

Computer-aided detection in musculoskeletal projection radiography: A systematic review

M. Gundry¹, K. Knapp¹, R. Meertens¹, J. R. Meakin²

¹University of Exeter Medical School

²College of Engineering Mathematics and Physical Sciences

Corresponding author Michael Gundry on mg361@exeter.ac.uk

University of Exeter Medical School

St Luke's

Heavitree Road

Exeter

UK

EX1 2LU

Karen Knapp email K.M.Knapp@exeter.ac.uk

University of Exeter Medical School

St Luke's

Heavitree Road

Exeter

UK

EX1 2LU

Rob Meertens email R.M.Meertens@exeter.ac.uk

University of Exeter Medical School

St Luke's

Heavitree Road

Exeter

UK

EX1 2LU

Judith R. Meakin email J.R.Meakin@exeter.ac.uk

University of Exeter

Stocker Road

Exeter

UK

EX4 4QL

Conflict of interest

Michael Gundry

Involvement in helping develop and test a form of cervical spine Computer Aided Detection (CAD) software, this was undertaken as part of my self-funded Masters by research degree at the University of Exeter, and in collaboration with City University and the Royal Devon and Exeter hospital. Additionally I was funded to help test a citizen segmentation program (Citseg), which was funded through the Wellcome trust.

Karen Knapp

Previous and current funding for the development of musculoskeletal focused CAD software from the EPSRC and Royal Devon and Exeter Foundation NHS Trust small grants scheme and Wellcome Trust Institutional Strategic Support Award (WT105618MA)) to investigate the use of citizen science in developing CAD software.

Robert Meertens

Has no conflict of interest.

Judith R Meakin

Funding from a Wellcome Trust Institutional Strategic Support Award (WT105618MA)) to investigate the use of citizen science in developing CAD software. Funding from the Royal Devon and Exeter Foundation NHS Trust small grants scheme to develop and test CAD software. Funding for a PhD studentship from the University of Exeter to develop automatic segmentation methods of MR images.

Highlights

Computer-aided detection in musculoskeletal projection radiography: A systematic review

- Systematic review of available CAD software within MSK projection radiography
- 22 papers retrieved post screening, three involved clinical testing
- The 3 papers showed an increase in sensitivity and specificity when using CAD
- CAD shows promise and variation but lack of clinical testing

Abstract

Objectives

To investigate the accuracy of computer-aided detection (CAD) software in musculoskeletal projection radiography via a systematic review.

Key findings

Following selection screening, eligible studies were assessed for bias, and had their study characteristics extracted resulting in 22 studies being included. Of these 22 three studies had tested their CAD software in a clinical setting; the first study investigated vertebral fractures, reporting a sensitivity score of 69.3% with CAD, compared to 59.8% sensitivity without CAD. The second study tested dental caries diagnosis producing a sensitivity score of 68.8% and specificity of 94.1% with CAD, compared to sensitivity of 39.3% and specificity of 96.7% without CAD. The third indicated osteoporotic cases based on CAD, resulting in 100% sensitivity and 81.3% specificity.

Conclusion

The current evidence reported shows a lack of development into the clinical testing phase; however the research does show future promise in the variation of different CAD systems.

Computer-aided detection in musculoskeletal projection radiography: A systematic review

Introduction

Even the best human observers make errors in the interpretation and classification of radiographs in making a diagnosis; be it a fracture, pathology or precursor to disease. These errors may be due to tiredness, inexperience, environmental disturbances or a combination of these [1]. As such, computers and software can potentially facilitate reducing these errors [1]. One of these facilitators is computer-aided detection (CAD), a technology designed to reduce observational oversights by using pattern recognition in order to bring attention to suspicious abnormalities within the image. CAD is designed to increase the sensitivity and specificity of a medical test [2]. CAD software has shown to increase diagnostic accuracy in many medical fields and thus helps physicians/radiologists to interpret medical images [2]. So far CAD has been integrated into some of the most common medical imaging examinations, for example:

- Mammography; improving the detection of micro calcifications [3-5].
- Chest computed tomography (CT) scans; identifying pulmonary nodules via their density and shape [6]
- CT colonography; identifying colorectal polyps [7]
- Magnetic resonance imaging (MRI): prostate cancer screening [8]
- CT cardiac scans investigating coronary artery stenosis [9]
- Nuclear medicine whole body scans; where CAD identifying bone metastases [10]
- CT spinal imaging: detecting sclerotic bone metastases, and vertebral fractures in the spine [11,12]

These CAD programs have been shown to improve diagnostic accuracy and sensitivity in these fields, and are a clinically proven technology [2-7]. However it must be acknowledged that data exist that suggest CAD systems do not statistically improve accuracy of diagnosis [13], and that CAD systems increase recall rates and reading times [14]. Although the majority of research shows positive results for CAD systems, there seems to be a lack of research regarding CAD software being used in musculoskeletal (MSK) medical imaging, this is especially important where inexperienced readers do more poorly at interpreting images in an acute trauma setting [15].

A systematic review was undertaken to investigate the use of CAD software within MSK projection radiographic imaging, compared to the reference standard of current practice or a radiologists report. In addition, the review aimed to highlight possible evidence for further research.

Methods

This systematic review was carried out according to the guidance provided by the Cochrane Collaboration with regards to systematic reviews and diagnostic test accuracy studies [16,17], whilst also utilising the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement [18].

Eligibility criteria

The participants included any patients with suspected MSK abnormalities/pathologies/injuries from any age, gender, or background. The index test/intervention was any form of CAD applied to projection radiography including dual energy x-ray absorptiometry (DXA) and orthopantomography (OPT). Exclusions included any other medical imaging modalities such as: CT, nuclear medicine, MRI, ultrasound, mammography and colon imaging. Papers which discussed the technical aspect of CAD but did not test the software were also excluded. Additionally the use of CAD had to be the primary focus of the paper to be included. The reference standard/comparator was the current practice being utilised, in most cases this was the diagnostic report created by a radiologist. The primary outcome measures were sensitivity and specificity scores, and area under a curve (AUC) differences involving CAD as defined by the study. Therefore studies which did not include these measures of diagnostic test accuracy were excluded. Secondary outcomes were differences in interpretation time and any issues or errors within the CAD system.

All relevant study designs were included with the exclusion of ideas, opinions, case studies and editorials. Only studies published after 2004 were included, due to the nature of CAD technology which is constantly evolving and improving. Thus anything prior to 2004 would be obsolete and simplified, furthermore its results would be outdated more error prone. Only articles published in English were included.

Information sources

To ensure all relevant research was identified, a wide selection of databases were searched: EMBASE, HMIC, MEDLINE (Ovid) (including Journals@Ovid full text, Your Journals@Ovid, Ovid MEDLINE corrections, Ovid MEDLINE Daily updates) Global Health, AMED, PubMed, ISI Web of Science, TRIP and Science Direct. In addition, references cited from included papers that were not retrieved utilising the search strategy but were deemed as relevant were included and subjected to the same study selection and extraction criteria.

Searching strategy

For each database a search strategy was performed, this included keyword terms, synonyms, and AND/OR qualifiers. These were grouped via their index test (e.g. “computer aided detection” OR “software aided diagnosis”) or their target condition (e.g. musculoskeletal “AND bone”). See Appendix A and Appendix B for examples of the database searches.

Study selection and data extraction

All results were extracted to EndNote (Endnote x7.0.1 Bld 7212 and Endnote x7.5 Bld 9325), and all duplicates were removed from the results pool and recorded in the PRISMA flow diagram (2.1.3 2009) as shown in Figure 1. Two independent reviewers double screened the remaining studies using the title and abstract, against the eligibility criteria. Any disagreements were debated over by the two reviewers, with a third independent reviewer arbitrating. The included papers were then screened for full text inclusion against the eligibility criteria by the same two independent reviewers. These results were again compared, and any disagreements discussed, with the third reviewer having the final decision of their eligibility. Each excluded full text was accompanied by a justification as to its exclusion (e.g. text not retrievable, not MSK, no sensitivity or specificity data). Prior to data extraction the extraction form was trialled on two of the included papers and modifications made prior to full extraction. This extraction form included data such as: title, date of publication, pathology, images/patients used, how patients/images were recruited, CAD sensitivity and specificity scores, details of the reference standard, differences in interpretation times, key conclusions, and miscellaneous comments by the author or the reviewer. A truncated version of this information is seen in Tables 1-3.

