RMD Open
Rheumatic & Musculoskeletal Diseases

REVIEW

Use of artificial intelligence in imaging in rheumatology – current status and future perspectives

Berend Stoel

To cite: Stoel B. Use of artificial intelligence in imaging in rheumatology – current status and future perspectives. *RMD Open* 2020;**6**:e001063. doi:10.1136/rmdopen-2019-001063

Received 5 December 2019
Revised 6 January 2020
Accepted 8 January 2020

ABSTRACT

After decades of basic research with many setbacks, artificial intelligence (AI) has recently obtained significant breakthroughs, enabling computer programs to outperform human interpretation of medical images in very specific areas. After this shock wave that probably exceeds the impact of the first AI victory of defeating the world chess champion in 1997, some reflection may be appropriate on the consequences for clinical imaging in rheumatology. In this narrative review, a short explanation is given about the various AI techniques, including 'deep learning', and how these have been applied to rheumatological imaging, focussing on rheumatoid arthritis and systemic sclerosis as examples. By discussing the principle limitations of AI and deep learning, this review aims to give insight into possible future perspectives of AI applications in rheumatology.

Key messages

- There is a long history of artificial intelligence (AI) in imaging in rheumatology, using classical AI methods, but most of them did not make it into clinical practice.
- Recently a subgroup of AI methods, called deep learning, has shown the potential to automatically interpret images beyond human-level accuracy.
- AI and deep learning have, however, practical and theoretical limitations, that may complicate a rapid and widespread use in clinical practice.
- AI will unlikely replace rheumatologists or radiologists in interpreting images, instead a hybrid solution will probably develop that benefits from both artificial and human intelligence.

INTRODUCTION

Currently, artificial intelligence (AI) receives much attention, in which it may sometimes be perceived as an obscure technology that may elusively affect our society. It is however also very clear, that AI can be used less controversially to improve efficacy and efficiency in medical care and research, with a special focus on medical imaging. This narrative review gives an overview of the background of AI and its use in clinical imaging in rheumatology, together with a discussion on its strengths and limitations, and an outlook on future applications.

What is AI?

Although it has become well-known to the general public only recently, AI is by no means new. The first attempts to mimic human reasoning by electrical circuits dates back to the 1940s.

AI can be defined loosely as the capability of a computer system to perform a task that usually or traditionally would require human intelligence. In the AI domain there are different subcategories (see figure 1). AI can include systems that simply execute tasks that would not require the system to learn at all. In medical imaging, for example, AI can be applied to detect anatomical structures using handcrafted algorithms that simply implement the idea of a software engineer. The vast majority of methods that have been developed during the past decades fall within this general category (blue area in figure 1). On the other hand, a subgroup of AI techniques, called 'machine learning' (ML), is able to learn automatically from the data it is presented with, mostly using ground truth data as training sets (ie, supervised learning). This collection of methods again consists of many different algorithms to automatically recognise patterns, also developed during the past decades, of which 'deep learning' is a subgroup. Deep learning is based on the concept of artificial neural networks, that mimic human learning capacity using mathematical representations of neurons and their interconnections. These neural networks exist already a couple of decades but obtained revolutionary results only in the past few years. AI has conquered many obstacles during its development, as overestimation of

Radiology, Division of Image Processing, Leiden University Medical Center, Leiden, The Netherlands

Correspondence to
Dr Berend Stoel;
B.C.Stoel@lumc.nl

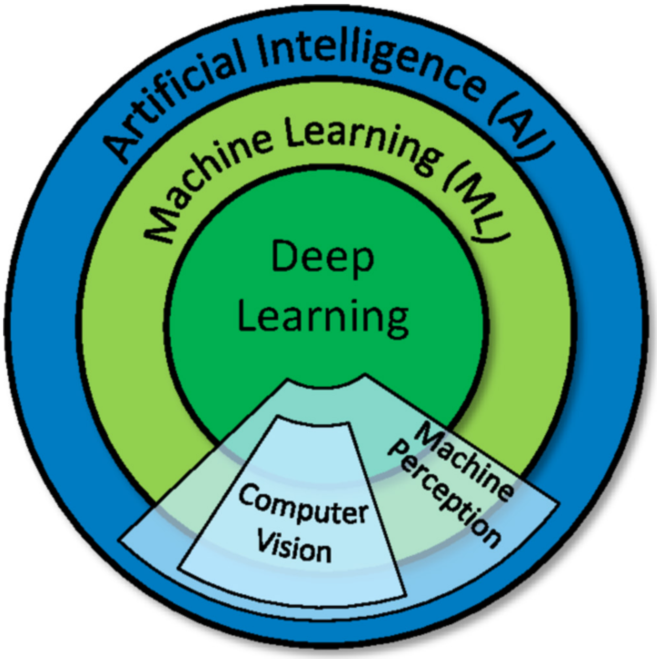

Figure 1 Different subgroups of artificial intelligence methods.

AI capabilities in those days (and maybe still today) and a lack of major breakthroughs led to several setbacks.

Within the domain of each of the above AI categories, there is a wide range of applications using different types of data. If AI methods mimic the way, in which humans use their sense organs to interpret their environment (eg, speech recognition), this is called 'Machine Perception'. If these senses are visual, it is called 'Computer Vision'.

Artificial neural networks

The basic building block of an artificial neural network (ANN) is a mathematical model of a neuron (see figure 2A). An artificial neuron can learn to give an output signal only if a certain combination of inputs occurs. In order to do this, it takes a number of input values, x_1, …, x_n, multiplies each with its corresponding weight, w_1, …, w_n, and accumulate them into a net input for an activation function. This activation function then produces the output 'activation' value, after applying a shift ('bias') to the net input value. The weights and bias are then the parameters that need to be set ('learnt') in order to produce the proper activation.

To tackle more complex problems, a network of neurons can be constructed containing many different layers of neurons, hence the term 'deep' learning. There can be millions of parameters of weights and biasses that then need training. In a hypothetical example of automatically diagnosing rheumatoid arthritis (RA), presented in figure 2B, the first layer may contain input values and the final layer may then provide the outcome of the classification (RA vs not-RA). The input values may represent relevant measurements, such as rheumatoid factor, erythrocyte sedimentation rate, C-reactive protein and anti-citrullinated protein antibodies, but also irrelevant biomarkers (hair colour). For a single subject, the set of input values can be considered a coordinate in a multidimensional 'feature space' (see figure 2C). Applying consecutive layers in an ANN can then be considered transformations of these coordinates, followed by nonlinear scaling. If input values turned out to be irrelevant, the dimension of this space will be reduced by setting the corresponding weights to zero. In the end, an ANN therefore repeatedly deforms the feature space in such a way that in the final layer there are (in this example) only two dimensions left, in which the two groups of patients can be distinguished by taking the label with the highest output value.

