



Deep learning analysis for rheumatologic imaging: current trends, future directions, and the role of human

Jucheol Moon, Ph.D.¹, Pratik Jadhav, M.S.¹, Sangtae Choi, M.D., Ph.D.²

¹Department of Computer Engineering and Computer Science, College of Engineering, California State University Long Beach, Long Beach, CA, USA, ²Division of Rheumatology, Department of Internal Medicine, Chung-Ang University College of Medicine, Seoul, Korea

Rheumatic diseases, such as rheumatoid arthritis (RA), osteoarthritis (OA), and spondyloarthritis (SpA), present diagnostic and management challenges due to their impact on connective tissues and the musculoskeletal system. Traditional imaging techniques, including plain radiography, ultrasounds, computed tomography, and magnetic resonance imaging (MRI), play a critical role in diagnosing and monitoring these conditions, but face limitations like inter-observer variability and time-consuming assessments. Recently, deep learning (DL), a subset of artificial intelligence, has emerged as a promising tool for enhancing medical imaging analysis. Convolutional neural networks, a DL model type, have shown great potential in medical image classification, segmentation, and anomaly detection, often surpassing human performance in tasks like tumor identification and disease severity grading. In rheumatology, DL models have been applied to plain radiography, ultrasounds, and MRI for assessing joint damage, synovial inflammation, and disease progression in RA, OA, and SpA patients. Despite the promise of DL, challenges such as data bias, limited explainability, and the need for large annotated datasets remain significant barriers to its widespread adoption. Furthermore, human oversight and value judgment are essential for ensuring the ethical use and effective implementation of DL in clinical settings. This review provides a comprehensive overview of DL's applications in rheumatologic imaging and explores its future potential in enhancing diagnosis, treatment decisions, and personalized medicine.

Keywords: Deep learning, Artificial intelligence, Rheumatology, Diagnostic imaging

INTRODUCTION

Rheumatic diseases are a group of autoimmune and chronic inflammatory conditions that affect the connective tissues and musculoskeletal system. This group includes conditions like rheumatoid arthritis (RA), osteoarthritis (OA), and spondyloarthritis (SpA), which result in joint destruction and impaired function. Accurate diagnosis, evaluation of disease activity, and understanding of the underlying pathophysiology are crucial for improving patient outcomes. Imaging studies, such as plain radiography, musculoskeletal ultrasounds, computed tomography (CT), and magnetic resonance imaging (MRI), play a key role in

assessing joint and soft tissue involvement, monitoring disease activity, and evaluating response to therapy in various rheumatic diseases [1-3]. However, traditional imaging techniques often encounter challenges such as repetitive and time-consuming manual assessments, variability between observers and within the same observer, and the requirement for highly specialized interpretation [4-7]. These limitations underscore the need for more advanced, efficient, and reliable imaging analysis methods.

In recent years, artificial intelligence (AI) has been increasingly applied across various sectors of society, including autonomous driving, language translation, and conversational models [8-10]. The medical field is no exception. Since the US Food

Received November 4, 2024; Revised December 13, 2024; Accepted December 29, 2024; Published online January 20, 2025

Corresponding author: Sangtae Choi, <https://orcid.org/0000-0002-2074-1733>

Division of Rheumatology, Department of Internal Medicine, Chung-Ang University College of Medicine, 102 Heukseok-ro, Dongjak-gu, Seoul 06973, Korea. **E-mail:** beconst@cau.ac.kr

Copyright © The Korean College of Rheumatology.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

and Drug Administration (FDA) first approved the use of AI for medical purposes in 1995, the number of AI and machine learning (ML) algorithms receiving FDA approval has grown significantly, particularly in the past decade. By June 2024, over 950 AI/ML-enabled medical devices had been approved by the FDA [11]. Deep learning (DL) algorithms learn patterns and features from large datasets [12]. DL is already being applied in various areas of medicine, such as diagnosis, disease activity assessment, medical image interpretation, treatment selection, and drug discovery [12-14]. Among these, medical imaging analysis is advancing rapidly, with significant progress in applying DL to rheumatology image interpretation [15-17]. While DL offers potential advantages in rheumatologic imaging, such as compensating for the limitations of traditional image interpretation and improving accuracy, efficiency, and consistency, several challenges remain. These issues must be addressed when applying DL to medical image analysis [15,18].

This review aims to provide rheumatologists with a comprehensive introduction to the basic concepts of DL and examine current trends in its application to rheumatologic image analysis. We will explore how DL techniques can enhance the diagnosis, management, and research of rheumatologic diseases. Additionally, we will address the risks and challenges associated with using DL in rheumatologic imaging, emphasizing the importance of the human role in its effective implementation.

MAIN SUBJECTS

Deep learning in medical imaging analysis

In recent years, DL has become a powerful tool for medical imaging analysis, offering significant improvements in diagnostic accuracy and efficiency. DL leverages neural networks (NNs) with multiple layers—hence the term “deep”—to process vast amounts of data and identify patterns. This technology has gained prominence due to its ability to handle the complex, high-dimensional data typically associated with medical images.

1) What is deep learning?

The field of AI can be visualized as a hierarchical structure, with traditional AI forming the outermost layer, followed by ML, and finally DL at the core. Traditional AI, also known as symbolic AI, relies on rule-based systems and expert knowledge, where pre-programmed logic is used for problem-solving and decision-making. ML, a subset of AI, marks a shift from

rule-based approaches to systems that learn from data, improving performance by recognizing patterns and making predictions. DL, a further specialization within ML, employs NNs with multiple layers to model complex data patterns, excelling in tasks such as image recognition, natural language processing, and autonomous decision-making. While traditional AI focuses on rule-based logic, both ML and DL emphasize learning from data. DL pushes the boundaries of AI by processing large amounts of unstructured data. Figure 1 illustrates the AI hierarchy.

NNs are inspired by the structure and function of biological neural systems found in animal brains. They consist of interconnected nodes, analogous to neurons, connected by edges that represent synapses. These edges carry signals, and each has a weight that determines the strength of its connection. NNs typically contain multiple layers: an input layer, an output layer, and one or more hidden layers in between. The edges are directed, meaning signals flow in one direction—from the input layer, through the hidden layers, to the output layer—in a feedforward manner without cycles. Each node receives signals from the previous layer, calculates a pre-activation value, and applies an activation function to produce a post-activation value, which is then passed to the next layer. The pre-activation value is the weighted sum of the incoming signals, calculated by multiplying the incoming signal by the weight of the edge connecting the nodes. The activation function is a nonlinear mathematical operation, with a finite range, and is continuously differentiable. In DL, “learning” refers to the process of optimizing the weights of the edges to improve the network’s performance in fitting the

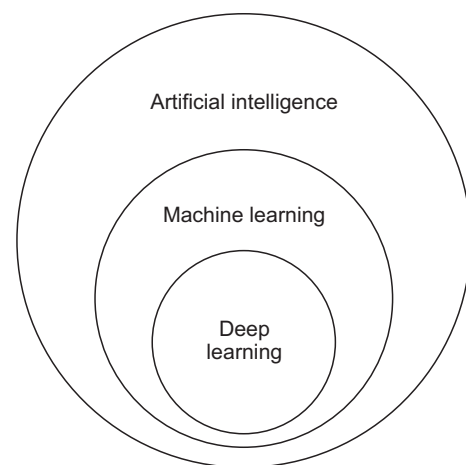


Figure 1. The hierarchy of artificial intelligence.

data. The Figure 2 shows the structure of the NNs.

NNs have achieved remarkable success across a wide range of domains. However, they also face notable challenges. One significant limitation is their susceptibility to overfitting, especially when trained on limited or imbalanced datasets. Overfitting occurs when a model memorizes the training data instead of generalizing to new, unseen data. Regularization techniques, such as dropout, weight decay, and early stopping, can mitigate this issue. Additionally, access to large, diverse, and high-quality datasets is essential for improving generalization and ensuring robust performance.

Convolutional neural networks (CNNs) are a specialized type of NN widely used for image recognition and processing tasks. Unlike traditional image analysis methods, which rely heavily on manual feature extraction, CNNs automatically learn the most relevant features directly from raw images during training, reducing the need for domain-specific expertise. One of the key advantages of CNNs is parameter sharing, which enables efficient pattern recognition in images. Instead of learning separate parameters for each pixel or region, CNNs use the same set of filters (or kernels) across different areas of the input image. As illustrated in Figures 3 and 4, this approach allows the network to use the same weights to detect features such as edges or tex-

tures, regardless of their position within the image. By sharing parameters, CNNs significantly reduce the number of parameters that need to be learned, improving both generalization and computational efficiency. This enables CNNs to recognize patterns, such as shapes or objects, in various locations within an image, making them highly effective for image recognition tasks. The combination of parameter sharing and the ability to learn feature hierarchies gives CNNs their powerful pattern rec-

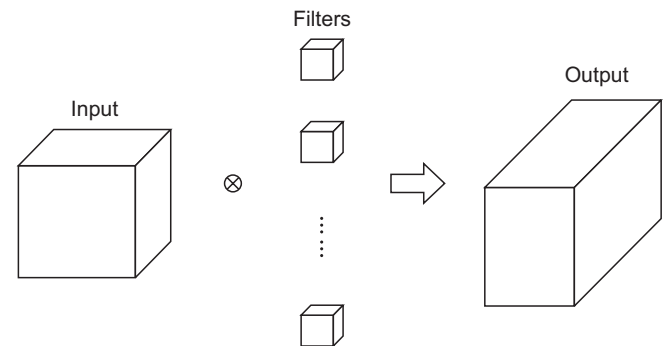


Figure 4. An example of 3D convolution operation. An RGB (red, green, and blue) color image can be viewed as a 3D tensor with three dimensions: width, height, and depth. In this process, multiple filters are typically applied to detect various patterns in the input tensor. The depth of the resulting output tensor corresponds to the number of filters used.