Risk of bias in individual studies

A protocol was developed and tested and modifications made prior to any data extraction, this mainly included introducing the Quality Assessment of Diagnostic Accuracy Studies

(QUADAS-2) tool [19] instead of the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) tool [20], due to QUADAS-2 being more specialised in diagnostic accuracy studies.

Prior to data extraction the QUADAS-2 signalling questions were modified to be more relevant to the CAD centred response, this tool was then employed on all eligible studies for assessing the quality and presence of bias in the included papers. It was piloted by two independent reviewers, and then applied by two independent reviewers.

Analysis

The main primary outcome measure in all studies was sensitivity and specificity, this included AUC scores. A meta-analysis was considered but due the wide range of pathologies and CAD type likely to be discovered this would lead to high heterogeneity and thus a narrative review was more appropriate.

Results

Study selection

6,253 studies were identified, 12 papers (0.19%) could not be accessed or retrieved despite several attempts and thus could not be included. Following the PRISMA flow diagram (Figure 1) primary screening resulted in 149 papers remaining for full text review, 19 of the 149 (7.84%) were meditated on by a third researcher (JM) due to disagreement between the two reviewers. A total of 24 papers were included in the final data extraction, upon extraction two of these papers were excluded, one due to being a form of literature review [21] and cited the papers mentioned within the review so did not provide any additional information, and the second [24] on closer inspection was a technical modelling paper, both papers passed the inclusion criteria but upon investigation failed to provide any new or relevant information so were excluded from the final data extraction.



PRISMA 2009 Flow Diagram

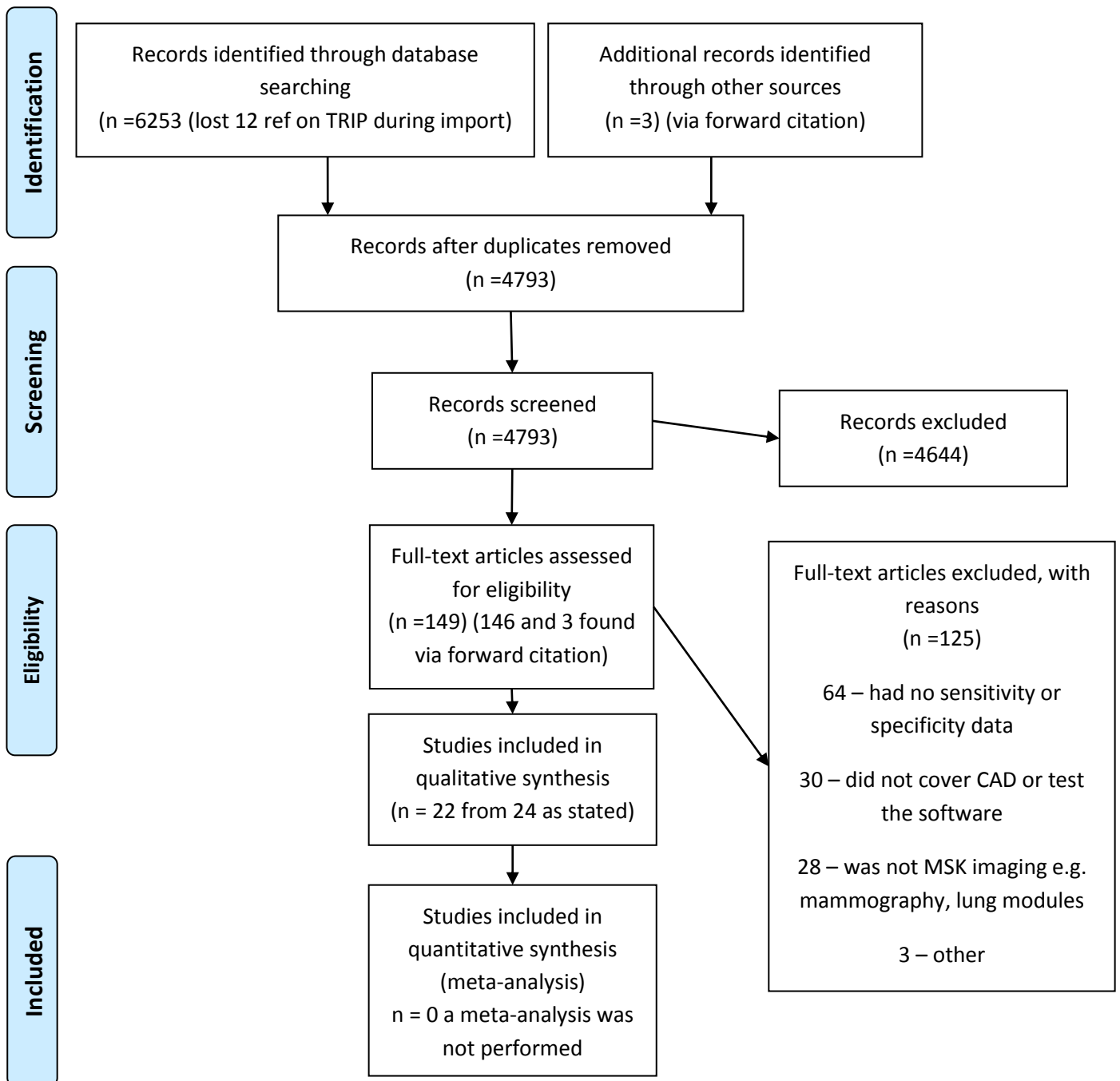


Figure 1. PRISMA flow diagram [23] showing study selection and exclusion/inclusion

Study characteristics

A condensed version of the characteristics of the included studies is divided into their different CAD pathologies is shown in Tables 1-3. Of the final 22 studies; five utilised CAD to determine vertebral fractures; of which three investigated CAD in lateral chest radiographs (all by the same author Kasai et al), and two reviewed CADs use in DXA. Nine papers

investigated the ability of CAD to determine osteoporotic risk, with eight using dental radiographs measuring mandibular cortical width (MCW) (including four papers by the same research team). Additionally two papers applied CAD to periprosthetic osteolysis (these two papers were both by Wilkie et al), and two papers investigated CAD in determining rheumatoid arthritis in the hands. Further papers included: one reviewing spine disorders and CAD, another investigated osteoarthritis and the knee, another reviewed fractures in long bones, and one utilised CAD to help diagnose dental caries. Of all 22 papers CAD was only directly tested clinically three times [24-26]. In Tracy et al [24] dentists diagnosed just from radiographs and then applied the CAD tool and made a new diagnosis. In Kasai et al [25] radiologists diagnosed first without the CAD software then with the software. The third paper by Matsumoto et al [26] conducted multiclinical trials on their CAD tool measuring MCW highlighting all suspected osteoporotic patients compared against a dental radiologist report. The rest of the papers primarily tested their CADs software ability to detect or highlight the injury or pathology on carefully selected radiographs against the highest reference standard, rather than being in clinical competition against it.

The papers have a high level of heterogeneity, this is shown in how many of the CAD systems work in different ways; certain papers measured vertebral height compared with an expected height [25,27,28], others used texture analysis [29,30] whilst some combined it with shape analysis [31] or cortical width [32]. Other types of software utilised width measurements between joint spaces; both in the hands for RA [33] and in the knees for OA [34]. Others measured cortical thickness [26] or cortical width [35-38] investigating bone mineral density (BMD) loss and osteoporosis. Other research appropriated orientation and alignment data for spine diagnosis [39] or used edge detection for long bone fracture [40]. Additionally density analysis and pattern recognition [25] have also been employed in order to determine dental caries. There was also wide variation in the countries of origin of the included studies with nine different countries across five continents. Japan and the United States of American (USA) lead the way by producing six papers each.

