study. The AI was trained on a large dataset of musculoskeletal radiographs (the MURA dataset (Stanford ML group, 2018)) and is currently being used to train other AI systems. This AI performed well on the test dataset provided as part of the MURA study (Cohen's kappa:0.65) and poorly on the clinical images described above (Cohen's kappa: 0.088).

Ethical implications

The medical images in the test dataset have all been anonymised. Ethical permission has been granted previously to use the images for research purposes as part of previous PhD research (Monash University, Clayton, Australia, 2011). The images have been reported on by 3-5 qualified professionals and consensus determined. The patient management will already be completed, there is therefore no risk of impact on patient pathway or new medical revelations being made from the image interpretations. All patients' identifiable information such as the patient's name, date of birth and health and care number have been removed from the images. Images do not contain any rare abnormalities or pathologies which could readily identify an individual.

There is a potential risk of embarrassment to the participants and a fear of error. The preamble to the survey will ensure participants that any information gathered will be anonymous. They will specifically be requested to not include any personal information or information about the region of the UK where they work. The resultant data will be stored securely on encrypted, password protected external hard drives, held in Ulster University and used only for the purposes of this study.

Ethical approval will be sought from the Institute of Nursing and Health Research Filter Committee at Ulster University. It is anticipated the application will be submitted at the beginning of this academic year (September 2020).

An information sheet will be provided to potential participants, explaining the project in full, albeit in a more succinct format. The potential participants will be informed that the AI decision may not be correct although the proportion of incorrect AI diagnoses will not be made available to the participants during the study.

It will be made clear to the participants that their responses will be confidential and there will be no means of identification of the participants.

Written consent will be obtained from each participant prior to commencement of the study.

Comments from the PPI steering group, as detailed below, will be incorporated prior to submission for ethical approval, specifically, due consideration will be given to the release of the accepted diagnoses following the study to allow for participants to use their participation in the study as a learning exercise.

Patient and Public Involvement

A PPI steering group has been set up to confirm and provide insight into the appropriateness of the currently designed study, using a 'Consultation' and 'Collaboration' approach as detailed in the INVOLVE Briefing notes for researchers (pp. 21-22, National Institute for Heath Research (2012)).

Individuals known to the Principal Investigator were approached to represent the demographic spread of the study participants. Members of the group are as follows:

Mrs B. McKeag, retired reporting radiographer and Practice Educator, with over 40 years clinical experience.

Mrs K. Kissack, MRI and general radiographer, with 15 years clinical experience.

Ms J. Kennedy, undergraduate radiography student, first year (progressing to second year in September 2020).

Ms G. Doherty, undergraduate radiography student, second year (progressing to third year in September 2020).

Mr H. Kilgore, patient and county hurling player with a history of complex tarsal bone injuries requiring multiple radiological investigations and hospital admissions.

This group has been provided with an introductory letter with points for thought for comment and later discussion (appendix (i)) along with an abbreviated proposal for this study for review and comment (appendix (ii)). The points for thought and structure of this discussion was based on recommendations contained within the NHS National Institute for Health Research INVOLVE Briefing notes for researchers (2012). A 'face to face' Microsoft Teams meeting has taken place (30/07/20) to further discuss the proposal.

Based on this guidance, the meeting began with a general discussion of the project and its aims with opportunity for the group to pose questions. No queries were raised, and all members felt that the proposal was clear and understandable.

The points addressed in the initial letter were posed (appendix (i)):

All members felt the research is relevant to the professional and public communities.

No additional ethical issues were raised, although members felt that participants in the study would be able to use the study for learning if they were provided with the diagnoses at a suitable time, following completion of the study.

All members felt the suggested questions posed in the Qualtrics survey are written at an appropriate level and will be easily understood by the intended participants.

All members felt that the study design will allow for the overall aim to be achieved and that reliable and honest data will be obtained.

Members agreed with the dissemination strategy but feel that a reach to all professional audiences could be best achieved through a publication in Synergy and suggested that a CPD piece would be both interesting and informative.

At conclusion of the meeting, all members indicated their willingness to be kept informed of developments in this study.

**Potential impact**

AI is present and will be increasingly more present in healthcare moving into the future. This study aims to clarify how practitioners and student practitioners are affected by an AI system. This is particularly important as the literature is awash with potentially promising results. This study uses an AI which performed well in the laboratory/hypothetical setting but poorly with more clinically relevant images. This study will investigate if there is any difference in trust and acceptance of AI diagnoses between differing experience levels.
This study hopes to provide direction for educating undergraduate and practicing clinicians in the promise and pitfalls of integrating AI into the clinical setting.