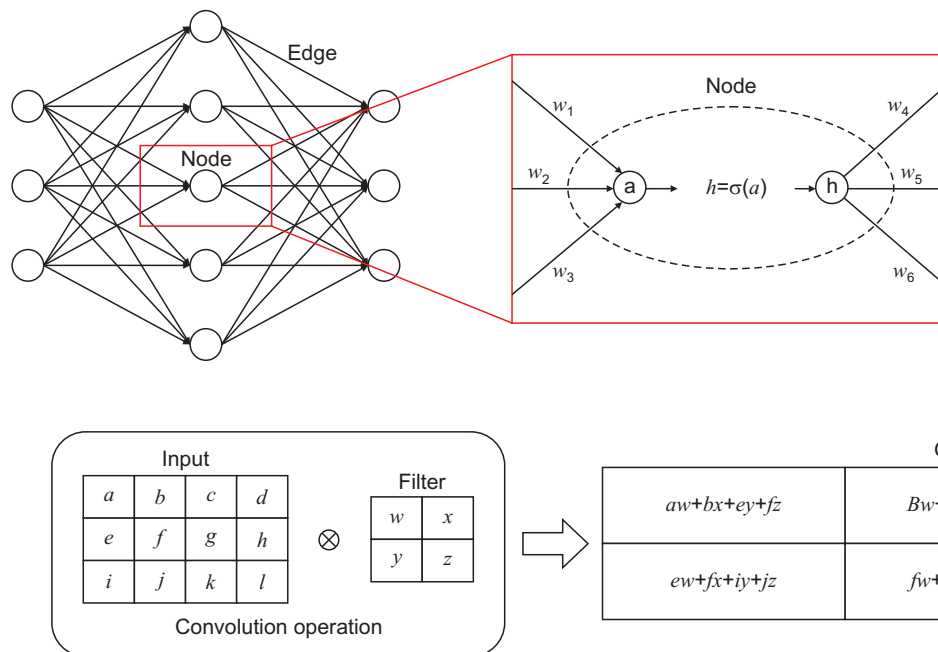


Figure 3. An example of a 2D convolution operation involves element-wise multiplication of two matrices. Here, the filter is a 2×2 matrix, and it is applied to corresponding segments of the input matrix with the same dimensions. The variables $a, b, \dots, l, w, x, y, z$ represent numerical values, and $\otimes$ denotes the convolution operator.

ognition capabilities.

Despite their strengths, CNNs are not without limitations. They can sometimes learn spurious correlations or focus on irrelevant features, such as imaging artifacts, instead of meaningful patterns. Explainability techniques can help identify whether the model is attending to the correct regions of an image, facilitating interpretation and debugging. Another limitation of CNNs is their inability to capture long-range dependencies or global context within images. This shortcoming can be addressed by integrating CNNs with transformers, which are adept at modeling global relationships. By combining these architectures, models can achieve a more holistic understanding of images, improving performance in tasks requiring global context.

While DL has transformed the field of AI with their powerful capabilities, addressing their limitations through robust regularization, explainability techniques, and advanced architectural innovations remains crucial. For further details on DL and its applications [19,20].

2) Impact of deep learning in medical imaging

Medical imaging has always been at the forefront of diagnostic medicine, but analyzing these images has traditionally been a time-consuming and error-prone process due to human limitations such as fatigue, bias, or subtle differences in pathology presentation. DL addresses these challenges by providing rapid and objective analysis, often surpassing human performance in certain tasks.

Several studies demonstrate the capabilities of DL in medical imaging. For instance, Gulshan et al. [21] showed that a DL system for detecting diabetic retinopathy in retinal fundus photographs performed comparably to ophthalmologists. Similarly, Esteva et al. [22] used a CNN model to classify skin cancer from dermoscopic images with performance on par with dermatologists. These studies underline the potential of DL to enhance diagnostic accuracy, reduce the workload of healthcare providers, and enable early detection of diseases.

One area where DL has had a particularly transformative effect is in radiology. Traditional radiological interpretation can be subjective, but DL can assist radiologists by providing more consistent and quantitative assessments. For example, in the detection of lung nodules in CT scans, models trained using datasets such as the LUNA16 dataset (Lung Image Database Consortium and Image Database Resource Initiative [LIDC-IDRI]) have shown great promise in early lung cancer screening, potentially

reducing mortality through earlier intervention [23].

Current applications of deep learning in imaging analysis

1) Tumor detection and classification

DL models have shown impressive results in identifying and classifying various types of tumors. For instance, researchers have used CNNs to analyze mammograms for breast cancer detection, achieving accuracy comparable to expert radiologists [24]. Similarly, DL has been applied to MRI scans for detecting and classifying brain tumors [25].

2) Segmentation tasks

In medical imaging, accurate segmentation of anatomical structures and pathological regions is crucial for diagnosis and treatment planning. DL-based segmentation models, such as the U-Net (U-shaped convolutional neural network) architecture (University of Freiburg, Freiburg, Germany), have been extensively used for segmenting organs, lesions, and tumors. U-Net, originally proposed by Ronneberger et al. [26], has become a benchmark model for medical image segmentation and is used in applications such as liver tumor segmentation in CT scans and cardiac segmentation in MRI scans.

3) Retinal image analysis

Retinal imaging has benefited from DL, particularly in screening for diabetic retinopathy, macular degeneration, and glaucoma. The IDx-DR system (Digital Diagnostics Inc., Coralville, IA, USA), which uses DL to analyze retinal images, was the first FDA-approved AI diagnostic tool for autonomous detection of diabetic retinopathy [27]. The system's ability to provide quick and accurate diagnoses without the need for a specialist is an example of how DL can help bridge the gap in healthcare access, particularly in underserved areas.

4) Enhancement of low-quality images

Another application of DL is in enhancing low-quality medical images. In cases where images are noisy or of low resolution, such as those obtained from portable plain radiography machines, DL can be used to improve image quality, enabling more accurate interpretation. Generative adversarial networks have been explored for this purpose, with studies showing that they can enhance the quality of low-dose CT images without com-

Table 1. Summary of references

Data modality	*Type	Body part	AI method
Rheumatoid arthritis			
Plain radiography	A	Foot	VGG [38]
		Hand	VGG [38], CNN [37], AlexNet, ResNet, DenseNet, and ViT [34], U-Net++, and ResNet [35]
	C	Cervical spine	CNN and HRNet [33]
		Hand	CNN and RetinaNet [30], YOLO [31]
MRI	A	Hand	Traditional AI [46], U-Net [50]
	C	Hand	U-Net [46], CNN and ResNet [48]
	P	Foot	ANN [51]
		Hand	U-Net [52], ANN [51]
Ultrasound	A	Hand	Traditional AI [40], VGG [42], Inception [41], CNN [43], ResNet [44]
	C	Hand/foot	Traditional AI [3]
Thermal image	C	Hand	RANet [54]
Osteoarthritis			
Plain radiography	A	Knee	PIM [63], ResNet, VGG, DenseNet, and Inception [57], ResNet [67], CNN [61], DenseNet [59], EfficientNet, RegNet, ResNet, and ResNext [55], Elastic Net and RF [58], YOLO [56]
	C	Hand	Cellular neural networks [69]
		Hip	VGG [70,71]
		Knee	CNN [64,65], Xception [62], VGG, EfficientNet, ResNet, and DenseNet [60]
		Shoulder	ResNet [72]
	P	Hip	EfficientNet [73], GAM [74], ResNet [75]
		Knee	CNN [66]
		Knee	GBM [68]
MRI	A	Knee	U-Net [79,82,85], CNN [84,87], VGG [80,81], U-Net++ [83]
	C	Knee	CNN [76,86], cGAN, U-Net [78], HNN [77]
	P	Knee	DenseNet [88]
		Hip	VGG, DenseNet, and ViT [92]
CT	C	Temporomandibular joint	YOLO [90,91]
Ultrasound	A	Hand	U-Net++ [93]
Rheumatoid arthritis /osteoarthritis			
Plain radiography	C	Hand	EfficientNet [32]
Spondyloarthritis			
Plain radiography	C	SIJ	DenseNet [113], Traditional AI [6], ResNet [110], MobileNet [111]
	P	SIJ	ANN [117], GLM, SVM, and RF [118], XGBoost [120], DBARTS and GAM [119]
		Spine	RF and XGBoost [120], RF, DBARTS, and GAM [119]
	A	SIJ	U-Net [97,105], Attention U-Net [108], ResNet [103]
	C	SIJ	U-Net [104], XGBoost [106], VGG [100], ANN [98], ResNet [96,102], DenseNet [107], SVM [95]
		Spine	Attention U-Net [109]
CT	P	SIJ	EfficientNet, U-Net, and ResNet [101]
	C	SIJ	CNN [116], U-Net [114,115]