Additional analysis

Although the secondary outcome was investigating interpretation times, and the issues and errors of the CAD software, these were not addressed in the relevant papers, and as such are not discussed within these results.

Table 1. Study characteristics osteoporosis studies

Study	CAD system	Participant and/or image characteristics	How were participants/patients/images recruited/gathered	Reference standard/comparator (e.g. radiologists report):
Devlin et al., 2007 (UK) [35]	Defining OP risk in the hip, spine and femoral neck, using measurements of the MCW on DPRs	652 participants all women, mean age 54.9. 140 had OP at one of the three measurement sites and 65 had OP at the hip.	Participants were recruited at 4 locations Belgium, UK, Sweden and Greece for DXA scans	Patients were diagnosed with OP according to WHO criteria BMD T-score values
Kavitha et al., 2011 (Japan) [37]	OP/Low BMD measured via MCW on DPRs	100 postmenopausal women, 50 allocated to training, and 50 to its validation	A total of 531 women underwent a skeletal BMD examination at an oral radiology clinic at Hiroshima University Hospital	DXA lumbar and femoral neck WHO criteria
Kavitha et al., 2012 (Japan) [36]	OP/Low BMD measured via MCW on DPRs	100 postmenopausal women, 50 allocated to training, and 50 to its validation	A total of 531 women underwent a skeletal BMD examination at an oral radiology clinic at Hiroshima University Hospital	DXA lumbar and femoral neck WHO criteria
Kavitha et al., 2013 (Korea) [41]	OP/Low BMD measured via MCW on DPRs	100 women; 60 for training software, 40 for testing. Had to meet inclusion criteria 1) postmenopausal 2) aged 50 or greater 3) no previous diagnosis of OP	A total of 531 women underwent a skeletal BMD examination at an oral radiology clinic at Hiroshima University Hospital	DXA lumbar and femoral neck WHO criteria
Kavitha et al., 2015 (Korea) [32]	OP/Low BMD measured via MCW on DPRs	141 female patients mean age 64.3, 121 did not have OP and 20 were determined to have OP on the basis of lumbar spine and femoral neck BMD scores	Recruited from the Kyungpook National University Hospital	DXA lumbar and femoral neck WHO criteria
Matsumoto et al., 2012 (Japan) [26]	OP screening via MCW on DPR	Originally 100 cases; 73 control 27 OP training set. Then 223 cases obtained were processed. 4 cases were assessed as suspected OP by a dental radiologist	Not stated. Although the additional 223 cases were obtained in collaboration with Gifu Prefecture Dental Association	Dental radiologist and manual measurements
Muramatsu et al., 2012 (Japan) [38]	Low BMD OP diagnosed via MCW on DPR	100 images; 17 as normal volunteers. Leaving 83 clinical cases; 26 had been diagnosed with OP on the basis of DXA	100 DPRs obtained at Asahi University Hospital, Gifu, Japan.	DXA scan
Nakamoto et al., 2008 (Japan) [42]	Identifying low BMD OP via MCW on DPR	DPR from 200 post-menopausal women. 100 women aged 50-84 mean age 59.5, for software learning, and 100 for validation study 50-74 mean age 57.5	DPR were obtained from women who visited Hiroshima hospital clinic for BMD assessment	BMD measurements taken via DXA scans and using the WHO criteria of classification
Sapthagirivasan et al., 2013 (India) [43]	OP risk (low BMD) in digital hip radiographs	50 hip radiographs of south Indian women (mean age of 50.7 years) with no previous history of OP fracture, 28 used to train the CAD and 22 were used to test it	A free OP camp was organised for participants to attend	BMD measurements via DXA

Table 2. Study characteristics of vertebral fracture studies

Study	CAD system	Participant and/or image characteristics	How were participants/patients/images recruited/gathered	Reference standard/comparator (e.g. radiologists report):
Kasai et al., 2006 (USA) [27]	Vertebral fractures on lateral chest radiographs	6 radiographs (3 with grade 3 vertebral fractures and 3 normals)	From a fracture database of 1000 images, although only severe fractures (grade 3) were included	2 Radiologists independently classified the severity of the vertebral fractures using the Genant scale
Kasai et al., 2006 (USA) [28]	Vertebral fractures on lateral chest radiographs	437 male, 563 female; mean age, 76 years. 20 participants had severe vertebral fractures (grade 3), 118 without fracture. In addition, for a final validation test 32 fracture cases were used	1000 lateral chest radiographs from the Department of Radiology at the University of Chicago Hospitals	A consensus by two radiologists

Kasai et al., 2008 (USA) [25]	Vertebral fractures in lateral chest radiographs	18 radiologists were tested on 31 images (21 with vertebral fractures). Participants were all over 65 (mean age 76)	From a fracture database consisting of 1000 images although only severe fractures (grade 3) were included in testing	2 Radiologists independently classified the severity of the vertebral fractures using the Genant scale
Massari et al., 2005 (Argentina) [44]	Vertebral fractures in DXA images	362 participants; 161 with vertebral fractures and 201 without. All were Caucasia women	Recruited from a database of participants with a previous dorsolumbar radiograph, who were participating in an OP detection campaign with DXA	Specialist in bone radiology and considered the highest reference standard
Roberts et al., 2008 (UK) [31]	OP vertebral fractures in DXA scans	360 lateral DXA scans	Not stated	vertebrae were given the highest reference standard a classification using a consensus reading by 2 radiologists

Table 3. Study characteristics of other CAD studies

Study	CAD system	Participant and/or image characteristics	How were participants/patients/images recruited/gathered	Reference standard/comparator (e.g. radiologists report):
Donnelley et al., 2008 (Australia) [40]	Fractures in the long bones using edge detection	6 training images and 44 to test. All training images contained at least one fracture, while 38 of the 44 contained at least one fracture	50 long bone images obtained from the flinders medical centre emergency department	Radiologist report from flinders medical centre emergency department
Langs et al., 2007 (Austria) [45]	CAD used to detect RA erosions on hand radiographs	Radiographs of 11 left and 6 right hands were used	17 hand radiographs were gathered from patients at tertiary care rheumatology clinic	Annotation of the erosions was done by a musculoskeletal radiologist
Mandal., 2014 (India) [37]	Spine disorder diagnosis on lateral panoramic spine radiographs	310 participants (60 with herniated discs) (150 with spondylolisthesis) (100 normals)	Collected from a medical residence period in spine surgery at The Group of Applied Research in Orthopaedics	Not stated, only states the diagnosis
Oka et al., 2009 (Japan) [34]	OA in the knee on AP radiographs. CAD measures regions of the knee to determine severity	3040 participants, 5844 knee images. 2,076 knees in men and 3,768 knees in women, mean age 70.2	Images were gathered from the Research on Osteoarthritis Against Disability study	KL grading system to determine the severity of the OA
Pfeil et al., 2013 (Germany) [33]	To determine RA on hand radiographs	458 participants Caucasian 248 healthy, average age 50.7, 210 RA patients average age 55.7. 252 female, 206 male	Patients admitted to the university clinic due to trauma	Each hand radiograph was read by two radiologists for evidence of OA using the KL grading. Verified RA diagnosis according to the revised criteria of the American college of rheumatology
Tracy et al., 2011 (USA) [24]	Diagnosing dental caries using density analysis and pattern recognition LCD	12 evaluators reviewed 17 radiographs. There were a total of 159 dental surfaces on 17 radiographs with 28 confirmed dental caries	Collected during routine by attending dentist and diagnosed by him for interproximal caries visually and using LCD.	Attending dentist diagnosed visually and LCD including restoring the surfaces, tactile inspection of soft spots and follow up radiographs to confirm diagnosis
Wilkie et al., 2006 (USA) [29]	Periprosthetic osteolysis	48 total hip replacement cases. 16 osteolysis cases and 32 normal cases met certain criteria, from a database of 77	Obtained at the Anderson Orthopaedic Research Institute.	An experienced orthopaedic surgeon determined the location of osteolysis on the most recent follow-up image
Wilkie et al., 2007 (USA) [30]	Detect periprosthetic osteolysis	101 cases met specific criteria, 35 in the osteolysis group and 66 in normal group	Radiographs of THR cases obtained at Anderson orthopaedic research institute	Stated in clinical notes