Convolutional neural networks

A special type of ANNs are convolutional neural networks (CNNs) that mimic biological vision by taking an image as input layer and adding layers that process images by filters (defined by convolution kernels, the weights of which are trainable). After training, these weights take values such that certain structures (like edges) are enhanced in the images, which are needed to recognise more complex structures by deeper layers in the network. Remarkably, this mechanism resembles the way, in which images are perceived in biological vision. In the human retina, the output of the receptors are preprocessed by retinal ganglion cells in a similar way.[1] Therefore, a CNN not only models learning mechanisms but also seems to mimic evolution of human vision into 'hard-wired' circuits, for example, in the retina (see figure 3, for an example of a CNN for detecting anatomical structures in the wrist).[2 3]

One of the most important steps of developing CNNs for a particular task, is not only designing the network architecture and their settings (hyperparameters), but also collecting and defining the proper data sets. Generally, three separate data sets are used: (1) for *training* the neural network, in order to set the proper weights and biasses; (2) for *validation* of the performance of a trained network using separate data, in order to determine the appropriate architecture and hyperparameters and to prevent overfitting and (3) for *testing* the ultimate network, since usually the design of a network is adapted after a couple of training-validation cycles. If a neural network is successful, a fourth data set may be needed for external validation, in order to prove its general applicability for other scanner manufacturers, imaging protocols or study populations.

Image interpretation

To analyse images with greater efficacy and efficiency, considerable effort has been put into the development of AI methods to perform image processing. It is only recently, however, that these automated methods have gained the accuracy that matches or even surpasses human performance in specific areas, that is, in classifying skin lesions[4] and detecting diabetic retinopathy.[5]

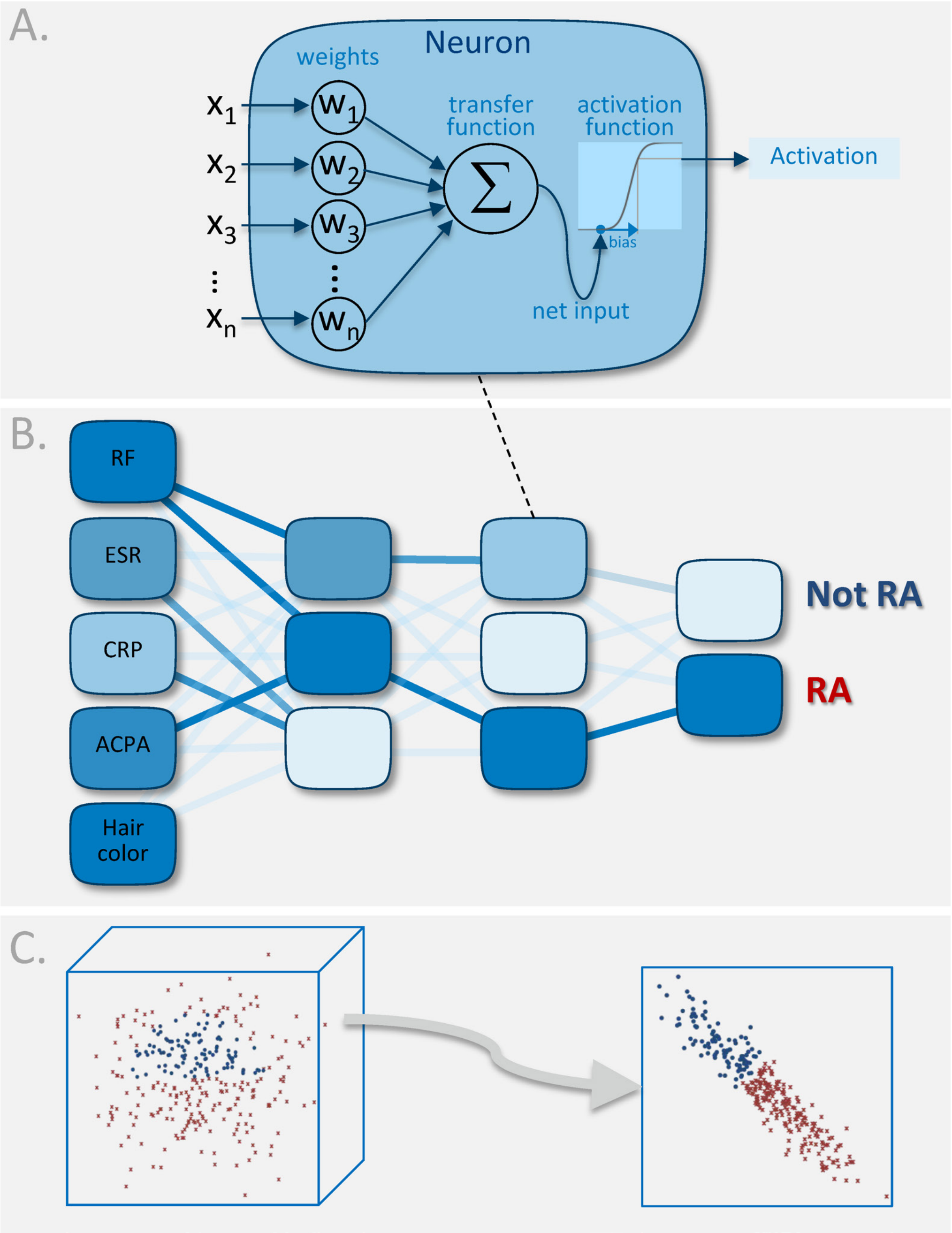

Figure 2 Diagram of an artificial neural network. (A) an artificial neuron takes a number of input values $x_{1,}$ …, x_n, multiplied by the corresponding weights, $w_{1,}$ …, w_n. The inputs, summed by the transfer function, is then input for the activation function that produces the output signal after applying a bias (or threshold). (B) A multi-layered network of artificial neurons could for example be trained to distinguish RA from non-RA, based on blood markers. since irrelevant biomarkers like hair colour obtain small weights, these input values are neglected. (C) The working of an artificial neural network can be considered a consecutive deformation of the feature space such that the last layer makes a clear distinction between RA and not-RA, based on only two values, that are non-linear combinations of the input values. RA,rheumatoid arthritis.