Table 1. Continued

Data modality	*Type	Body part	AI method
Others			
CT (systemic sclerosis)	P	Lung	AtlasNet [125]
DECT (gout)	C	Foot/hand	SegResNet and Swin UNETR [127]
MRI (avascular necrosis)	A	Hip	VGG-16, ResNet, and Inception [126]
MRI (myositis)	A	Muscle	V-Net [122]
	C	Muscle	ANN [121]
Nailfold capillaroscopy (systemic sclerosis)	C	Nailfold capillary	ViT [123], U-Net and ResNet [124]
Ultrasound (carpal tunnel syndrome)	C	Median nerve	ResNet [128]

AI: artificial intelligence, VGG: visual geometry group, CNN: convolution neural networks, ResNet: residual network, DenseNet: densely connected convolutional networks, ViT: vision transformer, U-Net: U-shaped convolutional neural network, HRNet: high-resolution network, RetinaNet: retina network, YOLO: you look only once, ANN: artificial neural network, RANet: region attention network, PIM: plug-in-module, EfficientNet: efficient network, RegNet: self-regulated network, ResNext: residual networks with external transformations, Elastic Net: elastic net regularization, RF: random forest, Xception: extreme inception, GAM: generalized additive models, GBM: boosting machine, cGAN: conditional generative adversarial network, HNN: holistically nested networks, MobileNet: mobile network, GLM: generalized linear model, SVM: support vector machine, XGBoost: extreme gradient boosting, DBARTS: Discrete Bayesian Additive Regression Trees, AtlasNet: atlas network, Segresnet: segmentation residual network, SwinUNETR: shifted window U-Net transformer, V-Net: V-shaped convolutional neural network. *A: disease activity or severity assessment, C: diagnosis or classification, P: prediction of disease progress.

promising diagnostic accuracy [28].

DL in rheumatic diseases

1) Rheumatoid arthritis

To date, most DL research on images of patients with RA has focused on plain radiography, ultrasonography, and MRI, particularly on the hands and wrists, with a few studies also examining the feet (Table 1) [15].

Plain radiography image analysis has primarily concentrated on detecting and classifying joint damage, including the presence or absence of bone erosions, as well as the automatic segmentation of specific joints or regions of interest (ROI) [29-31]. Published DL methods include distinguishing RA from OA on hand plain radiography and detection models for atlantoaxial subluxation on cervical spine plain radiography in RA patients [32,33]. Joint damage has been evaluated using scores like the erosion score based on the modified Sharp van der Heide method and the joint space narrowing score [34-37], or the Ratingen method to assess joint damage severity [38].

In ultrasonography studies, disease activity has been primarily evaluated using images of the hand. Synovial inflammation has been scored according to the EULAR-OMERACT synovitis scoring system, which assesses synovial hypertrophy and/or blood flow in the synovium [39,40]. DL-based methods for evaluating synovitis typically involve detecting the ROI, specifi-

cally the region of synovitis, followed by grading through DL techniques. Alternatively, entire images have been analyzed using regression networks with various DL models, including cascade CNNs, deep CNNs, and residual network [41-45].

MRI studies are also ongoing, as MRI is highly sensitive to early inflammatory changes in joints that are not detectable on conventional plain radiography. DL-based MRI analyses have been published, focusing on classifying bone marrow edema (BME), synovitis, erosions, osteitis, and other soft tissue abnormalities in early RA [46-48]. While most studies have concentrated on classification or segmentation [49], one study reported the quantification of tenosynovitis in wrist joints of RA patients using DL [50]. Moreover, DL models have been developed to predict progression to RA using extremity MRI scans in early arthritis and clinically suspected arthralgia [51]. Research is also exploring models that automatically detect changes in RA patients before and after treatment by comparing serial MRI scans [52,53].

Despite this progress, DL studies on RA images remain largely focused on the hands, particularly the phalanges and some wrist joints, with limitations such as partial manual selection of the ROI. Additionally, the range of imaging modalities explored remains narrow, although a CNN model for detecting RA in hand thermal images has been reported [54]. However, future advancements in DL could move beyond simple disease detection and activity assessment to cover a broader range of joints.

These developments could contribute to disease monitoring, predicting disease progression, differentiating RA from other conditions, and guiding treatment decisions by integrating clinical data.

2) Osteoarthritis

OA is the most actively researched condition in the field of rheumatologic imaging analysis using DL, with the majority of studies focusing on knee OA [15]. In studies using plain radiography, several approaches have been employed to first detect the ROI around the knee joint, classify knee OA, and then use regression NNs to automatically assign a Kellgren-Lawrence grade, which determines the severity of knee OA [55-65]. Some DL studies have also integrated plain radiography images with clinical data and MRI images to predict knee OA progression or the need for total knee replacement [66-68]. Beyond the knee, DL research has been conducted to detect osteophytes in hand radiographs [69], as well as identify hip and shoulder OA using plain radiography [70-72]. More recently, DL-based prediction models for total hip replacement have been published [73-75].

Early DL studies involving MRI images of OA patients primarily focused on detecting bone or cartilage abnormalities [76-78]. Later, segmentation studies emerged, aiming to measure cartilage thickness, bone flattening, and synovial fluid volume [78-84]. Various other DL research using MRI images has been conducted, including the use of 3D-CNN models for segmenting and staging the severity of meniscus and patellar cartilage lesions [85], quantifying femoral and tibial cartilage degeneration [86], and grading OA patients based on morphological phenotypes [87]. Additionally, DL models have been developed to predict the progression of knee OA using MRI images [88,89].

DL studies involving imaging modalities beyond plain radiography and MRI have also been explored recently. These include the classification of temporomandibular joint OA [90,91], the grading of hip OA on CT scans [92], and the development of a DL model for segmenting and scoring the severity of osteophytes in hand OA using ultrasound images [93].

3) Spondyloarthritis

In axial spondyloarthritis (axSpA), the diagnosis and grading of sacroiliitis using plain radiography often show low inter-clinician agreement, making early diagnosis difficult and complicating the assessment of disease activity. This is why MRI imaging is widely used in clinical practice [6,94]. The most actively

researched area in DL studies related to axSpA focuses on MRI image analysis, where various AI models are being developed for early detection, disease activity evaluation, and disease prediction in axSpA patients.

Early research aimed at distinguishing axSpA patients from healthy individuals by identifying BME in the sacroiliac joint, applying DL networks to classify active sacroiliitis [95,96]. Subsequent studies have concentrated on the automatic detection of abnormal lesions in the sacroiliac joint [97-102], quantifying BME to assess disease activity [103-105], and improving diagnostic accuracy for sacroiliitis on MRI using radiomics approaches [106,107]. Additionally, algorithms have been developed to classify both BME and synovitis in axSpA patients [97,103], and a DL model for the automatic segmentation of fat metaplasia has been introduced [108]. Other methods using deep NNs have been created to detect inflammatory or structural changes in the sacroiliac joint, which are indicative of axSpA [105], as well as a deep NN for detecting spinal inflammation in MRI STIR (Short Tau Inversion Recovery) images [109].

Beyond MRI, DL has also been applied to analyze plain radiography [110-113], and there have been studies focused on detecting erosions or ankylosis using CT scans [114-116]. Additionally, research has been conducted to predict radiographic spinal progression in axSpA by evaluating the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) in combination with clinical data [117-120].

4) Other rheumatologic conditions

Although less numerous than studies on more common diseases, DL applications in image analysis are also being explored in other rheumatic conditions. DL has been employed to analyze MRI scans for the quantitative assessment of muscular inflammation in myositis and to differentiate myositis from other types of muscular dystrophy [121,122]. Researchers have also applied DL techniques to analyze nailfold capillaroscopy images for the classification and quantification of microvasculopathies [123,124], and a DL-based method has been developed for risk stratification by quantifying the extent of interstitial lung disease on high-resolution CT scans in systemic sclerosis [125].

Furthermore, two studies have focused on detecting avascular necrosis in the hip using MRI and in the shoulder using plain radiography [72,126]. Other notable research includes identifying the optimal DL algorithm for segmenting green-encoded pixels of monosodium urate (MSU) crystals on dual-energy CT

scans, using two U-Net-based models to detect MSU deposits in gout patients [127]. Additionally, a CNN has been developed to analyze ultrasound images of the median nerve for the classification of carpal tunnel syndrome [128].