Key Terms; OP Osteoporosis, OA Osteoarthritis, RA Rheumatoid Arthritis, KL Kellgren Lawrence, LCD Logicon Caries Detector, BMD Bone Mineral Density, MCW Mandibular Cortical Width, DPR Dental Panoramic Radiographs, DXA Dual energy X-ray Absorptiometry, THR Total Hip Replacement

Risk of bias within studies

There was substantial variability in the quality of studies, as shown in table 4, with some having almost no bias present [35] and other studies lacking adequate description in key areas making assessment of bias impossible [39]. The majority of the studies had bias within image selection, as most studies utilised selective radiographs specific to their determined CAD limitations and parameters. For example, in one study [25] only grade 3 vertebral fractures were included, so the more subtle grade 1 and 2 were removed. This was addressed by the paper in the quote “*To avoid the effect of bias, we believe that all cases in all grades should be included in future studies*”. Additional studies removed images that were suboptimal [35], or removed for technical reasons [44]. This selective sampling bias within the studies was also coupled with bias within the ratio selection of pathological images to healthy images. In all testing scenarios (excluding clinical) the ratio of pathological images to healthy images was high, and did not reflect the prevalence of the disease in a clinic setting, this setup most likely overestimated the accuracy of the CAD software. Additionally the majority of studies did not state if application of the index test was blinded by the reference standard result, as this may influence the result if the operator applying the index test knows the true outcome they might optimise their selection or reapply CAD to until a more accurate diagnosis is reached.

Risk of bias across studies

The 22 studies were published across 16 different academic journals, combined with such a wide range of data gathered via different search engines; this should limit the presence of publication bias. Although it must be noted that authors would probably only seek publication if they have positive results [46].

Table 4. QUADAS-2 risk of bias across different studies

Study	CAD pathology																						
	Osteoporosis								Vertebral fractures				Other										
	Devlin et al., 2007 (UK) [35]	Kavitha et al., 2011 (Japan) [37]	Kavitha et al., 2012 (Japan) [36]	Kavitha et al., 2013 (Korea) [41]	Kavitha et al., 2015 (Korea) [32]	Matsumoto et al., 2012 (Japan) [26]	Muramatsu et al., 2012 (Japan) [38]	Nakamoto et al., 2008 (Japan) [42]	Sapthagirivasan et al., 2013 (India)[43]	Kasai et al., 2006 (USA) [27]	Kasai et al., 2006 (USA) [28]	Kasai et al., 2008 (USA) [25]	Massari et al., 2005 (Argentina) [44]	Roberts et al., 2008 (UK) [31]	Donnelley et al., 2008 (Australia) [40]	Mandal., 2014 (India) [39]	Langs et al., 2007 (Austria) [45]	Oka et al., 2009 (Japan) [34]	Pfeil et al.,2013 (Germany) [33]	Tracy et al., 2011 (USA) [24]	Wilkie et al., 2006 (USA) [30]	Wilkie et al., 2007 (USA) [29]	
Could the selection of patients/images have introduced bias? RISK: LOW/HIGH/UNCLEAR																							
Is there concern that the included participants/patients/images do not match the review question? CONCERN: LOW/HIGH/UNCLEAR																							
Could the conduct or interpretation of the index test (CAD) have introduced bias? RISK: LOW /HIGH/UNCLEAR																							
Is there concern that the index test (CAD), its conduct, or interpretation differ from the review question? CONCERN: LOW /HIGH/UNCLEAR																							
Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW /HIGH/UNCLEAR																							
Is there concern that the target condition (e.g. c-spine fractures) as defined by the reference standard does not match the review question? CONCERN: LOW/HIGH/UNCLEAR																							
Could the patient flow have introduced bias? RISK: LOW /HIGH/UNCLEAR																							

Key Terms: Black block - high risk of bias, grey block - unclear risk of bias, white block - low risk of bias

Choice of reference standard

During the conduct of the review there was concern about the validity of the choice of reference standard. The majority of the research studies utilised a radiologists report or DXA scan, but some studies were more unclear as to who defined the diagnosis, either stating it was from the clinical notes [25] or not stating where the diagnosis had come from at all [39]. In other papers [30] the reference standard (osteoporosis/low BMD) was determined by a dental radiologist or compared against manual measurements and not a DXA scan; which is considered the highest reference standard for defining this particular pathology. As such this may have introduced bias into the results.

Results of individual studies

A summary of the results of the primary outcome measures of the 22 eligible studies divided into their CAD pathologies is shown in tables 5-7. The majority of the studies were compared to their highest reference standard and had sensitivity from 40% [34] to 100% [26] and specificity from 1.4% [40] to 100% [26]. These wide ranges were in part due to being early versions in developmental systems. Results show:

- Vertebral fracture (five papers) sensitivity, specificity, and AUC ranged from respectively; sensitivity 62.5% [44] to 95% [39], specificity 92.1% [44] to 96.8% [44], AUC 0.951 [25]
- Osteoporotic risk (nine papers) sensitivity, specificity, and AUC ranged from respectively; sensitivity 74.4% [42] to 100% [26], specificity 43.8% [42] to 100% [26] AUC=0.759 [35] to AUC=0.97 [26]
- RA in the hands (two papers) sensitivity, specificity, and AUC ranged from respectively; sensitivity 61.4% [33] to 88.1% [33], specificity 77.8% [33] to 88.7% [33], AUC 0.878 [33] to 0.920 [45]
- Periprosthetic osteolysis (two papers) only AUC recorded: AUC 0.677 [29] to 0.88 [42]
- Other (four papers made up of: long bone fractures, OA in the knee, spinal disorders and dental caries) sensitivity, specificity, and AUC ranged from respectively; sensitivity 40% [12] to 96.6% [39], specificity 1.4% [40] to 98.7% [39], AUC 0.998 [39] 0.639 [12]

The lowest sensitivity and specificity figures come from the other group criteria; the 40% sensitivity score was only one result of many different CAD techniques used to define the severity of OA in the knee. This severity was measured via the Kellgren Lawrence score using the osteophyte area, in addition, the Oka et al study [34] also used the measurements of the medial minimum joint space width, and tibiofemoral angle which both scored higher sensitivity scores (42.7% and 58.3 % respectively), the AUC of this low sensitivity score was 0.739 due to a specificity score of 79%. The specificity score 1.4% [40] was due to overcalling on long bone fractures. The two highest scores in sensitivity and specificity are 100% in each in screening of osteoporosis [26], although these scores were independent with one during training and one during clinical testing.

Regarding the three clinically tested papers: Matsumoto et al [26] screened for osteoporosis via dental imaging and tested 223 cases reporting a sensitivity of 100% and a specificity of 81.3%. Kasai et al [25] tested 18 radiologists on 31 images (21 with vertebral fractures); radiologists scored an average sensitivity of 59.8% (226/378) without CAD, compared to an average sensitivity of 69.3% (262/378) with CAD, an AUC change of 0.906 to 0.951. The third paper by Tracy et al [24] tested 12 dentists to evaluate 17 digital radiographs (a total of 159 surfaces with 28 confirmed dental caries) compared to a Logicon caries detector (LCD) CAD tool. Sensitivity with the LCD CAD tool was 68.8% (CI 61.8-75.7) and specificity was 94.1% (CI 92-96.2), compared to the dentists highest sensitivity result when the image was sharpened of 39.3% (CI 30.2-48.4) (with a specificity of 93.1% (CI 88.2-97.9)), and highest specificity with the initial image of 96.7% (CI 93.6-99.8) (with a sensitivity score of 30.4% (CI 21.6-39.1)).