Ever since the advent of digital imaging in the 1970s, human interpretation of images gradually started to rely on image processing techniques. This may even go unnoticed, but digital images are nowadays always fully automatically enhanced or reconstructed first, before they are presented to the observer. Thus, human interpretation is now fully dependent on computerised image processing. But there are other levels, at which a computer can help the physician to subsequently interpret the resulting images (see figure 4). Dependent on the level of automation, these methods are divided into fully-human, semi-automatic and fully-automatic interpretation.

Fully-human interpretation

As performed mostly in daily clinical practice, a physician interprets the digital image by visually detecting lesions and directly classify them (route A-1 in figure 4, from visual detection to cognitive characterisation). Alternatively,

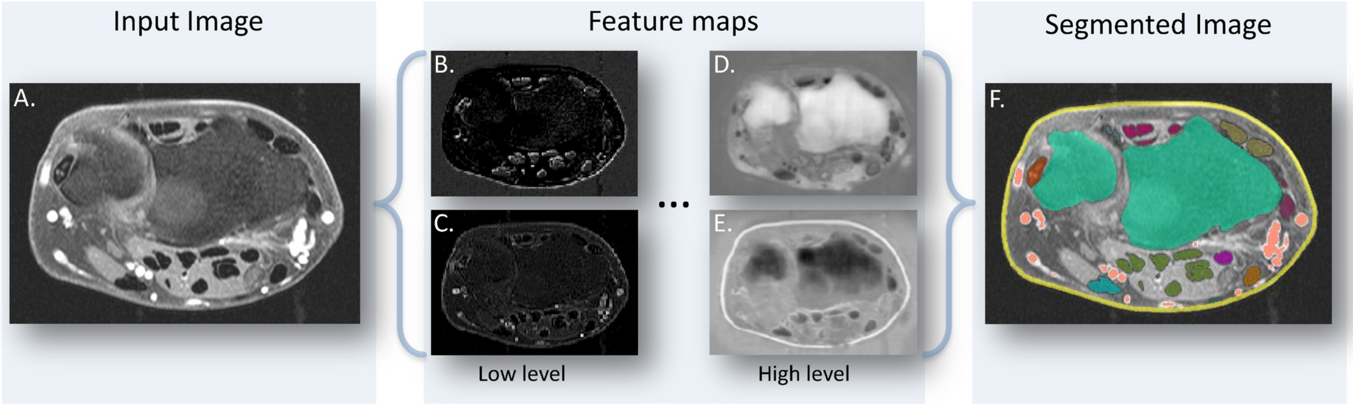

Figure 3 An axial MRI image of the wrist[2] (A), processed by a CNN. The neural network decomposes the image into different so-called 'feature maps'.[3] The CNN apparently learnt to apply low level filters first, by, for example, enhancing transitions from dark to bright in the vertical direction (B), or enhancing bright spots surrounded by a dark area (C). In the deeper layers of the network, it highlights higher level patterns, that seem to represent, for example, bony structures (D) and skin (E). In the final layers, the network combines all information into one output image that consists of labels for skin (yellow), bones (green), vessels (pink) and the different tendon groups (in separate colours) (F). ACPA,anti-citrullinated protein antibodies; CNN, convolutionalneural network; CRP, C-reactive protein; ESR, erythrocytesedimentation rate; RA, rheumatoid arthritis;RF, rheumatoid factor.

in order to come to a semi-quantitative interpretation, the severity of the detected lesion can be scored visually, based on visually estimating handcrafted features (such as joint space narrowing or image intensities that indicate inflammation). Based on this visual score the lesion can be classified (route A-I-2). These features can also be obtained from manual calliper measurements (route A-II-2). An example of the latter approach is using measurements of the joint space width in mm's from X-rays (for assessing joint space narrowing in hand and feet joints), or the extent of inflammatory region around a tendon from MRI also in mm's (to measure tenosynovitis in RA).

Semi-automatic interpretation

In a hybrid or 'computer-aided' approach, the computer can facilitate parts of the detection, quantification and classification tasks, in different ways:

a. The detection of lesions or anatomical structures is done manually, after which quantification is done automatically by computer. The subsequent classification can then be done either manually (route B-III-2) or automatically (route B-III-3). Early examples of the latter approach is the manual annotation of the synovial membrane in MRI scans, followed by an automatic calculation of the synovial volume to quantify synovitis in RA.[6]

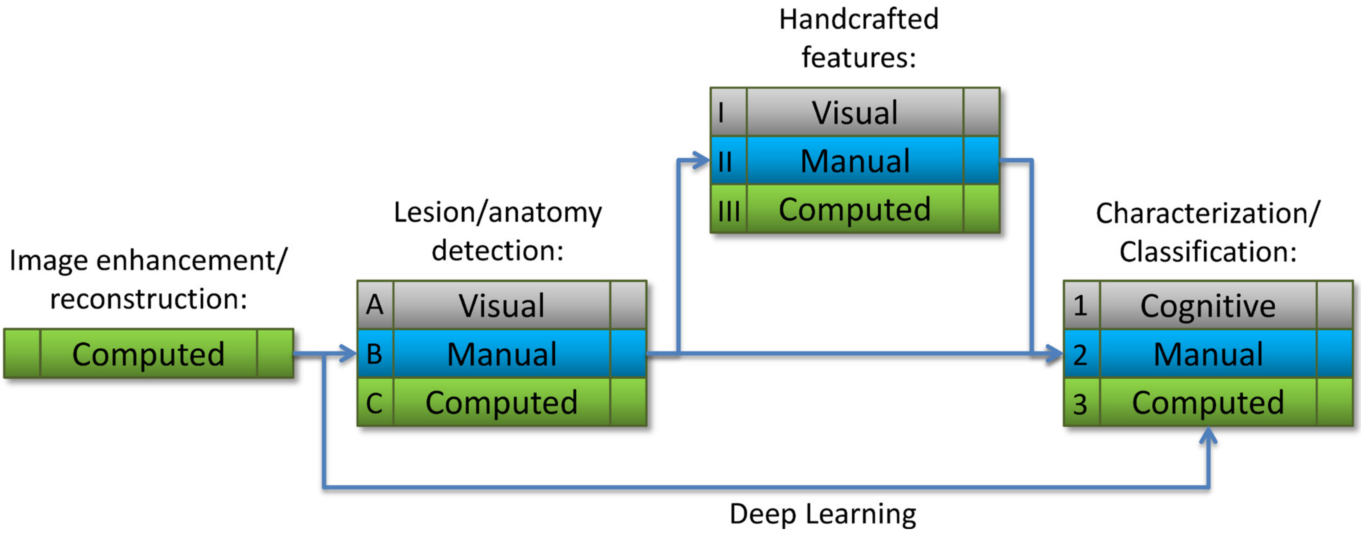

Figure 4 Possible pathways for classifying diseases based on imaging. Each phase (lesion detection, feature selection and classification) can be performed in three different ways: (1) By an observer using her/his knowledge, experience or 'gut feeling' (indicated by grey labels, A, I and 1, respectively); or (2) by an observer who manually annotates, measures and classifies (blue labels B, II and 2, respectively) or (3) by computer (green labels, C, III and 3, respectively).

b. Both detection and quantification of a lesion is performed automatically by AI methods, followed by a manual classification in order to come to a diagnosis (route C-III-2). Since the quantification is usually based on clearly defined (handcrafted) features, the quantification is performed automatically.