Benefits of using deep learning for image interpretation in rheumatology

DL technology, which has advanced rapidly in recent years, is driving revolutionary changes not only in various aspects of daily life but also in medicine. DL is expected to become widely used in rheumatologic imaging in the near future. Currently, DL is primarily being developed for tasks such as diagnosis and classification, assessment of disease activity or severity, and prediction of disease progression (Table 1). In the future, these technologies will likely be applied more broadly in fields such as research and education, beyond patient care. Additionally, DL could evolve into a tool for precision medicine through multimodal analysis, combining clinical data with image analysis [129].

The main advantages of using DL in medical imaging are summarized in the following sub-sections.

1) Accuracy

A common challenge in medical image interpretation is the variability between and within observers [6,7]. Factors such as the type of disease, imaging modality, and the interpreter's experience can affect consistency. DL algorithms can provide more precise and consistent analysis of medical images, enhancing diagnostic accuracy and reliability. For example, the EULAR-OMERACT synovitis score is used to assess disease activity in RA patients via ultrasonography, or the modified total Sharp score is used to evaluate joint damage progression on plain radiography images. DL can significantly reduce inter-observer variability in such cases. Furthermore, while the human eye struggles to objectively quantify pixel brightness, DL can accurately quantify absolute signal intensities [15]. These capabilities allow DL to surpass the intuitive differentiation abilities of humans when interpreting images, enabling the detection of subtle findings that might otherwise be missed. By combining precision and consistency, DL offers a transformative advantage in medical image interpretation.

2) Efficiency

Medical image interpretation is often a repetitive and time-

consuming task. AI's ability to process large volumes of data quickly is a key advantage. Segmentation, classification, and regression of medical images have already been partially or fully automated. As DL-based image interpretation technologies continue to evolve, they could drastically improve time and cost efficiency in diagnosing rheumatologic diseases and assessing disease activity, reducing the workload of clinicians, including radiologists and rheumatologists [129].

3) Personalized medicine

Precision medicine, a personalized approach to healthcare, is one of the key directions for future medical advancement. This requires integrating a wide range of data, including imaging analysis, genetic information, clinical history, biomarkers, and predictions of therapeutic responses. Handling this complex data is essential for patient-specific treatment. Although still in its early stages, DL-based research is being conducted to analyze medical imaging alongside clinical information to predict disease progression or treatment responses (Table 1). This approach could ultimately support personalized decision-making when selecting treatments for individual patients.

4) Advanced pattern recognition in research

DL models excel at analyzing large, multidimensional datasets, leveraging probabilistic approaches to identify complex patterns within vast amounts of imaging data. Unlike traditional methods, DL operates without inherent bias, enabling it to uncover previously undetected diseases or reveal new aspects of disease pathophysiology. By integrating DL into the classification of medical imaging and clinical data, more objective and precise evaluations of disease activity, severity, progression, and remission can be achieved [15]. DL in image interpretation could also advance rheumatology research by providing new insights, identifying novel imaging biomarkers, and deepening our understanding of disease progression and treatment responses [130].

5) Addressing healthcare disparities

Global healthcare disparities, particularly the shortage of specialized medical professionals in underdeveloped regions, are a significant challenge [131,132]. In such areas, general practitioners often handle tasks typically performed by specialists, such as diagnosing rheumatic diseases. Although it may not provide a complete solution, DL-based imaging analysis could assist in

diagnosing and treating patients in regions without access to specialists, helping to reduce global healthcare disparities [22].

The limitation of deep learning and the role of human

DL has transformed medical image analysis, significantly improving accuracy and enabling the automation of tasks such as disease detection and diagnosis. Despite these advancements, DL models developed in academic research have yet to achieve widespread application in clinical practice [133]. DL models still have notable limitations, particularly in areas like interpretability, generalization, and the need for large annotated datasets. These models can struggle to explain their predictions, making it difficult for healthcare professionals to fully trust or understand their decisions [134]. Moreover, DL systems can suffer from overfitting and poor performance when applied to out-of-distribution data, highlighting the role of human expertise in verifying and contextualizing results [135]. The dependency on massive labeled datasets also necessitates human involvement in data annotation, where domain-specific knowledge is essential to ensure quality and accuracy [136]. The practical deployment of DL models is also hindered by data availability challenges. The data used to develop these models is often not easily accessible in standard clinical environments, limiting their usability [133]. While DL models excel in pattern recognition, human experts are crucial for addressing ethical, clinical, and interpretative challenges in medical imaging.

1) Bias in input data

High-quality data is crucial for developing accurate DL models, as errors or biases in input data can lead to skewed outcomes. If training images are biased by confounders such as age, sex, race, or the distribution of disease types and severity, the DL model's results may be inaccurate [15]. The type of data used for training, especially in MRI image analysis, is another critical factor. Variations in MRI machines, imaging protocols, pixel size, and data heterogeneity across institutions can influence DL model performance [137]. Additionally, background features can inadvertently affect classification, leading to background bias, where differences between two groups are misinterpreted as lesion-related differences. Ensuring high-quality, unbiased data and accurate ground truth labeling is crucial, but these can be difficult to achieve in cases with high inter- and intra-observer variability [15,138]. Data leakage during testing can also inflate performance metrics, highlighting the need for care-

ful validation processes.

2) Limited data in rheumatologic diseases

While DL models traditionally required large datasets, advancements such as transfer learning and data augmentation have made it possible to train models with smaller amounts of data [139,140]. However, large, high-quality datasets still improve DL model performance. In rheumatologic imaging, data scarcity remains a significant challenge, as labeled datasets are often limited compared to conditions like malignancies or cardiovascular diseases. This scarcity increases the risk of overfitting and makes generalization to new data difficult, especially for rare rheumatic diseases. As a result, current DL research has primarily focused on more common diseases like OA, RA, and axSpA (Table 1).

3) Lack of explainability

DL models, particularly in medical imaging, are often referred to as black boxes, because their decision-making processes are opaque and difficult to interpret [141]. These models process input data through multiple layers of mathematical transformations, making it challenging to understand the specific features or patterns they rely on for predictions. While DL models excel at tasks like lesion detection and disease classification, the underlying reasoning behind their decisions remains unclear.

This lack of explainability is a significant concern in medical imaging, where clinicians need to trust and understand the basis for diagnostic decisions. The complexity of NNs makes it difficult to identify which features in the image are driving the model's output, raising concerns that the model could be focusing on irrelevant or biased aspects, such as imaging artifacts. This can lead to inaccurate or biased diagnoses if the model is influenced by clinically insignificant features.

Efforts to enhance explainability in DL models are ongoing, especially in the critical area of image reconstruction, where interpretability remains a persistent challenge. Visualization techniques, including saliency maps, class activation maps, deconvolution networks, and optimization-based methods, offer insights into model decision-making by highlighting image regions most influential to predictions [142]. Layer-wise relevance propagation builds on these approaches, generating heatmaps to pinpoint areas with the greatest impact on model outputs [143].

Despite these advancements, several significant challenges hinder the deployment of DL models in image reconstruction,

particularly in healthcare settings. For example, DL models are often sensitive to variations in image intensity and inconsistencies in imaging protocols, which can result in misinterpretations [15]. Unlike human experts, who can adapt to such variability, DL models require carefully calibrated and standardized input data. Additionally, the limited availability of large, high-quality medical datasets poses another critical barrier. In contrast to general-purpose DL applications that leverage extensive public datasets, medical imaging datasets are often constrained by privacy and ethical considerations. This scarcity not only increases the risk of overfitting but also limits the generalizability of DL models across different imaging modalities and patient populations.

4) Validation and generalizability

Validation and generalizability are crucial for the successful integration of DL models into clinical care. While internal validation methods like cross-validation are useful, they can still suffer from bias in the training data, and as a result, the quality of a DL model may not be fully assured. Therefore, external validation—using datasets from different institutions and sources—is essential for assessing the robustness of a DL model [144]. For reliable implementation in clinical settings, DL models should undergo validation with large-scale datasets that reflect diverse populations, and external multi-center datasets should be used to independently test the algorithms. It is crucial to compare DL algorithms systematically and apply randomized controlled trials, similar to those used in other clinical studies, to establish their effectiveness and reliability [145]. These approaches ensure that DL models can be trusted to perform consistently in real-world settings [15,129].

Generalizability remains a significant challenge. One of the primary barriers to applying DL-based decision-making in medicine is the difficulty models face in generalizing across different institutions, which may have varying patient populations, imaging machines, or protocols [146]. Even when MRI machines or imaging protocols differ across institutions, it is critical for DL models to maintain accuracy and consistency. This can be more difficult for DL models than for human experts, as common issues such as overfitting and underspecification hinder generalizability [145,147]. To address this, ongoing research focuses on improving model robustness across diverse clinical environments.