**Dissemination Strategy**

Although focus in this study is on the UK landscape, this team believe the results of this study will have worldwide impact as the increasing visibility of AI technology is not unique to the UK alone. Worldwide dissemination is appropriate in, not only radiography conferences and publications, but medical and technology spheres also e.g. publication in Radiography and presentation at conferences such as ISRRT, UKIO and ECR

It is hoped that this study will permit publication of the findings in several different ways:

- Firstly, how an AI impacts decision switching in both the student population (inexperienced clinicians) and practicing radiographers (experienced clinicians).
- Secondly, how levels of trust/acceptance of an AI diagnosis and information varies dependent on level of expertise.

This study is important and timely as AI development is incentivised by the NHS and large manufacturers and developers race for pole position in the potentially profitable field of AI assisted diagnostic tools. Radiographers need to have in depth understanding of the potentials and pitfalls of these systems (Tucker, 2020).

## References


Blüthgen, C., Becker, A.S., Vittoria de Martini, I., Meier, A., Martini, K. and Frauenfelder, T. (2020) Detection and localization of distal radius fractures: Deep learning system versus radiologists. European Journal of Radiology, 126, 108925.

Brahim A., Jennane R., Riad R., Janvier T., Khedher L., Toumi H. and Lespessailles, E. (2019) A decision support tool for early detection of knee OsteoArthritis using X-ray imaging and machine learning: Data from the OsteoArthritis Initiative. Computerized Medical Imaging and Graphics, 73, 11-18.

Badgeley M.A., Zech, J.R, Oakden-Rayner, L., Glicksberg, B.S, Liu, M., Gale, W., McConnell M.V., Percha, B., Snyder,T.M., Dudley, J.T (2019) Deep learning predicts hip fracture using confounding patient and healthcare variables. NPJ Digital Medicine. Available at: https://www.nature.com/articles/s41746-019-0105-1.pdf [Accessed: 9th September 2019]

Brealey, S., King, D., Warnock, N. (2002) Methodological standards in radiographer plain file reading performance studies. British Journal of Radiology. 75 107-113.

Cheung, S.W., Han, S.S, Lee, J.W., Oh, K.S, Kim, N.R., Yoon, J.P. (2018) Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Radiologica, 89(4), 468-473

Department of Health and Social Care (2018) News story: Matt Hancock launches tech vision to build the most advanced health and care system in the world. Available: https://www.gov.uk/government/news/matt-hancock-launches-tech-vision-to-build-the-most-advanced-health-and-care-system-in-the-world [Accessed 5th June 2019]

Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. (2012) Journal of the American Medical Informatics Association : JAMIA. 19 (1): 121-7.

Guan Q., Huang Y., Zhong Z., Zheng Z., Zheng L., Yang Y. (2018) Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification. Available at: https://arxiv.org/pdf/1801.09927.pdf [Accessed 20th August 2019]

Hirchman A., Cyriac, J, Stieltjes, B., Kober, T., Richiardi, J., Omoumi, P. (2019) Artificial intelligence in musculoskeletal imaging: review of current literature, challenges and trends. Seminars in Musculoskeletal Radiology, 23(03), 304-311.

Islam, M.T., Aowal, M.A., Minhaz, A.T., Ashraf, K (2017) Abnormality detection and localization in chest x-rays using deep convolutional neural networks. Cornell University. Available at: https://arxiv.org/abs/1705.09850 [Accessed 15th July 2019]

Kim D.H. and MacKinnon, T. (2018) Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clinical Radiology, 73(5), 439-445.

Kumar, D., Wong, A., Taylor, G.W. (2018). Explaining the unexplained: a Class-Enhances Attentive Response (CLEAR) approach to understanding deep neural networks. Available at: https://ieeexplore.ieee.org/Xplore/home.jsp [Accessed: 10th August 2019]

Lakhani, P. , Sundaram, B. (2017) Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology, 284(2), 574-582.

Lee J.S., Adhikari S., Liu L., Jeong H.G., Kim H. and Yoon, S.J. (2018) Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study. Dento Maxillo Facial Radiology, , 20170344.

Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., Hotchkiss, R. and Potter, H. (2018) Deep neural network improves fracture detection by clinicians. Proceedings of the National Academy of Sciences of the United States of America, 115(45), 11591-11596.

Lodwick, G., Keats, T. E., , Dorst, J. P. (1963) The coding of roentgen images for computer analysis as applied to lung cancer. Radiology, 81 (2) [online] Available: https://pubs.rsna.org/doi/10.1148/81.2.185 [Accessed 15th June 2019]

McConnell, J, Devaney, C., Gordon, M. (2013) Queensland radiographer clinical descriptions of adult appendicular musculo-skeletal trauma following a condensed education programme. Radiography, 19(1), 48-55.