Fully-automatic interpretation

Here one can take one of two approaches.

a. In the 'classical' image processing approach, the lesions or anatomical structures are first detected automatically, followed by quantification of handcrafted features, and subsequently classification (route C-III-3); or
b. By using machine (deep) learning the handcrafted features are replaced by automatically learnt features. This means that detection, quantification and classification are performed in one step.

APPLICATIONS OF AI IN IMAGING IN RHEUMATOLOGY

In most publications, the goal of developing and applying AI in medical imaging is described along the same pattern. Sometimes, automation is required as an alternative to visual scoring, because automation would be less expensive, time-consuming and/or subjective, producing less interobserver and intraobserver variability; and it may require less extensive training and/or specific examiner skills. This goal is especially relevant for clinical trials, where subtle changes or treatment effects are expected that need to be detected accurately. Other reasons for automation are that these measurements could be more specific than visual scoring because a computer program is more consistent and less distracted by other image information.[7]

In some situations however, automatic quantification is not an alternative to visual scoring but a measurement on its own, because the imaging modality is already quantitative by nature (eg, CT densitometry or compositional MRI). The same applies when an automated method can quantify features that are impractical or impossible to assess by humans, because of the sheer amount of imaging data, or limitations in the human visual system, for example, difficulties in detecting subtle differences in absolute intensities.

The next three paragraphs give some examples of automatic quantification or classification in two distinct rheumatic disorders, RA and systemic sclerosis (SSc), representing primarily joint disorders and typical multiorgan diseases, respectively, followed by a special AI topic on comparative imaging.

Rheumatoid arthritis

As imaging of RA is mainly focused on (teno)synovitis, bone marrow oedema, bone erosions and cartilage loss, the developed automated image analysis methods are aimed to detect and assess these manifestations. Synovitis and bone marrow oedema can be assessed directly, as they invoke an increase or decrease in image intensity. In other words, the lesion creates its own signal; if there is no signal, there is no pathology. Bone erosions and cartilage loss would require a slightly different approach where bone or cartilage is first detected, followed by a comparison with an estimated original (or normal) volume of bone or cartilage. In other words, there is always a signal, but the question is whether its volume has decreased after disease onset. If a comparison with the original situation is not possible, the bone/cartilage measurement can only be used for monitoring disease progression, for example by quantifying changes in bone lesions over time.[8]

(Teno)synovitis

One of the first attempts to generally quantify inflammation in contrast-enhanced MRI scans of the hand was made in 2007, by classifying the image intensities into normal and inflamed using ML (Fuzzy clustering).[9] Alternatively, dynamic contrast enhanced (DCE-)MRI of the hand and wrist have been used to quantify synovitis. These methods have been developed to align the time sequence of images through automatic image registration and subsequently classify the intensity versus time curves of each voxel into different patterns. The resulting perfusion maps can be used to detect synovitis and quantify their extent.[10] As both of the above approaches may produce false positives outside the synovium, research has also been focussed on separately detecting and quantifying the synovium, based on prior segmentation of the bones, in, for example, the wrist.[11] Deep learning has not been applied yet to classify MR images in terms of synovitis.

In ultrasound (power Doppler), the amount of available automated methods to quantify synovitis is relatively small compared with other imaging modalities, probably because of the technical challenges in analysing these images, due to image noise. Using classical image processing and ML, a method has been presented to automatically detect skin, bones and synovitis regions, the latter of which was quantified by their area.[12] As in MRI, efforts have been made to detect bones as a preprocessing step for accurately defining the synovium, with classical AI methods[13] and deep learning.[14] Recently, applications of deep learning have been proposed to directly quantify synovitis from the entire image, by training a CNN with visual scoring as ground truth.[15]

Similarly to synovitis, quantification of tenosynovitis requires an automatic segmentation of tendons, in order to define the surrounding synovium and quantify inflammation in contrast-enhanced MRI. Our group has developed such an automatic framework in order to quantify tenosynovitis in the wrist.[2] Next steps in our research are to apply deep learning in order to detect tendons (see figure 3) and quantify tenosynovitis, but also to classify the entire image by a single CNN.

Bone lesions

In order to quantify bone marrow oedema (BME), the first classical AI methods were based on DCE-MRI of the wrist.[16] Using static postcontrast MRI scans of the wrist, BME could also be assessed automatically, specifically in early RA.[17] So far, however, deep learning has not been investigated yet in measuring BME.

Early AI developments for automatically quantifying bone erosion use fractal analysis to characterise the texture of trabecular bones in the radius, based on plain radiographs.[18] This type of analysis has only been employed in research, probably because measures from fractal analysis may be difficult to interpret clinically, especially if there are no normal values.

An alternative method detects the bones from hand radiographs and compare their shape with a statistical shape model from healthy bones and determine deviations from that model to detect and quantify erosions.[19] More recently, deep learning has been employed to detect bone erosions from radiograph of hands.[20] Using peripheral CT of the hand, classical AI-based[21] and CNN-based[22] methods were able to detect bone erosions.

Cartilage loss

Of the different RA features, most effort has been put into automatic quantification of cartilage loss, probably because of the wide availability of radiographs and because cartilage loss is also relevant for osteoarthritis.

From plain radiographs, cartilage loss can be assessed only indirectly, by measuring the joint space, since cartilage itself is not visible. Early AI development to quantify joint space width in the hand[23] and knee[24] date back to 1989. More recently, classical AI methods have been proposed to automatically measure joint spaces within the wrist.[25]

In MRI scans, cartilage thickness and volume have been measured using a wide variety of classical AI methods.[26] Deep learning has been used to detect knee cartilage,[27] which has been developed further more recently in order to classify cartilage lesions.[28] Similarly, CNNs have also been applied to detect cartilage in the wrist joints.[29]

Systemic sclerosis

As SSc involves multiorgans, research on automated imaging biomarkers for vasculopathy and fibrosis has been focussed on different anatomical sites, such as skin, heart and lungs.