Additionally, after validation, continuous monitoring and

updating of models are necessary to ensure quality assurance. When new research findings or changes to clinical criteria for diagnosis and disease activity emerge, DL models must be modified and revalidated to ensure continued reliability and effectiveness [148].

5) Value judgment

Despite the rapid advancement of DL in medical imaging, it cannot replace the role of human judgment. DL models rely on probabilistic calculations, while human clinicians apply value-based judgments, taking into account multiple clinical factors. For instance, in assessing RA disease activity, clinicians integrate a variety of indicators, such as the disease activity score 28 (DAS28), simplified disease activity index (SDAI), and clinical disease activity index (CDAI), depending on the specific context of treatment [149]. These indices require the weighing of variables based on clinical judgment, which DL models cannot fully replicate. Different DL models may yield different outcomes depending on how data is selected, weighted, and analyzed, so human decision-making remains vital [150].

Moreover, when DL models integrate medical images with clinical data to recommend treatment options, results can vary depending on the treatment guidelines used. Ultimately, human clinicians must make the final decisions, ensuring that AI tools serve as supportive aids rather than as autonomous decision-makers [148]. The role of thoughtful physicians will continue to be essential, even with the growing use of AI in patient care.

6) Ethical and legal issues

The use of DL in rheumatologic imaging raises important ethical and legal considerations. Key questions include determining responsibility in the event of errors during patient care, setting boundaries for AI use in medical research, and defining who has the authority to use specific DL models [151,152]. Additionally, the economic aspects of DL-based decision-making, such as determining who benefits financially, need careful evaluation [153]. Protecting patient privacy is another critical concern, particularly as DL involves handling large datasets, increasing the risk of data breaches [154]. To effectively implement DL technology in real-world medical settings, appropriate legal regulations must be in place, alongside addressing these ethical issues. Ultimately, the goal of using DL in medicine should be to enhance human values and dignity by improving patient care.

CONCLUSION

The development of DL models in rheumatologic imaging analysis has advanced rapidly in recent years. While current efforts have largely focused on building classification and prediction models for RA, OA, and SpA, the field is expected to evolve toward multimodal analysis, which will integrate various types of clinical, imaging, and genomic data. Despite some existing limitations, DL is poised to become a vital tool in rheumatologic patient care and research. However, the successful integration of DL into clinical practice will require rigorous validation to ensure accuracy and generalizability across different settings. Moreover, value judgment will remain essential, as human expertise is necessary to interpret results, guide treatment decisions, and ensure that AI complements clinical judgment rather than replacing it. Ethical considerations—such as maintaining patient privacy, ensuring fairness, and determining accountability—must also be addressed. In conclusion, while DL will play an increasingly significant role in healthcare, the involvement of human clinicians is crucial in maintaining oversight, ensuring ethical use, and ultimately guiding the application of these technologies in a way that enhances patient outcomes.

FUNDING

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (RS-2025-00519376).

ACKNOWLEDGMENTS

None.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

AUTHOR CONTRIBUTIONS

J.M. and S.C. conducted the investigation and prepared the original draft. P.J. summarized the references. J.M., P.J., and S.C. reviewed and revised the manuscript. S.C. supervised the work. All authors approved the final version of the manuscript.

ORCID

Jucheol Moon, <https://orcid.org/0000-0002-2885-0627>

Pratik Jadhav, <https://orcid.org/0009-0003-3386-6273>

Sangtae Choi, <https://orcid.org/0000-0002-2074-1733>

REFERENCES

1. Mohana-Borges AVR, Chung CB. Imaging of rheumatic diseases affecting the lower limb. *Rheum Dis Clin North Am* 2024;50:463-82.
2. Braun J, Kiltz U, Baraliakos X. Significance of structural changes in the sacroiliac joints of patients with axial spondyloarthritis detected by MRI related to patients symptoms and functioning. *Ann Rheum Dis* 2022;81:11-4.
3. Filippucci E, Cipolletta E, Mashadi Mirza R, Carotti M, Giovagnoni A, Salaffi F, et al. Ultrasound imaging in rheumatoid arthritis. *Radiol Med* 2019;124:1087-100.
4. Piórkowski A, Obuchowicz R, Urbanik A, Strzelecki M. Advances in musculoskeletal imaging and their applications. *J Clin Med* 2023;12:6585.
5. Deodhar A, Strand V, Kay J, Braun J. The term 'non-radiographic axial spondyloarthritis' is much more important to classify than to diagnose patients with axial spondyloarthritis. *Ann Rheum Dis* 2016;75:791-4.
6. van Tubergen A, Heuft-Dorenbosch L, Schulpen G, Landewé R, Wijers R, van der Heijde D, et al. Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality? *Ann Rheum Dis* 2003;62:519-25.
7. van den Berg R, Lenczner G, Thévenin F, Claudepierre P, Feydy A, Reijnierse M, et al. Classification of axial SpA based on positive imaging (radiographs and/or MRI of the sacroiliac joints) by local rheumatologists or radiologists versus central trained readers in the DESIR cohort. *Ann Rheum Dis* 2015;74:2016-21.
8. Picazo-Sanchez P, Ortiz-Martin L. Analysing the impact of Chat-GPT in research. *Appl Intell* 2024;54:4172-88.
9. Yang J, Gao S, Qiu Y, Chen L, Li T, Dai B, et al. Generalized predictive model for autonomous driving. Paper presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2024 Jun 16-22; Seattle, USA. p. 14662-72.
10. Xu H, Kim YJ, Sharaf A, Awadalla HH. A paradigm shift in machine translation: boosting translation performance of large language models. *arXiv [Online]*. 2023 [cited 2024 Oct 24]. Available from: <https://doi.org/10.48550/arXiv.2309.11674>.
11. U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML)-enabled medical devices [Internet]. Silver Spring: U.S. Food and Drug Administration, c2024 [cited 2024 Oct 24]. Available from: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>.
12. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
13. Koetzier LR, Mastrodicasa D, Szczykutowicz TP, van der Werf NR, Wang AS, Sandfort V, et al. Deep learning image reconstruc-

- tion for CT: technical principles and clinical prospects. *Radiology* 2023;306:e221257.
14. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers* 2021;25:1315-60.
15. Stoel BC, Staring M, Reijnierse M, van der Helm-van Mil AHM. Deep learning in rheumatological image interpretation. *Nat Rev Rheumatol* 2024;20:182-95.
16. Botnari A, Kadar M, Patrascu JM. A comprehensive evaluation of deep learning models on knee MRIs for the diagnosis and classification of meniscal tears: a systematic review and meta-analysis. *Diagnostics (Basel)* 2024;14:1090.
17. Fritz B, Fritz J. Artificial intelligence for MRI diagnosis of joints: a scoping review of the current state-of-the-art of deep learning-based approaches. *Skeletal Radiol* 2022;51:315-29.
18. McMaster C, Bird A, Liew DFL, Buchanan RR, Owen CE, Chapman WW, et al. Artificial intelligence and deep learning for rheumatologists. *Arthritis Rheumatol* 2022;74:1893-905.
19. Aggarwal CC. *Neural networks and deep learning*. Vol. 10. Cham, Springer, 2018.
20. Goodfellow I, Bengio Y, Courville A. *Deep learning*. Cambridge, MIT press, 2016.
21. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402-10.
22. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8.
23. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard CVD, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Med Image Anal* 2017;42:1-13.
24. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafi H, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89-94.
25. Pérez-García F, Sparks R, Ourselin S. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput Methods Programs Biomed* 2021;208:106236.
26. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Paper presented at: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 2015 Oct 5-9; Munich, Germany. p. 234-41.
27. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018;1:39.
28. Wolterink JM, Leiner T, Viergever MA, Isgum I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging* 2017;36:2536-45.
29. Radke KL, Kors M, Müller-Lutz A, Frenken M, Wilms LM, Baraliakos X, et al. Adaptive IoU thresholding for improving small object detection: a proof-of-concept study of hand erosions classification of patients with rheumatic arthritis on X-ray images. *Diagnostics (Basel)* 2022;13:104.
30. Murakami S, Hatano K, Tan J, Kim H, Aoki T. Automatic identification of bone erosions in rheumatoid arthritis from hand radiographs based on deep convolutional neural network. *Multimed Tools Appl* 2018;77:10921-37.
31. Fung DL, Liu Q, Islam S, Lac L, O'Neil L, Hitchon CA, et al. Deep learning-based joint detection in rheumatoid arthritis hand radiographs. *AMIA Jt Summits Transl Sci Proc* 2023;2023:206-15.
32. Ma Y, Pan I, Kim SY, Wieschhoff GG, Andriole KP, Mandell JC. Deep learning discrimination of rheumatoid arthritis from osteoarthritis on hand radiography. *Skeletal Radiol* 2024;53:377-83.
33. Okita Y, Hirano T, Wang B, Nakashima Y, Minoda S, Nagahara H, et al. Automatic evaluation of atlantoaxial subluxation in rheumatoid arthritis by a deep learning model. *Arthritis Res Ther* 2023;25:181.
34. Izumi K, Suzuki K, Hashimoto M, Endoh T, Doi K, Iwai Y, et al. Detecting hand joint ankylosis and subluxation in radiographic images using deep learning: a step in the development of an automatic radiographic scoring system for joint destruction. *PLoS One* 2023;18:e0281088.
35. Wang H, Ou Y, Fang W, Ambalathankandy P, Goto N, Ota G, et al. A deep registration method for accurate quantification of joint space narrowing progression in rheumatoid arthritis. *Comput Med Imaging Graph* 2023;108:102273.
36. Üreten K, Erbay H, Maraş HH. Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network. *Clin Rheumatol* 2020;39:969-74.
37. Hirano T, Nishide M, Nonaka N, Seita J, Ebina K, Sakurada K, et al. Development and validation of a deep-learning model for scoring of radiographic finger joint destruction in rheumatoid arthritis. *Rheumatol Adv Pract* 2019;3:rkz047.
38. Rohrbach J, Reinhard T, Sick B, Dürr O. Bone erosion scoring for rheumatoid arthritis with deep convolutional neural networks. *Comput Electr Eng* 2019;78:472-81.
39. Terslev L, Naredo E, Aegerter P, Wakefield RJ, Backhaus M, Balint P, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce-part 2: reliability and application to multiple joints of a standardised consensus-based scoring system. *RMD Open* 2017;3:e000427.
40. D'Agostino MA, Terslev L, Aegerter P, Backhaus M, Balint P, Bruyn GA, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce-part 1: definition and development of a standardised, consensus-based scoring system. *RMD Open* 2017;3:e000428.
41. Tang J, Jin Z, Zhou X, Zhang W, Wu M, Shen Q, et al. Enhancing convolutional neural network scheme for rheumatoid arthritis grading with limited clinical data. *Chin Phys B* 2019;28:038701.
42. Andersen JKH, Pedersen JS, Laursen MS, Holtz K, Grauslund J, Savarimuthu TR, et al. Neural networks for automatic scoring of arthritis disease activity on ultrasound images. *RMD Open* 2019;5:e000891.
43. Christensen ABH, Just SA, Andersen JKH, Savarimuthu TR. Applying cascaded convolutional neural network design further enhances automatic scoring of arthritis disease activity on ultrasound images from rheumatoid arthritis patients. *Ann Rheum Dis* 2020;79:1189-93.