Table 5. Study results of the osteoporosis CAD studies

Study	CAD	comparator	Key conclusions of the study authors:
Devlin et al., 2007 (UK) [35]	For the automatically initialized searches at any site (femoral neck, lumbar or spine) AUC=0.759 (95% CI=0.724 to 0.791). For the automatically initialized searches for just the femoral neck AUC=0.805 (CI=0.773 to 0.835)	MCW was measured using the manually initialized (semi-automatic) method, gave an AUC =0.816 (CI=0.784 to 0.845) at any site. Manually initialized fit gave an AUC =0.835 (CI=0.805 to 0.863) when just using the femoral neck	We concluded that measurement of MCW using active shape modelling is capable of diagnosing skeletal OP with good diagnostic ability and repeatability
Kavitha et al., 2011 (Japan) [37]	Sens lumbar spine 90% (CI 81.7-98.3) spec 75% (CI 63-87), femoral neck sens 81.8% (CI 70.1-91.8) spec 69.2% (CI 56.2-81.8). Validation results lumbar spine sens 93.3% (CI 85.9-100) spec 82.9% (CI 71.4 -92.7), femoral neck sens 92.3% (CI 85.9-99.5), spec 75.7% (CI 63-87)	DXA lumbar and femoral neck WHO criteria	Our new CAD system is a useful tool in screening for OP. Additional studies with a large number of postmenopausal women would be necessary to overcome this system limitation
Kavitha et al., 2012 (Japan) [36]	Sens score of 90.5% (CI 83.8-95.1) and spec 70.9% (CI 61.4-79.0). The corresponding values in the validation subjects were sens 92.9% (CI 86.2-96.5) and spec 77.8% (CI 68.9-84.9), respectively	DXA lumbar and femoral neck WHO criteria	The sensitivity of our current CAD system was almost the same, but the specificity was much higher
Kavitha et al., 2013 (Korea) [41]	The sens and spec using the HAC-SVM model were 95.8% (CI 91.9-99.7) and 86.6% (CI 79.9-93.3), respectively, at the lumbar spine; and 96.0% (CI 92.2-99.8) and 84.0% (CI 76.8-91.2), respectively, at the femoral neck. AUC of 0.886 (95% CI: 0.816-0.944) for femoral neck and AUC of 0.871 for the lumbar spine (95% CI: 0.804-0.936). Another piece of software called BP neural network, reported lumbar sens 93.3% (CI 88-98), spec 83.2% (CI 75.6-90.4), femoral neck sens 93.8% (CI 89-98) spec 82% (CI 74.5-89.5)	DXA lumbar and femoral neck WHO criteria	Our experimental results predict that the proposed HAC-SVM model combination applied on DPRs could be useful to assist dentists in early diagnosis and help to reduce the morbidity and mortality associated with low BMD and osteoporosis
Kavitha et al., 2015 (Korea) [32]	The best results for both the lumbar and femoral BMD scores were MCW combined with FD, resulting in lumbar AUC 0.922 sens 94% spec 82.8%. Femoral AUC of 0.947, sens 96.1% spec 84.7%	DXA lumbar and femoral neck WHO criteria	Our findings suggest that a combination of mandibular cortical bone textural features and MCW yields an improved assessment of OP or low BMD compared with the use of individual textural features or MCW
Matsumoto et al., 2012 (Japan) [26]	Sens and spec for identifying OP patients were 92.6% (25 out of 27) and 100% (73 out of 73), respectively. AUC 0.97. For the 223 cases including 4 fractures, the sens was 100% (4/4 cases), and the spec was 81.3% (178/219 cases)	dental radiologist, and manual measurements AUC scored 0.98	The result of clinical trials indicates that our proposed scheme may have a potential to identify OP patients at an early stage
Muramatsu et al., 2012 (Japan) [38]	The sens and spec for identifying OP patients were 88.5% and 97.3%, respectively	DXA	An automated MCW measurement technique is feasible using DPRs, and this method has a potential to identify asymptomatic OP patients
Nakamoto et al., 2008 (Japan) [42]	The sens and spec for identifying women with low skeletal BMD were 74.4% (CI 64.7-84) and 54.5% (CI 33.7-75.3) in the development group, and 76.8% (CI 67.7-86) and 61.1% (CI 38.6-83.6) in the validation group (for T-scores less than -1.0). In addition, the respective sens and spec were 96.8% and 44.9% in the development group, and 94.4% and 43.8% in the validation group (for T-scores that are less than -2.5).	BMD measurements taken via DXA scans and using the WHO criteria of classification	Post-menopausal women with undetected OP may be identified by CAD with sufficient performance and reproducibility in comparison with the questionnaire- based screening tools used worldwide.
Sapthagirivasan et al., 2013 (India) [43]	90% sens (95% CI of 82-98%) spec of 87% (95% CI of 78-96%). AUC 86.3% (95% CI of 79.3-93.5)	CAD compared to the highest reference standard the DXA scan	Findings suggest that the proposed CAD system would be useful for spotting women vulnerable to OP risk

Table 6. Study results of the vertebral fracture CAD studies

Study	CAD	comparator	Key conclusions of the study authors:
Kasai et al., 2006 (USA) [27]	3 fractures were indicated out of 4 fractures (75% sens), including one FP. Additionally all 3 normals were also classified correctly	Two radiologists independently classified the severity of the vertebral fractures using the Genant scale	Although the number of cases used in our study is very small, we believe that the preliminary results are encouraging
Kasai et al., 2006 (USA) [28]	The sens for detection of fracture cases was 95% (19/20), with 1.03 139/135 FP fractures per image. 32 additional fracture cases were used in a validation test. The sens for these cases was 75% (24/32) at 1.03 33/32 FP fractures per image	A consensus by two radiologists.	The preliminary results indicate that our scheme would be useful in assisting radiologists in the detection of vertebral fractures, and thus providing an early diagnosis of OP
Kasai et al., 2008 (USA) [25]	18 radiologists with CAD. Sens was 81% and a FP rate of 0.78 per case. Specs was not stated. AUC was 0.951	18 radiologists without CAD AUC score was 0.906.	The use of CAD with lateral chest radiographs can improve radiologists' image interpretation in the detection of vertebral fractures
Massari et al., 2005 (Argentina) [44]	The CAD program showed a sens of 62.5% and spec of 92.1%. When divided into regions the lumbar spec for fracture detection increased to 96.8%	Specialist in bone radiology and considered the highest reference standard, reviewing the images using the Genant scale	This method could be considered as a screening tool to discard vertebral fractures
Roberts et al., 2008 (UK) [31]	Automatic segmentations the appearance classifier sensitivity was 86% at 5% FP rate	A consensus reading by 2 radiologists using the algorithmically based qualitative method	Reasonable sensitivity can be achieved using an automatic segmentation, but occasional segmentation failures would require manual correction

Table 7. Study results of other CAD studies

Study	CAD	comparator	Key conclusions of the study authors:
Donnelley et al., 2008 (Australia) [40]	Sens is 83% spec is 1.4%	Not stated, although the images were gathered from flinders medical centre emergency department	These results will further expand the capabilities of today's CAD systems, and result in more accurate diagnosis of fractures and a reduction of the fracture miss rate
Langs et al., 2007 (Austria) [45]	Sens 85% spec 84% (FP rate of 16% FN rate of 15%) AUC of 0.92	Annotation of the erosions was done by a MSK radiologist	The automatic spotting of erosions provides promising results and the visualisation of the deviation from health anatomy aids clinicians in the evaluation of the erosions and in the reviewing of automatic detection results
Mandal., 2014 (India) [39]	Seven different methods /algorithms were used. Results of the seven 1) sens 0.948 spec 0.863 ROC 0.982 2) sens 0.914 spec 0.793 ROC 0.966 3) sens 0.948 spec 0.987 ROC 0.98 4) sens 0.9310 spec 0.920 ROC 0.926 5) sens 0.966 spec 0.868 ROC 0.917 6) sens 0.966 spec 0.929 ROC 0.998 7) sens 0.948 spec 0.863 ROC 0.966	Presumed standard clinical diagnosis but this is not stated	The primary aim of the work is to upgrade the software reliability of the spine diagnosis and reduce the rate of misdiagnosis and the obtained test and experimental results indicate that the goal is reached.
Oka et al., 2009 (Japan) [34]	The CAD program measured OA parameters. The AUC, sens, spec of medial minimum joint space width for KL \geq 2 and KL \geq 3 OA was 0.739 and 0.856, 58.3% and 76.2%, 79.0% and 83.9%, respectively, and those of the osteophyte area were 0.663 and 0.767, 40% and 63.3%, 92.2% and 85.8% and tibiofemoral angle were 0.639 and 0.729, 42.7% and 54.9%, 79% and 84.8%, respectively	KL grading system to determine the severity of the OA	This system is useful as an objective and accurate method for measurement of the structural severity of the knee, and can be a surrogate measure for the development of disease-modifying drugs for OA, just as BMD is in OP