Vasculopathy

Nailfold capillaroscopy is an established modality to distinguish and classify sequential patterns of capillary abnormalities that are considered representative to the general microvascular involvement of SSc. With early computerised tools, the observer was able to perform calliper measurements (capillary dimensions and density) on screen, based on completely manually drawn contours[30] or manually defined landmarks.[31] Initially, the developed image processing methods closely followed the way, in which human observers would recognise morphological abnormalities (loop width, capillary density, etc), but researchers have also designed morphological features that are less obvious to quantify by hand, such as tortuosity[32] and texture,[33] followed by automated classification by ML.

Using chest CT, vasculopathy can be quantified with AI by detecting pulmonary vessels that can be resolved by CT (ie, down to ~0.5 mm in diameter) and subsequently quantifying their morphology to produce imaging biomarkers.[34] These biomarkers estimate the number of small vessels and the difference in occurrence between small vessels and large vessels, which are found to be correlated with functional impairment of the lungs, in the absence of lung fibrosis. This is a typical example where a method is intended to be complementary to visual scoring, since it is practically impossible to measure all vessels in the lungs by hand or score them visually. Currently, our research group is working on applying neural networks to automatically distinguish pulmonary arteries from veins,[35] so that it becomes feasible to measure the pulmonary vasculature separately, to increase its specificity (see figure 5).

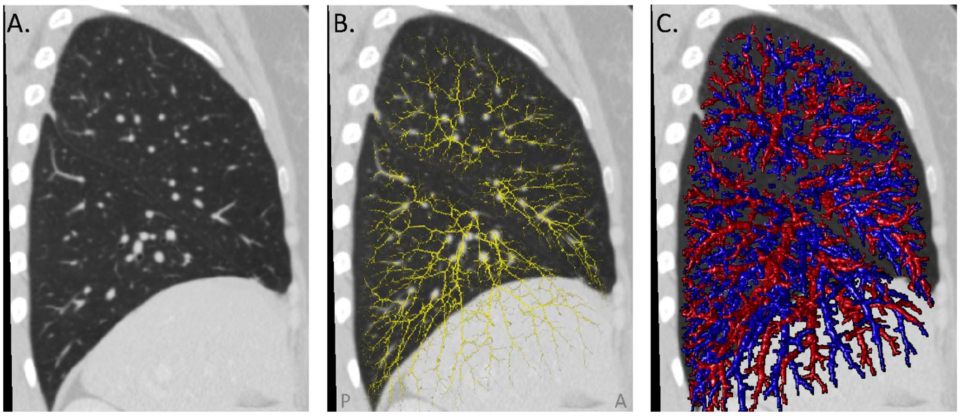

Figure 5 Deep learning-based detection of the arterial and venous pulmonary vascular.[35] (A) Sagittal cross-section of a right lung on chest CT; (B) detection of all centre lines of the pulmonary vessels; (C) classification into arteries (red) and veins (blue) by deep learning.

Cardiac complications of SSc can be quantified automatically from imaging by methods that rely on a long history of general applications in cardiology. ML methods have already been able to automatically quantify left and right ventricular function, based on ultrasound,[36] MRI[37] and CT.[3835]

Fibrosis

In estimating fibrosis as manifested in changes in skin thickness, the modified Rodnan Skin Score methods based on palpation is the gold standard. Fully-quantitative alternatives have been proposed to measure skin thickness more directly from imaging.[39] These methods employ for example ultrasound, MRI or optical coherence tomography (OCT) to determine skin thickness and other characteristics. AI-based image analysis methods have been introduced, to fully automatically measure skin thickness[40] and other (epi)dermal characteristics[41] from ultrasound images, and MRI[42] and OCT.[43] These automated imaging biomarkers, however, still need to be evaluated because the relatively new imaging techniques require further optimisation and standardisation, and have not yet been applied in SSc.

In chest CT, fibrosis has also been assessed automatically in the lungs, as an objective and more reproducible alternative to the visual Goh scoring system. Lung densitometry may be used to quantify interstitial lung disease, since lung tissue mass is expected to increase over time, which has been proven to correlate with a decline in lung function.[44]

With the success of deep learning in classifying images, an obvious next step is to develop neural networks that can distinguish different patterns for differential diagnosis and SSc staging from the different image modalities.

Comparative imaging

Whenever the question arises whether RA has progressed or not, a specific task is to basically ‘spot the differences’ in consecutive images. The evolution of the human visual system has however not been optimised to look for differences between two images (this is probably the reason why ‘spot the difference’ is a popular children’s game). On the other hand, computer vision is better equipped for detecting differences in image intensities. Therefore, a separate field of AI research, called ‘comparative imaging’ is focussed on providing difference images to the radiologist/rheumatologist to help find locations of RA progression or regression. In our research group for example, software has been developed to show inflammatory changes in vertebral lesions in axial spondyloarthropathy by matching baseline and follow-up MRI scans through image registration and subsequently display the changes over time by image subtraction and colour coding[45] (see figure 6A). For SSc we have developed software to map the local changes in lung density over time (figure 6B). Other researchers have developed similar