44. He X, Wang M, Zhao C, Wang Q, Zhang R, Liu J, et al. Deep learning-based automatic scoring models for the disease activity of rheumatoid arthritis based on multimodal ultrasound images. *Rheumatology (Oxford)* 2024;63:866-73.
45. Hemalatha RJ, Vijaybaskar V, Thamizhvan 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.
46. Wong LM, Shi L, Xiao F, Griffith JE. Fully automated segmentation of wrist bones on T2-weighted fat-suppressed MR images in early rheumatoid arthritis. *Quant Imaging Med Surg* 2019;9:579-89.
47. Crowley AR, Dong J, McHaffie A, Clarke AW, Reeves Q, Williams M, et al. Measuring bone erosion and edema in rheumatoid arthritis: a comparison of manual segmentation and RAMRIS methods. *J Magn Reson Imaging* 2011;33:364-71.
48. Schlereth M, Mutlu MY, Utz J, Bayat S, Heimann T, Qiu J, et al. Deep learning-based classification of erosion, synovitis and osteitis in hand MRI of patients with inflammatory arthritis. *RMD Open* 2024;10:e004273.
49. Adams LC, Bressemer KK, Ziegeler K, Vahldiek JL, Poddubnyy D. Artificial intelligence to analyze magnetic resonance imaging in rheumatology. *Joint Bone Spine* 2024;91:105651.
50. Shamonin D, Li Y, Hassanzadeh T, Bakker ME, Reijnierse M, Van der Helm van Mil A, et al. Quantification of tenosynovitis in RA from wrist MRIs, based on deep learning. *Ann Rheum Dis* 2023;82(Suppl 1):770-1.
51. Li Y, Shamonin D, Hassanzadeh T, Reijnierse M, Van der Helm van Mil A, Stoel B. Exploring the use of artificial intelligence in predicting rheumatoid arthritis, based on extremity MR scans in early arthritis and clinically suspect arthralgia patients. *Ann Rheum Dis* 2023;82(Suppl 1):1-2.
52. Hassanzadeh T, Shamonin DP, Li Y, Krijbolder DI, Reijnierse M, Van der Helm-van Mil AHM, et al. A deep learning-based comparative MRI model to detect inflammatory changes in rheumatoid arthritis. *Biomed Signal Process Control* 2024;88:105612.
53. Hassanzadeh T, Shamonin D, Li Y, Reijnierse M, Van der Helm-van Mil A, Stoel B. Treatment effects in wrist MRIs, determined by deep learning. *BMJ* 2023;82(Suppl 1):1286.
54. Ahalya RK, Almutairi FM, Snehalatha U, Dhanraj V, Aslam SM. RANet: a custom CNN model and quantvolutional neural network for the automated detection of rheumatoid arthritis in hand thermal images. *Sci Rep* 2023;13:15638.
55. Abedin J, Antony J, McGuinness K, Moran K, O'Connor NE, Rebholz-Schuhmann D, et al. Predicting knee osteoarthritis severity: comparative modeling based on patient's data and plain X-ray images. *Sci Rep* 2019;9:5761.
56. Chen N, Feng Z, Li F, Wang H, Yu R, Jiang J, et al. A fully automatic target detection and quantification strategy based on object detection convolutional neural network YOLOv3 for one-step X-ray image grading. *Anal Methods* 2023;15:164-70.
57. Chen P, Gao L, Shi X, Allen K, Yang L. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. *Comput Med Imaging Graph* 2019;75:84-92.
58. Liu B, Luo J, Huang H. Toward automatic quantification of knee osteoarthritis severity using improved Faster R-CNN. *Int J Comput Assist Radiol Surg* 2020;15:457-66.
59. Norman B, Pedroia V, Noworolski A, Link TM, Majumdar S. Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs. *J Digit Imaging* 2019;32:471-7.
60. Fei M, Lu S, Chung JH, Hassan S, Elsisy J, Schneiderman BA. Diagnosing the severity of knee osteoarthritis using regression scores from artificial intelligence convolution neural networks. *Orthopedics* 2024;47:e247-54.
61. Westbury LD, Fuggle NR, Pereira D, Oka H, Yoshimura N, Oe N, et al. Machine learning as an adjunct to expert observation in classification of radiographic knee osteoarthritis: findings from the Hertfordshire Cohort Study. *Aging Clin Exp Res* 2023;35:1449-57.
62. Touahema S, Zaimi I, Zrira N, Ngote MN, Doulhousne H, Aouial M. MedKnee: a new deep learning-based software for automated prediction of radiographic knee osteoarthritis. *Diagnostics (Basel)* 2024;14:993.
63. Lee DW, Song DS, Han HS, Ro DH. Accurate, automated classification of radiographic knee osteoarthritis severity using a novel method of deep learning: plug-in modules. *Knee Surg Relat Res* 2024;36:24.
64. Naguib SM, Kassem MA, Hamza HM, Fouda MM, Saleh MK, Hosny KM. Automated system for classifying uni-bicompartamental knee osteoarthritis by using redefined residual learning with convolutional neural network. *Heliyon* 2024;10:e31017.
65. Subha B, Jeyakumar V, Deepa SN. Gaussian Aquila optimizer based dual convolutional neural networks for identification and grading of osteoarthritis using knee joint images. *Sci Rep* 2024;14:7225.
66. Yin R, Chen H, Tao T, Zhang K, Yang G, Shi F, et al. Expanding from unilateral to bilateral: a robust deep learning-based approach for predicting radiographic osteoarthritis progression. *Osteoarthritis Cartilage* 2024;32:338-47.
67. Leung K, Zhang B, Tan J, Shen Y, Geras KJ, Babb JS, et al. Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: data from the osteoarthritis initiative. *Radiology* 2020;296:584-93.
68. Tiulpin A, Klein S, Bierma-Zeinstra SMA, Thevenot J, Rahtu E, Meurs JV, et al. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci Rep* 2019;9:20038.
69. Banerjee S, Bhunia S, Schaefer G. Osteophyte detection for hand osteoarthritis identification in X-ray images using CNNs. *Annu Int Conf IEEE Eng Med Biol Soc* 2011;2011:6196-9.
70. Üreten K, Arslan T, Gültekin KE, Demir AND, Özer HF, Bilgili Y. Detection of hip osteoarthritis by using plain pelvic radiographs with deep learning methods. *Skeletal Radiol* 2020;49:1369-74.
71. Xue Y, Zhang R, Deng Y, Chen K, Jiang T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLoS One* 2017;12:e0178992.
72. Magnéli M, Axenhus M, Fagrell J, Ling P, Gislén J, Demir Y, et al. Artificial intelligence can be used in the identification and classification of shoulder osteoarthritis and avascular necrosis on plain radiographs: a training study of 7,139 radiograph sets. *Acta Orthop* 2024;95:319-24.
73. Xu Y, Xiong H, Liu W, Liu H, Guo J, Wang W, et al. Development and