Pfeil et al., 2013 (Germany) [33]	total JSD MCP sens 88.1% and spec 77.8%, AUC 0.920 (CI 0.896-0.943) total JSD PIP sens 61.4% and spec 88.7%, AUC 0.878 (CI 0.846-0.910)	Read by two radiologists for evidence of OA using the KL grading system, verified RA diagnosed according to the revised criteria of the American college of rheumatology	CAJSA method presented a reliable assessment of disease related joint space narrowing suffering from RA
Tracy et al., 2011 (USA) [24]	Sens was 68.8% (CI 61.8-75.7) spec was 94.1% (CI 92-96.2) when the LCD CAD dentist analysis tool was used	Sens among all 12 blinded evaluator dentists was 30.4% (CI 21.6-39.1) with the initial image, 34.2% (CI 26.4-42.1) with brightness and contrast adjusted, 39.3% (CI 30.2-48.4) when the image was sharpened. Spec was found to be 96.7% (CI 93.6-99.8) with the initial image 95.3% (CI 90.9-99.8) with the brightness and contrast adjusted and 93.1% (CI 88.2-97.9) with the sharpened image	Compared to the unaided eye the LCD can significantly improve dentists' ability to detect and classify caries.
Wilkie et al., 2006 (USA) [30]	3 CAD methods: linear regression AUC max was 0.68, LDA max was 0.78, BANN max was 0.88	CAD compared to an experienced orthopaedic surgeon who determined the location of osteolysis	The preliminary analysis in this research was based over a small database, and results cannot be considered conclusive
Wilkie et al., 2007 (USA) [29]	Different methods of rRTA were tested. For the FMP LDA method the AUC was 0.677, FMP BANN method 0.695, for the FMP angular LDA was AUC 0.798, and AUC of 0.780 for FMP angular BANN method	Stated diagnosis from clinical notes	tRTA has shown promise as a method to help detect periprosthetic osteolysis. One of the biggest drawbacks is that many cases do not have complete data available from reasonable follow-up intervals and must be excluded from the analysis

Key terms:

CAD Computer Aided Detection, AUC Area Under the Curve, DXA Dual energy X-ray Absorptiometry, CI Confidence Intervals, WHO World Health Organisation, ROC Receiver Operating Characteristic, MSK Musculoskeletal, JSD MCP Joint Space Width Metacarpal Phalangeal, JSD PIP Joint Space Width Proximal Interphalangeal, CAJSA Computer Aided Joint Space Analysis, RA Rheumatoid Arthritis, OP Osteoporosis, BMD Bone Mineral Density, DPR Dental Panoramic Radiographs, HAC-SVM Histogram based Automatic Clustering Support Vector Machine, BP Back Propagation, FP False Positive, FN False Negative, MCW Mandibular Cortical Width, FD Fractal Dimensions, Sens Sensitivity, Spec Specificity, KL Kellgren Lawrence, OA Osteoarthritis, LCD Logicon Caries Detector rRTA temporal Radiographic Texture Analysis, BANN Bayesian Artificial Neural Network, LDA Linear Discrimination Analysis, FMP First Moment of the Power spectrum

Discussion

Summary of evidence

The primary findings of this systematic review suggests that CAD is in its early incarnation within planar MSK imaging, as such clinical testing has not been well established having only been attempted in three small studies.

The results of the three clinically tested papers had a small numbers of images; Matsumoto et al [26] tested 223 cases, but only four of those were positive as osteoporotic cases, so the reported sensitivity score of 100% could easily be misinterpreted or more likely an inflated figure. Kasai et al study [25] involved a small number of images (31), and only included grade 3 fractures; which again could possibly be misinterpreted in favour of the CAD system which might not be able to distinguish the more subtle grade 1 or grade 2 fractures. The third paper by Tracy et al [24] utilised dentists in their study. These dentists were reported to have limited knowledge of digital radiography, having come from an area that had not integrated it into their practice (the testing involved reviewing digital radiographs). This is in conjunction with the knowledge that none of the dentists had the opportunity to do an oral examination or review the patients' history both of these would clearly affect the sensitivity and specificity results, and thus the CAD tools generalizability and real world application.

Generally the studies identified focused on the importance of the sensitivity result. If this sensitivity rate is low then a patient with an injury/disease might be allowed to be discharged due to a test incorrectly revealing the absence of the injury/disease. However most of these papers focus on the sensitivity improvements at the expense of specificity reductions, mainly due to overcalling of CAD software on normal variants. This is most likely due to the implications of the thresholds used within the CAD systems, were a normal variation is just above the density or morphology threshold to be defined as pathology, this results in a higher sensitivity whilst simultaneously decreasing the specificity. This is especially evident in the Donnelley et al 2008 paper [40] which reports a sensitivity of 83% and a specificity of 1.4%, where 55.9% of the reported causes of false detection were biological features not related to the fracture [40].

With sensitivity being the primary interest it is not surprising that most of the CAD developers have targeted a screening type CAD system, alongside the fact CAD typically only assess one specific pathology.

This is reflected in the diverse types of CAD software utilised, although as shown in tables 1-3 and 5-7, nine papers investigated screening for osteoporosis, and five investigated vertebral fractures. In the osteoporosis studies there is a similar trend in investigation, with eight of the nine studies using MCW to screen for osteoporosis, with the ninth using modelling of the hip. Of the eight papers the lowest sensitivity and specificity scores are in the oldest paper from 2008 [42], with further published studies showing gradual improvements with the latest 2015 paper reporting a sensitivity of 94% and 96.1% [32] and specificity of 82.2 and 84.7% [32] for the lumbar and femoral BMD scores respectively when compared to DXA. In the five studies investigating vertebral fracture the oldest paper [44] had the poorest sensitivity, a score of 62%, this compared to the latest studies reporting sensitivity scores of 81% [25] and 86% [31].

The majority of the research shows CAD has promise in distinguishing certain pathologies, and improves through each incarnation. This is elaborated on by the research by Kasai et al [25], who over two years and three papers showed increasing results to the point of clinical testing. Data by Kavitha et al also supports this argument of improvement, showing a gradual increase in accuracy of their CAD system between the four papers covering 2011-2013. Although this idea must be approached with caution, as if this was the case for all papers, there should be a larger amount of follow-up research from other studies that showed similar original promise, such as Kasai et al's clinical testing in 2008 which has produced no further publications. However this lack of continuation and evolution of these systems might be due to the researchers taking other avenues of investigation, or the possibility that particular CAD system did not work when clinically tested, and thus no further papers were published, which may in itself be due to high levels of CAD complexity. Thus this systematic review results might be an exaggeration of CADs success.

Limitations of this review

A comprehensive and systematic approach was used, with at least two researchers screening the studies throughout. The loss of 12 research papers during the import of the original 6253 may have introduced bias into this study, as well as any studies registered under a different name other than CAD or its derivatives. Additionally the modification to the search criteria to remove other modalities from the results might have introduced bias. Furthermore the two papers that made it through the inclusion criteria but not in the data extraction shows weakness in the search criteria design. In future the exclusion criteria should be more stringent and include the exclusion of literature reviews if they cover all the

papers under investigation, and modelling papers as well as technical papers should be excluded unless the technical aspects can be understood clearly by the researcher. It is also acknowledged an author with a computer science background could have helped provide a better understanding of some of the technical aspects of the CAD systems identified.