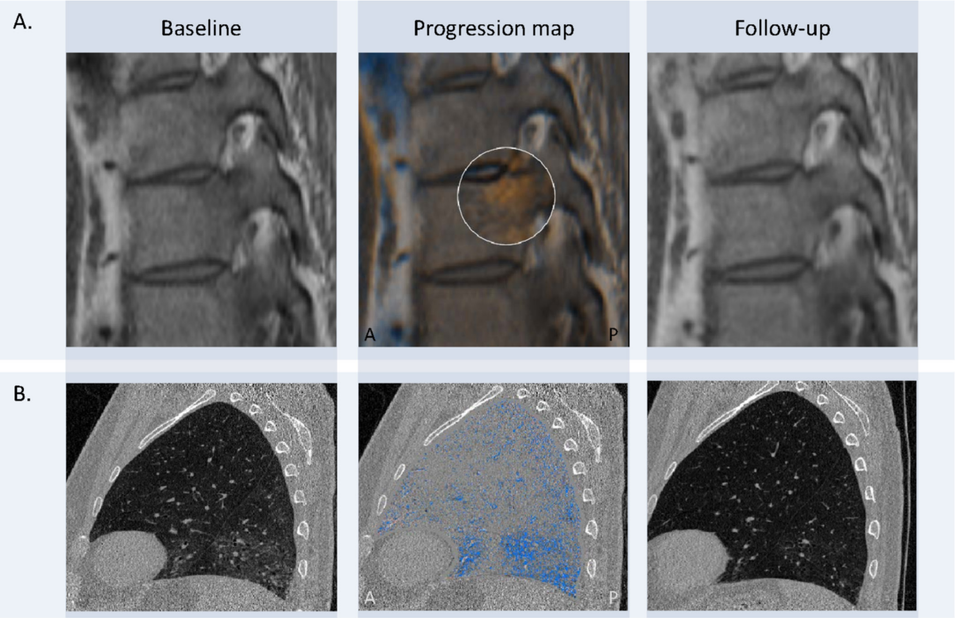

Figure 6 Two examples of comparative imaging. (A) Baseline (left panel) and follow-up (right) sagittal T1-weighted turbo spin echo (TSE) MRI scan of three thoracic vertebrae of a patient with suspected axial spondyloarthritis. Change in low signal intensity at the posterior vertebral body (high on stir – not shown) consistent with bone marrow oedema is appreciated. The middle panel shows the progression map, where a decrease in intensity is displayed in orange, and an increase in blue.[45] (B) Baseline (left) and follow-up (right) CT scan of a patient with diffuse systemic sclerosis[44] (left lung). The progression map in the middle panel shows the local density changes, with the same colour code; blue: decrease in density, orange: density increase. The map demonstrates that the ground glass densities were partly resolved by treatment. These two examples show clear changes in the images as an illustration, but more subtle difference that are less obvious to the human eye can also be displayed in this manner.[47]

comparative imaging techniques to display changes in joint spaces in sequential hand radiographs[46] and in bone lesions[8] in MRI of the hand. It is very likely that deep learning methods will also enter this area of research, where CNNs could learn to distinguish relevant from irrelevant changes in image intensity over time.[44 45 47]

FUTURE PERSPECTIVES

From the above overview, it is clear that there is extensive literature on imaging biomarkers in rheumatology, mostly using classical AI methods (using separate segmentation, feature extraction and classification steps). Applications of deep learning in rheumatology is sparse, but this will grow rapidly.

Until recently radiologists and rheumatologists were the only interpreters of clinical rheumatological images, and these images were digital solely because of their ease of archiving and transferring data. Recently, however, neural networks started to outperform human interpretation of images in very specific tasks.[4 5] It is however unlikely that AI will replace humans, since generalisation of computerised image interpretation is much more difficult than in very narrow areas of expertise. A more likely scenario would be the development of a hybrid solution, which benefits from both artificial and human intelligence. As humans are less capable of analysing large amounts of multidimensional data, AI may help reducing the dimensionality or recognising patterns that are not apparent to the human eye and brain. This is especially true when image data is combined with other clinical data, producing even more big data, which cannot be handled with human power only.

Predicting the future of AI is tempting but difficult, as the time scale at which developments will occur is very difficult to estimate, considering the many setbacks AI experienced during its development. Since overestimation of AI's capabilities is lurking, we need to be critical and be aware of the drawbacks and limitations of AI and deep learning.

Limitations of deep learning

All deep learning methods presented in this paper are examples of supervised learning, meaning that large amounts of ground truth data are needed to train the system. If these ground truth data can only be obtained from human interpretation, we are actually back to square one. Therefore, much effort needs to be put into manually annotating anatomical structures and lesions in images or classifying them (in consensus), in order to provide an accurate ground truth. Without a proper quality control, however, deep learning will learn to copy human mistakes. Furthermore, visual scoring may not be considered ground truth anymore, if our goal were to surpass human interpretation. Therefore alternative reference outcome measures will be needed, for example from longitudinal data, where the task will shift towards prediction of clearly defined outcomes instead of static assessment of disease severity.

Neural networks are certainly not off-the-shelf solutions, as there is not a generic network that solves all possible problems, like the human brain does. Therefore, today human expertise is still needed to design a neural network, including its settings, and to perform proper training, validation and testing. Evolutionary algorithms may, however, help designing networks in the future, in order to find the optimum architecture and hyperparameters automatically.

Much like humans may sometimes rely on their experience or 'gut feeling' without being able to clearly explain their decision, neural networks can also be considered 'black boxes'. There are developments towards opening the 'black box' by providing so-called class-discriminating attention maps,[48] which may give at least an indication of where the network had focussed on, in order to come to a certain classification. When it has been proven irrefutably that a neural network outperforms humans in a certain task, we will probably need to adapt and trust these systems in much the same way we would need to trust human experts. Legal and ethical issues are still open, however, especially on liability of machine-based clinical decisions and in collecting large amounts of confidential data.

It is unlikely that deep learning will completely replace all classical AI algorithms. For simple problems, deep learning may be too complex and non-linear to produce the proper solution. Furthermore, deep learning invents and extracts its own features from the images to come to a classification, whereas classical AI uses man-made features to measure. In many cases, however, there is a need for measurements, of which the physical meaning is clear. For example, the dimensions of a vessel is much more informative than a classification of this vessel by a neural network. In those cases, deep learning can still be very useful in the first part of the process, where the vessels need to be detected first.

Similar to the education and certification of a biological neural network (ie, the rheumatologists), an artificial neural network also needs to be trained and validated extensively, before it can be applied in the clinic. There is, however, an imbalance between the amount of published technical innovations, as described above, and the amount of techniques that actually make it into clinical practice. Due to publication bias, a large part of the innovations simply does not pass external validation, or only works on very specific data sets. But it also seems that performing validation studies is a limiting factor in itself. Researchers may only be interested in the purely technical innovation without aiming for a clinical application ('hit and run' research), and software validation is less interesting to basic researchers because of the amount of administration needed to obtain clearance from regulatory organisations for such a medical device. Software validation is therefore mostly performed by commercial companies, but still only if there is clear market viability for this product. Thus, a closer collaboration between universities and industry may be needed, as well as between technical and clinical researchers. For these types of collaborations, interdisciplinary university studies are essential to bridge this gap further. There are already several initiatives from the different professional organisations to create new training programme in the use of AI in medical imaging.