- validation of a deep-learning model to predict total hip replacement on radiographs. *J Bone Joint Surg Am* 2024;106:389-96.
74. Jang SJ, Fontana MA, Kunze KN, Anderson CG, Sculco TP, Mayman DJ, et al. An interpretable machine learning model for predicting 10-year total hip arthroplasty risk. *J Arthroplasty* 2023;38(7S):S44-50.e6.
75. Chen CC, Wu CT, Chen CPC, Chung CY, Chen SC, Lee MS, et al. Predicting the risk of total hip replacement by using a deep learning algorithm on plain pelvic radiographs: diagnostic study. *JMIR Form Res* 2023;7:e42788.
76. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Med Image Comput Comput Assist Interv* 2013;16(Pt 2):246-53.
77. Cheng R, Alexandridi NA, Smith RM, Shen A, Gandler W, McCree E, et al. Fully automated patellofemoral MRI segmentation using holistically nested networks: implications for evaluating patellofemoral osteoarthritis, pain, injury, pathology, and adolescent development. *Magn Reson Med* 2020;83:139-53.
78. Gaj S, Yang M, Nakamura K, Li X. Automated cartilage and meniscus segmentation of knee MRI with conditional generative adversarial networks. *Magn Reson Med* 2020;84:437-49.
79. Norman B, Pedoia V, Majumdar S. Use of 2D U-net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology* 2018;288:177-85.
80. Panfilov E, Tiulpin A, Nieminen MT, Saarakkala S, Casula V. Deep learning-based segmentation of knee MRI for fully automatic subregional morphological assessment of cartilage tissues: data from the Osteoarthritis Initiative. *J Orthop Res* 2022;40:1113-24.
81. Jaremko JL, Felfeliyan B, Hareendranathan A, Thejeel B, Vanessa QL, Østergaard M, et al. Volumetric quantitative measurement of hip effusions by manual versus automated artificial intelligence techniques: an OMERACT preliminary validation study. *Semin Arthritis Rheum* 2021;51:623-6.
82. Eckstein F, Chaudhari AS, Fuerst D, Gaisberger M, Kemnitz J, Baumgartner CF, et al. Detection of differences in longitudinal cartilage thickness loss using a deep-learning automated segmentation algorithm: data from the foundation for the national institutes of health biomarkers study of the osteoarthritis initiative. *Arthritis Care Res (Hoboken)* 2022;74:929-36.
83. Guo J, Yan P, Qin Y, Liu M, Ma Y, Li J, et al. Automated measurement and grading of knee cartilage thickness: a deep learning-based approach. *Front Med (Lausanne)* 2024;11:1337993.
84. Felfeliyan B, Forkert ND, Hareendranathan A, Cornel D, Zhou Y, Kuntze G, et al. Self-supervised-RCNN for medical image segmentation with limited data annotation. *Comput Med Imaging Graph* 2023;109:102297.
85. Pedoia V, Norman B, Mehany SN, Bucknor MD, Link TM, Majumdar S. 3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. *J Magn Reson Imaging* 2019;49:400-10.
86. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, et al. Deep learning approach for evaluating knee mr images: achieving high diagnostic performance for cartilage lesion detection. *Radiology* 2018;289:160-9.
87. Namiri NK, Lee J, Astuto B, Liu F, Shah R, Majumdar S, et al. Deep learning for large scale MRI-based morphological phenotyping of osteoarthritis. *Sci Rep* 2021;11:10915.
88. Hu J, Zheng C, Yu Q, Zhong L, Yu K, Chen Y, et al. DeepKOA: a deep-learning model for predicting progression in knee osteoarthritis using multimodal magnetic resonance images from the osteoarthritis initiative. *Quant Imaging Med Surg* 2023;13:4852-66.
89. Hu J, Peng J, Zhou Z, Zhao T, Zhong L, Yu K, et al. Associating knee osteoarthritis progression with temporal-regional graph convolutional network analysis on MR images. *J Magn Reson Imaging* 2025;61:378-91.
90. Talaat WM, Shetty S, Al Bayatti S, Talaat S, Mourad L, Shetty S, et al. An artificial intelligence model for the radiographic diagnosis of osteoarthritis of the temporomandibular joint. *Sci Rep* 2023;13:15972.
91. Eşer G, Duman ŞB, Bayrakdar İŞ, Çelik Ö. Classification of temporomandibular joint osteoarthritis on cone beam computed tomography images using artificial intelligence system. *J Oral Rehabil* 2023;50:758-66.
92. Masuda M, Soufi M, Otake Y, Uemura K, Kono S, Takashima K, et al. Automatic hip osteoarthritis grading with uncertainty estimation from computed tomography using digitally-reconstructed radiographs. *Int J Comput Assist Radiol Surg* 2024;19:903-15.
93. Overgaard BS, Christensen ABH, Terslev L, Savarimuthu TR, Just SA. Artificial intelligence model for segmentation and severity scoring of osteophytes in hand osteoarthritis on ultrasound images. *Front Med (Lausanne)* 2024;11:1297088.
94. Ramiro S, Nikiphorou E, Sepriano A, Ortolan A, Webers C, Baraliakos X, et al. ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. *Ann Rheum Dis* 2023;82:19-34.
95. Faleiros MC, Nogueira-Barbosa MH, Dalto VF, Júnior JRF, Tenório APM, Luppino-Assad R, et al. Machine learning techniques for computer-aided classification of active inflammatory sacroiliitis in magnetic resonance imaging. *Adv Rheumatol* 2020;60:25.
96. Lee KH, Choi ST, Lee GY, Ha YJ, Choi SI. Method for diagnosing the bone marrow edema of sacroiliac joint in patients with axial spondyloarthritis using magnetic resonance image analysis based on deep learning. *Diagnostics (Basel)* 2021;11:1156.
97. Zheng Y, Bai C, Zhang K, Han Q, Guan Q, Liu Y, et al. Deep-learning based quantification model for hip bone marrow edema and synovitis in patients with spondyloarthritis based on magnetic resonance images. *Front Physiol* 2023;14:1132214.
98. Ożga J, Wyka M, Raczko A, Tabor Z, Oleniack Z, Korman M, et al. Performance of fully automated algorithm detecting bone marrow edema in sacroiliac joints. *J Clin Med* 2023;12:4852.
99. Lee GE, Kim SH, Cho JC, Choi ST, Choi SI. Text-guided cross-position attention for segmentation: case of medical image. Paper presented at: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023; 2023 Oct 8-12; Vancouver, Canada. p. 537-46.
100. Lee S, Jeon U, Lee JH, Kang S, Kim H, Lee J, et al. Artificial intelligence for the detection of sacroiliitis on magnetic resonance imaging in patients with axial spondyloarthritis. *Front Immunol*