Conclusion

Implications for clinical practice

There is potential for CAD to be instigated into MSK imaging especially in osteoporosis and to a smaller degree in vertebral fracture imaging, with its primary facility being as a screening tool. This has been indicated in the studies gathered from this review with nine papers investigating CADs application in osteoporosis and five papers reviewing its application in vertebral fractures. The main two studies covering these two topics show CAD implicating osteoporosis via MCW measurements or being used to improve results in vertebral fracture diagnosis. Although caution must be applied as the papers published lacked data on interpretation times and possible CAD system errors. Also it must be acknowledged that the CAD tested within the studies is not a fair reflection of a true ratio of pathological to non-pathological images this is particularly evident in the Kasai et al [25] study utilising 31 images 21 of which had vertebral fractures. Thus due to CAD still being in its infancy and the current lack of follow-up and clinical testing its impact for clinical practice is currently extremely limited.

Implications for future research

As stated the investigation into osteoporosis and vertebral fracture make up over 63% of the papers within this review and with the advent of more digital systems becoming advanced and automated, and resources becoming redistributed CADs intertwinement in medical imaging will continue to increase. Future research needs to be undertaken in a clinical setting, and avoiding the common trends of selection biases and high pathology rates that exists in current literature. Until there is a stronger foundation of evidence with methodologically sound studies that begins to challenge the highest reference standard, the best current diagnostic practice will always be seen as the greater diagnostic tool.

Appendices

Appendix A. Example of First search across EMBASE, HMIC, MEDLINE (ovid) (including Journals@ovid full text, Your Journals@ovid, ovid MEDLINE corrections, Ovid MEDLINE daily updates) Global Health

The screenshot displays the OvidSP search results page. At the top, there are navigation tabs for 'Search', 'Journals', 'My Workspace', and 'Primal Pictures'. Below this, a search bar and various tool options like 'Print', 'Email', and 'Export' are visible. The main content area shows two search results:

- Result 232:** "Trabecular morphometry by fractal signature analysis is a novel marker of osteoarthritis progression." by Kraus, Virginia Byers 1; Feng, Sheng 1; Wang, ShengChu 1; White, Scott 1; Ainslie, Maureen 1; Brett, Alan 2; Holmes, Anthony 2; Charles, Cecil H. 1. Published in *Arthritis & Rheumatism*, 60(12):3711-3722, December 2009. AN: 00000889-200912000-00023.
- Result 233:** "Comparison Study of Different Bowel Preparation Regimens and Different Fecal-Tagging Agents on Tagging Efficacy, Patients' Compliance, and Diagnostic Performance of Computed Tomographic Colonography: Preliminary Study." by Yoon, Soon Ho MD *; Kim, Se Hyung MD, PhD *; Kim, Sang Gyun MD, PhD *; Kim, Soo Jin MD, PhD *; Lee, Jeong Min MD, PhD *; Lee, Jae Young MD, PhD *; Han, Joon Koo MD, PhD *; Choi, Byung Ihn MD, PhD *. Published in *Journal of Computer Assisted Tomography*, 33(5):657-665, September/October 2009.

Each result includes a 'View Abstract' button and a 'Buy Now' option. The interface also features a 'Search Information' sidebar on the left with filters for search terms and returned results.

Appendix B. Example of ScienceDirect database search

The screenshot shows the ScienceDirect search results page. The search criteria are: "computer assisted detection" OR "computer aided detection" OR "computer aided diagnosis" OR "software aided detection" OR "software assisted detection" OR "software assisted diagnosis" AND bone OR muscle OR musculoskeletal OR vertebra. The results are filtered by year (2015-2004) and publication title. The main content area displays several search results:

- Result 1:** "Évaluation de l'impact d'un système CAD sur la performance des radiologues pour la détection des nodules pulmonaires sur des examens scanographiques multicoups de thorax." by B. Brochu, C. Beigelman-Aubry, J.-L. Goldpard, P. Raffy, P.A. Grenier, O. Lucidarme. Published in *Journal de Radiologie*, Volume 88, Issue 4, April 2007, Pages 573-578.
- Result 2:** "16 - Computer-Aided Diagnosis." by Maryellen L. Giger, Kenji Suzuki. Published in *Biomedical Information Technology*, 2008, Pages 359-374, X001.
- Result 3:** "3 - Mass Detection Scheme for Digitized Mammography." by Bin Zheng. Published in *Cancer Imaging*, Volume 1, 2008, Pages 325-337.
- Result 4:** "Informatique et imagerie médicale : les tendances décelables à Chicago en 2004." by J. Chabrias. Published in *Journal de Radiologie*, Volume 86, Issue 7-8, August 2003, Pages 864-867.
- Result 5:** "7 - Lung Cancer: Computer-aided Diagnosis with Computed Tomography." by Yusalygale Abe, Katsuna Tamura, Rieko Sakata, Jiro Ishida, Masayoshi Nagata, Masato Nakamura, Kikuo Machida, Toshiro Ogata. Published in *Cancer Imaging*, Volume 1, 2008, Pages 209-214.
- Result 6:** "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review." by Afaneh Jalalian, Syamsiah B. T. Mashohor, Hajah Rozi Mahmud, M. Iqbal B. Saripan, Abdul Rahman B. Ramli, Babak Karasi. Published in *Cancer Imaging*, Volume 3, Part 2, April 2009, Pages 7242-7251.
- Result 7:** "Computer aided diagnosis system of medical images using incremental learning method." by M. Park, B. Kang, S. J. Jin, S. Luo. Published in *Expert Systems with Applications*, Volume 36, Issue 3, Part 2, April 2009, Pages 7242-7251.

The interface includes a 'Refine filters' sidebar on the left and a 'Purchase' button for each result. An advertisement for Elsevier Exhibitions Survey is visible on the right side of the page.

References

1. Erickson BJ, Bartholmai B. Computer-aided detection and diagnosis at the start of the third millennium. *Journal Digit Imaging*. 2002;15(2):59-68.
2. Castellino RA. Computer aided detection (CAD): an overview. *Cancer Imaging*. 2005;5:17-9.
3. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*. 2001;220(3):781-6.
4. Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. *AJR Am J Roentgenol*. 2005;185(4):944-50.
5. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: interpretation with computer-aided detection--prospective evaluation. *Radiology*. 2006;239(2):375-83.
6. Beigelman-Aubry C, Raffy P, Yang W, Castellino RA, Grenier PA. Computer-aided detection of solid lung nodules on follow-up MDCT screening: evaluation of detection, tracking, and reading time. *AJR Am J Roentgenol*. 2007;189(4):948-55.
7. Petrick N, Haider M, Summers RM, Yeshwant SC, Brown L, Iuliano EM, et al. CT colonography with computer-aided detection as a second reader: observer performance study. *Radiology*. 2008;246(1):148-56.
8. Wang S, Burt K, Turkbey B, Choyke P, Summers RM. Computer aided-diagnosis of prostate cancer on multiparametric MRI: a technical review of current research. *BioMed Res Int*. 2014;2014:789561.
9. Rief M, Kranz A, Hartmann L, Roehle R, Laule M, Dewey M. Computer-aided CT coronary artery stenosis detection: comparison with human reading and quantitative coronary angiography. *Int J Cardiovasc Imaging*. 2014;30(8):1621-7.