Dependency on image acquisition and reconstruction

Although this may sound trivial, it is good to keep realising that AI – as any assessment – is bounded by the somewhat denigrating concept of 'garbage in, garbage out'. The accuracy of AI in image analysis is in the end limited to the accuracy of the image modality it takes as input. In cases where human image interpretation is the limiting factor, AI can help improving sensitivity and specificity of a particular assessment, by extracting additional information from the images. If not, the best AI can do is to improve reproducibility and time-efficiency and cost-efficiency. But even then, overall reproducibility may be limited by physiological variations (respiration or perfusion, for example, in pulmonary vascular imaging), or by operator dependency, for example, in ultrasound imaging.[49] But even for the latter, CNNs may help reducing variability by automatic selection of the most informative frame.[50] It is therefore important for AI experts to be on the watch for new image acquisition and reconstruction methods that may be more tailored to detect the specifics of a disease with higher reproducibility, and continuously adapt their AI methods to these new modalities or imaging protocols.

Dependency on image acquisition settings can also be more subtle, especially in follow-up analysis. As image analysis becomes more and more quantitative and sensitive, these methods may also pick up subtle changes in the imaging modality itself instead of the study subjects. In lung CT densitometry for example, the measured densities may shift or drift, simply by an update of the scanner's reconstruction software or due to ageing of the X-ray tube. Therefore, quantitative analysis will put stricter requirements on the constancy of imaging devices. Moreover, so far image acquisition protocols have been optimised primarily for human interpretation. It could well be, however, that the optimal settings for image acquisition and reconstruction are fundamentally different for AI methods. Therefore, this has opened a new field of research aimed at formalising these image quality requirements and developing standardisation to guarantee stable diagnostic accuracy of AI.

CONCLUSIONS

Artificial intelligence has already been applied in imaging in rheumatology for a long time. In clinical practice however, only few of these methods have been adopted. The recent successes of deep learning will dramatically change this, where AI will be supplementary to human image interpretation and clinical reasoning.

Acknowledgements The author would like to thanks Denis Shamonin, Evgeni Aizenberg, Zhiwei Zhai, Irene Hernández-Girón and Marius Staring for their technical contributions to this review paper, and Anne Schouffoer and Monique Reijnierse for reviewing the clinical aspects.

Contributors BCS performed the literature search and wrote the manuscript.

Funding Parts of the presented illustrations originate from the APPEAR project and ESMIRA project, both funded by the Dutch Technology Foundation (currently TTW) as part of the Netherlands Organization for Scientific Research (NWO), under grant number LPG. 07998 and 13329, respectively, and from support by China Scholarship Council scholarship number 201406120046.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

Data availability statement Not applicable

ORCID iD

Berend Stoel http://orcid.org/0000-0002-5975-8559

REFERENCES

1 Rodieck RW, Stone J. Analysis of receptive fields of cat retinal ganglion cells. *J Neurophysiol* 1965;28:833–49.

2 Aizenberg E, Shamonin DP, Reijnierse M, *et al*. Automatic quantification of tenosynovitis on MRI of the wrist in patients with early arthritis: a feasibility study. *Eur Radiol* 2019;29:4477–84.

3 Yosinski J, Clune J, Nguyen A, *et al*. Understanding neural networks through deep visualization. *arXiv*. *preprint arXiv* 2015;150606579.

4 Esteva A, Kuprel B, Novoa RA, *et al*. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.

5 Gulshan V, Peng L, Coram M, *et al*. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus Photographs. *JAMA* 2016;316:2402–10.

6 Ostergaard M*et al*. Quantitative assessment of the synovial membrane in the rheumatoid wrist: an easily obtained MRI score reflects the synovial volume. *Rheumatology* 1996;35:965–71.

7 van 't Klooster R, Hendriks EA, Watt I, *et al*. Automatic quantification of osteoarthritis in hand radiographs: validation of a new method to measure joint space width. *Osteoarthritis Cartilage* 2008;16:18–25.

8 Carano RAD, Lynch JA, Redei J, *et al*. Multispectral analysis of bone lesions in the hands of patients with rheumatoid arthritis. *Magn Reson Imaging* 2004;22:505–14.

9 Tripoliti EE, Fotiadis DI, Argyropoulou M. Automated segmentation and quantification of inflammatory tissue of the hand in rheumatoid arthritis patients using magnetic resonance imaging data. *Artif Intell Med* 2007;40:65–85.

10 Kubassova O, Boesen M, Boyle RD, *et al*. Fast and robust analysis of dynamic contrast enhanced MRI datasets. *Med Image Comput Comput Assist Interv* 2007;10:261–9.

11 Czaplicka K, Wojciechowski W, Włodarczyk J, *et al*. Automated assessment of synovitis in 0.2T magnetic resonance images of the wrist. *Comput Biol Med* 2015;67:116–25.

12 Cupek R, Ziębiński A. Automated assessment of joint synovitis activity from medical ultrasound and power Doppler examinations using image processing and machine learning methods. *R* 2016;5:239–42.

13 Cao K, Mills DM, Thiele RG, *et al*. Toward quantitative assessment of rheumatoid arthritis using volumetric ultrasound. *IEEE Trans Biomed Eng* 2016;63:449–58.

14 Hemalatha RJ, Vijaybaskar V, Thamizhvani TR. Automatic localization of anatomical regions in medical ultrasound images of rheumatoid arthritis using deep learning. *Proc Inst Mech Eng H* 2019;233:657–67.

15 Andersen JKH, Pedersen JS, Laursen MS, *et al*. Neural networks for automatic scoring of arthritis disease activity on ultrasound images. *RMD Open* 2019;5:e000891.