Nehrer S., Ljuhar R., Steindl P., Simon R., Maurer D., Ljuhar D., Bertalan Z., Dimai H.P., Goetz C. and Paixao, T. (2019) Automated Knee Osteoarthritis Assessment Increases Physicians' Agreement Rate and Accuracy: Data from the Osteoarthritis Initiative. Cartilage

'NHS News' (2019) NHS aims to be a world leader in artificial intelligence and machine learning within 5 years. Available at: https://www.england.nhs.uk/2019/06/nhs-aims-to-be-a-world-leader-in-ai-and-machine-learning-within-5-years/ [Accessed 5 June 2019]

NHS (2019) Long term plan; NHS long term plan implementation framework Available at: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/06/long-term-plan-implementation-framework-v1.pdf [Accessed 7 June 2019]

NHS (2019) Diagnostic imaging dataset statistical release. Available: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2019/08/provisional-monthly-diagnostic-imaging-dataset-statistics-2019-08-22.pdf [Accessed 15th June 2019]

Obuchowski, N.A. (2004) How Many Observers Are Needed in Clinical Studies of Medical Imaging? American Journal of Roentgenology 182 (4), 867-869

Olczak, J., Fahlberg, N., Maki, A., Ali, S.R., Jilert, A., André Stark, Olof Sköldenberg and Gordon, M. (2017) Artificial intelligence for analyzing orthopedic trauma radiographs: Deep learning algorithms—are they on par with humans for diagnosing fractures? Acta Orthopaedica, (6), 581.

Philpotts, L.E. (2009) Can computer-aided detection be detrimental to mammographic interpretation? Radiology, 253 (1), 17-22.

Qin, C., Yao, D., Shi, Y., Song, Z., (2018). Computer-aided detection in chest radiography based on artificial intelligence: a survey. BioMedical Engineering OnLine 17, 113

Society of Radiographers (2013) Preliminary Clinical Evaluation and Clinical Reporting by Radiographers: Policy and practice guidance. Available: https://www.sor.org/learning/document-library/preliminary-clinical-evaluation-and-clinical-reporting-radiographers-policy-and-practice-guidance [accessed 29th November 2019]

Stanford ML group (2018) Available: Bone X-Ray Deep Learning Competition. Available at: https://stanfordmlgroup.github.io/competitions/mura/ [Accessed 16th August 2019]

Stephens, S. (2019) Reform Health Conference: Unlocking the promise of digital health. London, 5th June 2019. Available at: https://reform.uk/events/health-conference-unlocking-promise-digital-health [Accessed 10 June 2019]

Tecle, N.(.1.)., Morris, M.R.(.1.)., Mitten, D. ( 1,2 ), Hammert, W.C.(.1.)., Teitel, J.(.2.). and Sani, N.(.2.). (2020) Convolutional Neural Network for Second Metacarpal Radiographic Osteoporosis Screening. Journal of Hand Surgery, 45(3), 175-181.

Tiulpin, A., Thevenot, J., Rahtu, E., Lehenkari, P. and Saarakkala, S. (2018) Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. Scientific Reports, 8(1), 1727.

Tucker, K (2020) The SCoR's Position on AI : InHealth Conference "Artificial Intelligence in Radiography - The Robots are coming! March 2nd 2020, online.

## Appendix (i)

Letter to PPI steering group

Dear All,

Thank you for your interest in assisting with the development of this research project. Your participation is sincerely appreciated. You have been asked to represent the study participants (i.e. qualified radiographers of various levels of experience and student radiographers) and the patient population. Your feedback and comments will ensure this study is appropriately designed to answer a research question which is relevant to the clinical, patient and public communities.

Please find the research proposal attached for your review. This study aims to investigate the impact of AI feedback on image comment, decision switching and trust perception in student and experienced radiographers. A background and rationale for this study is provided in the attached proposal.

Feedback and comment are welcomed on all aspects of the study design, for example:

- Is the research relevant to the professional and public communities?
- Are there any ethical issues not raised in the proposal which you feel should be addressed?
- Are the questions worded appropriately and at a level all participants will be able to understand?
- Will the study design allow for reliable, honest data to be obtained?
- Will the research aim be achieved through this study design?
- How might the findings be disseminated to best include any interested parties?

Following your perusal of this proposal I would like to invite you to a brief meeting to discuss any issues arising. I would propose this take place via Microsoft Teams at a date and time suitable for you. I envisage this taking no longer than 30 minutes. I will be in contact in due course to arrange this.

With renewed thanks for your participation,

Clare Rainey FHEA PgD (Adv Prac)

Lecturer in Diagnostic Radiography