- 2023;14:1278247.
101. Roels J, De Craemer AS, Renson T, de Hooge M, Gevaert A, Van Den Berghe T, et al. Machine learning pipeline for predicting bone marrow edema along the sacroiliac joints on magnetic resonance imaging. *Arthritis Rheumatol* 2023;75:2169-77.
 102. Bordner A, Aouad T, Medina CL, Yang S, Molto A, Talbot H, et al. A deep learning model for the diagnosis of sacroiliitis according to Assessment of SpondyloArthritis International Society classification criteria with magnetic resonance imaging. *Diagn Interv Imaging* 2023;104:373-83.
 103. Han Q, Lu Y, Han J, Luo A, Huang L, Ding J, et al. Automatic quantification and grading of hip bone marrow oedema in ankylosing spondylitis based on deep learning. *Mod Rheumatol* 2022;32:968-73.
 104. Bressemer KK, Adams LC, Proft F, Hermann KGA, Diekhoff T, Spiller L, et al. Deep learning detects changes indicative of axial spondyloarthritis at MRI of sacroiliac joints. *Radiology* 2022;305:655-65.
 105. Lin Y, Cao P, Chan SCW, Lee KH, Lau VWH, Chung HY. Deep learning algorithm of the SPARCC scoring system in SI joint MRI. *J Magn Reson Imaging* 2024;60:1390-9.
 106. Triantafyllou M, Klontzas ME, Koltsakis E, Papakosta V, Spanakis K, Karantanas AH. Radiomics for the detection of active sacroiliitis using MR imaging. *Diagnostics (Basel)* 2023;13:2587.
 107. Zhang K, Liu C, Pan J, Zhu Y, Li X, Zheng J, et al. Use of MRI-based deep learning radiomics to diagnose sacroiliitis related to axial spondyloarthritis. *Eur J Radiol* 2024;172:111347.
 108. Li X, Lin Y, Xie Z, Lu Z, Song L, Ye Q, et al. Automatic segmentation of fat metaplasia on sacroiliac joint MRI using deep learning. *Insights Imaging* 2024;15:93.
 109. Lin Y, Chan SCW, Chung HY, Lee KH, Cao P. A deep neural network for MRI spinal inflammation in axial spondyloarthritis. *Eur Spine J* 2024;33:4125-34.
 110. Bressemer KK, Vahldiek JL, Adams L, Niehues SM, Haibel H, Rodriguez VR, et al. Deep learning for detection of radiographic sacroiliitis: achieving expert-level performance. *Arthritis Res Ther* 2021;23:106.
 111. Li H, Tao X, Liang T, Jiang J, Zhu J, Wu S, et al. Comprehensive AI-assisted tool for ankylosing spondylitis based on multicenter research outperforms human experts. *Front Public Health* 2023;11:1063633.
 112. Üreten K, Maraş Y, Duran S, Gök K. Deep learning methods in the diagnosis of sacroiliitis from plain pelvic radiographs. *Mod Rheumatol* 2023;33:202-6.
 113. Lee KH, Lee RW, Lee KH, Park W, Kwon SR, Lim MJ. The development and validation of an AI diagnostic model for sacroiliitis: a deep-learning approach. *Diagnostics (Basel)* 2023;13:3643.
 114. Van Den Berghe T, Babin D, Chen M, Callens M, Brack D, Maes H, et al. Neural network algorithm for detection of erosions and ankylosis on CT of the sacroiliac joints: multicentre development and validation of diagnostic accuracy. *Eur Radiol* 2023;33:8310-23.
 115. Liu L, Zhang H, Zhang W, Mei W, Huang R. Sacroiliitis diagnosis based on interpretable features and multi-task learning. *Phys Med Biol* 2024;69:045034.
 116. Zhang K, Luo G, Li W, Zhu Y, Pan J, Li X, et al. Automatic image segmentation and grading diagnosis of sacroiliitis associated with AS using a deep convolutional neural network on CT images. *J Digit Imaging* 2023;36:2025-34.
 117. Baek IW, Jung SM, Park YJ, Park KS, Kim KJ. Quantitative prediction of radiographic progression in patients with axial spondyloarthritis using neural network model in a real-world setting. *Arthritis Res Ther* 2023;25:65.
 118. Joo YB, Baek IW, Park YJ, Park KS, Kim KJ. Machine learning-based prediction of radiographic progression in patients with axial spondyloarthritis. *Clin Rheumatol* 2020;39:983-91.
 119. Garofoli R, Resche-Rigon M, Roux C, van der Heijde D, Dougados M, Moltó A. Machine-learning derived algorithms for prediction of radiographic progression in early axial spondyloarthritis. *Clin Exp Rheumatol* 2023;41:727-34.
 120. Koo BS, Jang M, Oh JS, Shin K, Lee S, Joo KB, et al. Machine learning models with time-series clinical features to predict radiographic progression in patients with ankylosing spondylitis. *J Rheum Dis* 2024;31:97-107.
 121. Fabry V, Mamalet F, Laforet A, Capelle M, Acket B, Sengenès C, et al. A deep learning tool without muscle-by-muscle grading to differentiate myositis from facio-scapulo-humeral dystrophy using MRI. *Diagn Interv Imaging* 2022;103:353-9.
 122. Wang F, Zhou S, Hou B, Santini F, Yuan L, Guo Y, et al. Assessment of idiopathic inflammatory myopathy using a deep learning method for muscle T2 mapping segmentation. *Eur Radiol* 2023;33:2350-7.
 123. Garaiman A, Nooralahzadeh F, Mihai C, Gonzalez NP, Gkikopoulos N, Becker MO, et al. Vision transformer assisting rheumatologists in screening for capillaroscopy changes in systemic sclerosis: an artificial intelligence model. *Rheumatology (Oxford)* 2023;62:2492-500.
 124. Bharathi PG, Berks M, Dinsdale G, Murray A, Manning J, Wilkinson S, et al. A deep learning system for quantitative assessment of microvascular abnormalities in nailfold capillary images. *Rheumatology (Oxford)* 2023;62:2325-9.
 125. Le Gall A, Hoang-Thi TN, Porcher R, Dunogué B, Berezne A, Guillemin L, et al. Prognostic value of automated assessment of interstitial lung disease on CT in systemic sclerosis. *Rheumatology (Oxford)* 2024;63:103-10.
 126. Klontzas ME, Vassalou EE, Spanakis K, Meurer F, Woertler K, Zibis A, et al. Deep learning enables the differentiation between early and late stages of hip avascular necrosis. *Eur Radiol* 2024;34:1179-86.
 127. Faghani S, Nicholas RG, Patel S, Baffour FI, Moassefi M, Rouzrokh P, et al. Development of a deep learning model for the automated detection of green pixels indicative of gout on dual energy CT scan. *Res Diagn Interv Imaging* 2024;9:100044.
 128. Smerilli G, Cipolletta E, Sartini G, Moscioni E, Di Cosmo M, Fiorentino MC, et al. Development of a convolutional neural network for the identification and the measurement of the median nerve on ultrasound images acquired at carpal tunnel level. *Arthritis Res Ther* 2022;24:38.
 129. Minopoulou I, Kleyer A, Yalcin-Mutlu M, Fagni F, Kemenes S, Schmidkonz C, et al. Imaging in inflammatory arthritis: progress towards precision medicine. *Nat Rev Rheumatol* 2023;19:650-65.
 130. Waldstein SM, Seeböck P, Donner R, Sadeghipour A, Bogunović H, Osborne A, et al. Unbiased identification of novel subclinical imaging biomarkers using unsupervised deep learning. *Sci Rep* 2020;10:12954.
 131. GBD 2015 Healthcare Access and Quality Collaborators. Healthcare

- Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990-2015: a novel analysis from the Global Burden of Disease Study 2015. *Lancet* 2017;390:231-66.
132. Boniol M, Kunjumen T, Nair TS, Siyam A, Campbell J, Diallo K. The global health workforce stock and distribution in 2020 and 2030: a threat to equity and 'universal' health coverage? *BMJ Glob Health* 2022;7:e009316.
133. Markowetz F. All models are wrong and yours are useless: making clinical prediction models impactful for patients. *NPJ Precis Oncol* 2024;8:54.
134. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. *IEEE Trans Neural Netw Learn Syst* 2021;32:4793-813.
135. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys* 2019;29:102-27.
136. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60-88.
137. Kiryu S, Akai H, Yasaka K, Tajima T, Kunitatsu A, Yoshioka N, et al. Clinical impact of deep learning reconstruction in MRI. *Radiographics* 2023;43:e220133.
138. Abdalla M, Fine B. Hurdles to artificial intelligence deployment: noise in schemas and "gold" labels. *Radiol Artif Intell* 2023;5:e220056.
139. Goceri E. Medical image data augmentation: techniques, comparisons and interpretations. *Artif Intell Rev* 2023;56:12561-605.
140. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. *BMC Med Imaging* 2022;22:69.
141. Qamar T, Bawany NZ. Understanding the black-box: towards interpretable and reliable deep learning models. *PeerJ Comput Sci* 2023;9:e1629.
142. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. *arXiv [Online]*. 2014 [cited 2024 Oct 24]. Available from: <https://doi.org/10.48550/arXiv.1312.6034>.
143. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* 2015;10:e0130140.
144. Ho SY, Phua K, Wong L, Bin Goh WW. Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns (N Y)* 2020;1:100129.
145. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195.
146. Eche T, Schwartz LH, Mokrane FZ, Derclé L. Toward generalizability in the deployment of artificial intelligence in radiology: role of computation stress testing to overcome underspecification. *Radiol Artif Intell* 2021;3:e210097.
147. Li Z, Kamnitsas K, Glocker B. Analyzing overfitting under class imbalance in neural networks for image segmentation. *IEEE Trans Med Imaging* 2021;40:1065-77.
148. Choi YS. Concepts, characteristics, and clinical validation of IBM Watson for oncology. *Hanyang Med Rev* 2017;37:49-60.
149. England BR, Tiong BK, Bergman MJ, Curtis JR, Kazi S, Mikuls TR, et al. 2019 update of the American College of Rheumatology recommended rheumatoid arthritis disease activity measures. *Arthritis Care Res (Hoboken)* 2019;71:1540-55.
150. Yu KH, Healey E, Leong TY, Kohane IS, Manrai AK. Medical artificial intelligence and human values. *N Engl J Med* 2024;390:1895-904.
151. Naik N, Hameed BMZ, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? *Front Surg* 2022;9:862322.
152. Shaw J, Ali J, Atuire CA, Cheah PY, Español AG, Gichoya JW, et al. Research ethics and artificial intelligence for global health: perspectives from the global forum on bioethics in research. *BMC Med Ethics* 2024;25:46.
153. Caruso PF, Greco M, Ehm C, Angelotti G, Cecconi M. Implementing artificial intelligence: assessing the cost and benefits of algorithmic decision-making in critical care. *Crit Care Clin* 2023;39:783-93.
154. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics* 2021;22:122.