10. Sadik M, Suurkula M, Höglund P, Järund A, Edenbrandt L. Improved classifications of planar whole-body bone scans using a computer-assisted diagnosis system: a multicenter, multiple-reader, multiple-case study. *Journal of nuclear medicine: official publication, J Nucl Med.* 2009;50(3):368-75.
11. O'Connor SD, Yao J, Summers RM. Lytic metastases in thoracolumbar spine: computer-aided detection at CT—preliminary study. *Radiology.* 2007;242(3):811-6.
12. Burns JE, Yao J, Muñoz H, Summers RM. Automated detection, localization, and classification of traumatic vertebral body fractures in the thoracic and lumbar spine at CT. *Radiology.* 2016;278(1):64-73
13. Cole EB, Zhang Z, Marques HS, Edward Hendrick R, Yaffe MJ, Pisano ED. Impact of computer-aided detection systems on radiologist accuracy with digital mammography. *AJR Am J Roentgenol.* 2014;203(4):909-16.
14. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation?. *Radiology.* 2009;253(1):17-22.
15. Hlongwane ST, Pitcher RD. Accuracy of after-hour 'red dot' trauma radiograph triage by radiographers in a south african regional hospital. *S Afr Med J.* 2013;103(9):638-40.
16. Higgins JPT, Green S (editors). *Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011].* The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org>. (Accessed 26th June 2017).
17. Deeks JJ, Wisniewski S, Davenport C. Chapter 4: Guide to the contents of a cochrane diagnostic test accuracy protocol. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0.0.* The Cochrane Collaboration, 2013. Available from: <http://srdta.cochrane.org/>. (Accessed 26th June 2017).
18. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *J Clin Epidemiol* 2009; doi:10.1016/j.jclinepi.2009.06.005

19. University of Bristol. QUADAS-2. Available from: <http://www.bristol.ac.uk/social-community-medicine/projects/quadas/quadas-2/> (Accessed 26th June 2017).
20. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-6.
21. Peloschek PH, Nemecek S, Widhalm P, Donner R, Birngruber H, Thodberg F, et al. Computational radiology in skeletal radiography. *Eur J Radiol*. 2009;72(2):252-257.
22. Sahiner B, Petrick N, Paquerault S, Chen W, Nguyen T. Agreement between two versions of CADx system a simulation study. *Proc of SPIE Med Imaging 2011*;7966.
23. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7).
24. Tracy KD, Dykstra BA, Gakenheimer DC, Scheetz JP, Lacina S, Scarfe WC, et al. Utility and effectiveness of computer-aided diagnosis of dental caries. *Gen Dent*. 2011;59(2):136-44.
25. Kasai S, Li F, Shiraishi J, Doi K. Usefulness of computer-aided diagnosis schemes for vertebral fractures and lung nodules on chest radiographs. *Am J Roentgenol*. 2008;191(1):260-5.
26. Matsumoto T, Hayashi T, Hara T, Katsumata A, Muramatsu C, Zhou X et al. Automated scheme for measuring mandibular cortical thickness on dental panoramic radiographs for osteoporosis screening. *Proc of SPIE Med Imaging 2012*;8315.
27. Kasai S, Li F, Shiraishi J, Li Q, Nie Y, Doi K. Development of computerized method for detection of vertebral fractures on lateral chest radiographs. *Proc of SPIE Med Imaging 2006*;6144.
28. Kasai S, Li F, Shiraishi J, Li Q, Doi K. Computerized detection of vertebral compression fractures on lateral chest radiographs: Preliminary results with a tool for early detection of osteoporosis. *Med Phys*. 2006;33(12):4664-74.

29. Wilkie JR, Giger ML, Pesce LL, Engh Sr CA, Hopper Jr RH, Martell JM. Imputation methods for temporal radiographic texture analysis in the detection of periprosthetic osteolysis. *Proc of SPIE Med Imaging* 2007;6514.
30. Wilkie JR, Giger ML, Engh Sr CA, Martell JM. Investigation of temporal radiographic texture analysis for the detection of periprosthetic osteolysis. *Proc of SPIE Med Imaging* 2006;6144.
31. Roberts MG, Cootes T, Pacheco E, Adams JE. Computer aided diagnosis of osteoporotic vertebral fracture using appearance models and an automatic segmentation. *Calcif Tissue Int.* 2008;1(83):20.
32. Kavitha MS, An SY, An CH, Huh KH, Yi WJ, Heo MS, et al. Texture analysis of mandibular cortical bone on digital dental panoramic radiographs for the diagnosis of osteoporosis in Korean women. *Oral Surg Oral Pathol Oral Radiol.* 2015;119(3):346-56.
33. Pfeil A, Renz DM, Hansch A, Kainberger F, Lehmann G, Malich A, et al. The usefulness of computer-aided joint space analysis in the assessment of rheumatoid arthritis. *Joint Bone Spine.* 2013;80(4):380-5.
34. Oka H, Akune T, Muraki S, Nakamura K, Kawaguchi H, Yoshimura N. Automated quantification and diagnosis of the severity of the knee on plain radiographs: the ROAD study. *Osteoarthritis and Cartilage.* 2009;17(S1):S185.
35. Devlin H, Allen PD, Graham J, Jacobs R, Karayianni K, Lindh C et al. Automated osteoporosis risk assessment by dentists: a new pathway to diagnosis. *Bone.* 2007;40(4):835-42.
36. Kavitha MS, Samopa F, Asano A, Taguchi A, Sanada M. Computer-aided measurement of mandibular cortical width on dental panoramic radiographs for identifying osteoporosis. *J Investig Clin Dent.* 2012;3(1):36-44.
37. Kavitha MS, Asano A, Taguchi A. Continuous measurements of mandibular cortical width on dental panoramic radiographs for computer-aided diagnosis of osteoporosis. *Proc of SPIE Med Imaging* 2011;7963.

38. Muramatsu C, Matsumoto T, Hayashi T, Hara T, Katsumata A, Zhou X, et al. Automated measurement of mandibular cortical width on dental panoramic radiographs. *Int Comput Assist Radiol Surg*. 2013;8(6):877-85.
39. Mandal I. Developing new machine learning ensembles for quality spine diagnosis. *Knowledge Based Systems*. 2015;73:298-310.
40. Donnelley M, Knowles G, Hearn T. A CAD system for long-bone segmentation and fracture detection. *Image Signal Process*. 2008;5099:153-162.
41. Kavitha MS, Asano A, Taguchi A, Heo MS. The combination of a histogram-based clustering algorithm and support vector machine for the diagnosis of osteoporosis. *Imaging Sci Dent*. 2013;43(3):153-61.
42. Nakamoto T, Taguchi A, Ohtsuka M, Suei Y, Fujita M, Tsuda M, et al. A computer-aided diagnosis system to screen for osteoporosis using dental panoramic radiographs. *Dentomaxillofac Radiol*. 2008;37(5):274-81.
43. Sathagirivasan V, Anburajan M. Diagnosis of osteoporosis by extraction of trabecular features from hip radiographs using support vector machine: An investigation panorama with DXA. *Comput Biol Med*. 2013;43(11):1910-9.
44. Massari FE, Cachizumba MO, Gonçalves S, Gallo RA, Agüero RG, et al. Evaluation of a new vertebral fracture assessment method (CADfx) versus conventional radiography. *ASBMR 27th Annual Meeting*. 2005;S156.
45. Langs G, Peloschek P, Bischof H, Kainberger F. Model based erosion spotting and visualization in rheumatoid arthritis. *Acad Radiol*. 2007;14(10):1179-88.
46. Tripathy D. Negative is positive: a plea to publish all studies regardless of outcome. *American Journal of Hematology/Oncology*. 2015;11(4).

Computer-aided detection in musculoskeletal projection radiography: A systematic review

Acknowledgements

There are a great many people to acknowledge and thank for their assistance in this paper. I would primarily like to thank my supervisor Professor Karen Knapp for her perseverance in screening and reviewing thousands of papers and never once trying to lobotomise herself with a pencil. I would also like to thank Robert Meertens for the help with writing and reviewing the protocol, and assessing the quality of the cases. I would also like to thank Dr Judith Meakin for arbitrating on this review and providing feedback during its different incarnations. Finally I would like to thank my girlfriend Charlotte Lovell for her support, time, and for proofreading this article.

Funding: This work originally formed part of my Masters' thesis in medical imaging which was self-funded. As such this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.