16 Boesen M, Kubassova O, Bouert R, *et al*. Correlation between computer-aided dynamic gadolinium-enhanced MRI assessment of inflammation and semi-quantitative synovitis and bone marrow oedema scores of the wrist in patients with rheumatoid arthritis--a cohort study. *Rheumatology* 2012;51:134–43.
17 Aizenberg E, Roex EAH, Nieuwenhuis WP, *et al*. Automatic quantification of bone marrow edema on MRI of the wrist in patients with early arthritis: a feasibility study. *Magn. Reson. Med.* 2018;79:1127–34.
18 Disini L, Foster M, Milligan PJ, *et al*. Cancellous bone changes in the radius of patients with rheumatoid arthritis: a cross-sectional quantitative macroradiographic study. *Rheumatology* 2004;43:1150–7.
19 Langs G, Peloschek P, Bischof H, *et al*. Automatic quantification of joint space narrowing and erosions in rheumatoid arthritis. *IEEE Trans Med Imaging* 2009;28:151–64.
20 Rohrbach J, Reinhard T, Sick B, *et al*. Bone erosion scoring for rheumatoid arthritis with deep convolutional neural networks. *Computers & Electrical Engineering* 2019;78:472–81.
21 Topfer D, Finzel S, Museyko O, *et al*. Segmentation and quantification of bone erosions in high-resolution peripheral quantitative computed tomography datasets of the metacarpophalangeal joints of patients with rheumatoid arthritis. *Rheumatology* 2014;53:65–71.
22 Ren J, Moaddel A, Hauge EM, *et al*. Automatic detection and localization of bone erosion in hand HR-pQCT. medical imaging 2019: computer-aided diagnosis. *Int Soc Opt Photonics* 2019:1095022.
23 Allander E, Forsgren PO, Pettersson H, *et al*. Computerized assessment of radiological changes of the hand in rheumatic diseases. *Scand J Rheumatol* 1989;18:291–6.
24 Dacree JE, Huskisson EC. The automatic assessment of knee radiographs in osteoarthritis using digital image analysis. *Rheumatology* 1989;28:506–10.
25 Huo Y, Vincken KL, van der Heijde D, *et al*. Automatic quantification of radiographic wrist joint space width of patients with rheumatoid arthritis. *IEEE Trans Biomed Eng* 2017;64:2695–703.
26 Eckstein F, Peterfy C. A 20 years of progress and future of quantitative magnetic resonance imaging (qMRI) of cartilage and articular tissues—personal perspective. *Semin Arthritis Rheum* 2016;45:639–47.
27 Prasoon A, Petersen K, Igel C, *et al*. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Med Image Comput Comput Assist Interv* 2013;16:246–53.
28 Liu F, Zhou Z, Samsonov A, *et al*. Deep learning approach for evaluating knee Mr images: achieving high diagnostic performance for cartilage lesion detection. *Radiology* 2018;289:160–9.
29 Brui E, Efimtcev AY, Fokin VA, *et al*. Deep learning-based fully automatic segmentation of wrist cartilage in Mr images. *arXiv preprint arXiv* 2018.
30 Scheja A, Akesson A, Niewierowicz I, *et al*. Computer based quantitative analysis of capillary abnormalities in systemic sclerosis and its relation to plasma concentration of von Willebrand factor. *Ann Rheum Dis* 1996;55:52–6.
31 Anderson ME, Allen PD, Moore T, *et al*. Computerized nailfold video capillaroscopy-a new tool for assessment of Raynaud's phenomenon. *J Rheumatol* 2005;32:841–8.
32 Jones BF, Oral M, Morris CW, *et al*. A proposed taxonomy for nailfold capillaries based on their morphology. *IEEE Trans Med Imaging* 2001;20:333–41.
33 Schaefer G, Krawczyk B, Doshi NP, *et al*. Scleroderma capillary pattern identification using texture descriptors and ensemble classification. *Conf Proc IEEE Eng Med Biol Soc* 2013;2013:5473–6.
34 Zhai Z, Staring M, Ninaber MK, *et al*. Pulmonary vascular morphology associated with gas exchange in systemic sclerosis without lung fibrosis. *J Thorac Imaging* 2019;34:373–9.
35 Zhai Z, Staring M, Zhou X, *et al*. Linking convolutional neural networks with graph convolutional networks: application in pulmonary artery-vein separation. medical image computing and computer assisted intervention. *Shenzhen* 2019.
36 Genovese D, Rashedi N, Weinert L, *et al*. Machine Learning-Based three-dimensional echocardiographic quantification of right ventricular size and function: validation against cardiac magnetic resonance. *J Am Soc Echocardiogr* 2019;32:969–77.
37 Tao Q, Yan W, Wang Y, *et al*. Deep Learning-based method for fully automatic quantification of left ventricle function from cine Mr images: a Multivendor, multicenter study. *Radiology* 2019;290:81–8.
38 Mao SS, Li D, Vembar M, *et al*. Model-Based automatic segmentation algorithm accurately assesses the whole cardiac volumetric parameters in patients with cardiac CT angiography: a validation study for evaluating the accuracy of the workstation software and establishing the reference values. *Acad Radiol* 2014;21:639–47.
39 Kang T, Abignano G, Lettieri G, *et al*. Skin imaging in systemic sclerosis. *Eur J Rheumatol* 2014;1:111–6.
40 Lagarde J-M, George J, Soulcie R, *et al*. Automatic measurement of dermal thickness from B-scan ultrasound images using active contours. *Skin Res Technol* 2005;11:79–90.
41 Sciolla B, Le Digabel J, Josse G, *et al*. Joint segmentation and characterization of the dermis in 50 MHz ultrasound 2D and 3D images of the skin. *Comput Biol Med* 2018;103:277–86.
42 Ognard J, Mesrar J, Benhoumich Y, *et al*. Edge detector-based automatic segmentation of the skin layers and application to moisturization in high-resolution 3 tesla magnetic resonance imaging. *Skin Res Technol* 2019;25:339–46.
43 Abignano G, Aydin SZ, Castillo-Gallego C, *et al*. Virtual skin biopsy by optical coherence tomography: the first quantitative imaging biomarker for scleroderma. *Ann Rheum Dis* 2013;72:1845–51.
44 Ninaber MK, Stolk J, Smit J, *et al*. Lung structure and function relation in systemic sclerosis: application of lung densitometry. *Eur J Radiol* 2015;84:975–9.
45 Aizenberg E, van den Berg R, Ez-Zaitouni Z, *et al*. Computer-Aided evaluation of inflammatory changes over time on MRI of the spine in patients with suspected axial spondyloarthritis: a feasibility study. *BMC Med Imaging* 2017;17:55.
46 Ichikawa S, Kamishima T, Sutherland K, *et al*. Computer-Based radiographic quantification of joint space narrowing progression using sequential hand radiographs: validation study in rheumatoid arthritis patients from multiple institutions. *J Digit Imaging* 2017;30:648–56.
47 Staring M, Bakker ME, Stolk J, *et al*. Towards local progression estimation of pulmonary emphysema using CT. *Med Phys* 2014;41:021905.
48 Tiulpin A, Thevenot J, Rahtu E, *et al*. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep Learning-Based approach. *Sci Rep* 2018;8:1727.
49 Scheel AK, Schmidt WA, Hermann K-GA, *et al*. Interobserver reliability of rheumatologists performing musculoskeletal ultrasonography: results from a EULAR "Train the trainers" course. *Ann Rheum Dis* 2005;64:1043–9.
50 Fiorentino MC, Moccia S, Cipolletta E, *et al*. A Learning Approach for Informative-Frame Selection in US Rheumatology Images. In: *International Conference on image analysis and processing*. Springer, 2019: 228–36.