ARTIFICIAL INTELLIGENCE AND INFORMATICS

REVIEW

Artificial intelligence in musculoskeletal applications: a primer for radiologists

Michelle W. Tong[1,2,3]

Jiamin Zhou[4]

Zehra Akkaya[1,5]

Sharmila Majumdar[1,2]

Rupsa Bhattacharjee[1]

[1]University of California San Francisco, Department of Radiology and Biomedical Imaging, San Francisco, USA

[2]University of California San Francisco, Department of Bioengineering, San Francisco, USA

[3]University of California Berkeley, Department of Bioengineering, Berkeley, USA

[4]University of California San Francisco, Department of Orthopaedic Surgery, San Francisco, USA

[5]Ankara University Faculty of Medicine, Department of Radiology, Ankara, Türkiye

ABSTRACT

As an umbrella term, artificial intelligence (AI) covers machine learning and deep learning. This review aimed to elaborate on these terms to act as a primer for radiologists to learn more about the algorithms commonly used in musculoskeletal radiology. It also aimed to familiarize them with the common practices and issues in the use of AI in this domain.

KEYWORDS

Artificial intelligence, deep learning, machine learning, musculoskeletal, review

Approximately 1.71 billion people have musculoskeletal (MSK) conditions worldwide.[1] The need for imaging on MSK disorders is increasing in parallel with the rising and progressively aging global population,[2] posing a significant threat of fatigue in radiologists and unmet needs for patients.[3,4] The evolution of MSK radiology traces back to the inception of the field of radiology itself with the discovery of X-rays in 1895. On a separate trajectory, the 1950s witnessed the introduction of the first programming languages and software, raised by Turing's[5] question, "can machines think?". However, it was not until 1992, nearly a century later, that these two fields merged, culminating in the first research into artificial intelligence (AI) in radiology.[6] Today, AI has become an ever-growing field and is reshaping the world, including medicine, with radiology at the forefront, evidenced by Food and Drug Administration (FDA)-approved AI-based tools. The first AI-based algorithm was approved by the FDA in 2017. By 2022, radiology dominated the medical field by a striking 87% of all FDA-authorized AI-based devices.[7] In 2017, MSK applications were the second most common subject of AI-related publications in radiology, second only to neuroradiology.[8]

Thus far, AI research in radiology has primarily focused on interpretive tasks, including fracture detection, osteoarthritis detection and grading (cartilage and meniscal lesions), bone age determination, osteoporosis and bone quality assessment, tissue/region identification and segmentation, radiographic angle and bone measurements, clinical decision making on various bone and ligament anomalies, lesions characterization and diagnosis of infectious, oncological or rheumatological diseases, quantitative analysis and radiomics, and estimation of patient demographics.[9] However, AI also offers promising solutions for non-interpretive tasks, which aim to ensure high-quality care and time-efficient outputs for the rising demands on imaging.[10,11] Indeed, non-interpretive tasks, such as protocoling, quality control, and overseeing imaging studies, comprise 44% of a radiologist's daily workload.[12] However, most of these tasks are neglected where productivity is mainly assessed by the number of produced reports. Research in the emergency radiology department shows that for every 1 minute spent on the phone by radiologists, the report turnaround time increases by approximately 4 minutes.[12] Therefore, it is imperative to create time-efficient solutions to meet the rising demand in the field, where AI offers revolutionary solutions.

Therefore, radiologists must embrace a comprehensive understanding of AI and machine learning (ML) to integrate these technologies into their practice effectively, as described in Figure 1. Proficiency in data interpretation and validation will ensure the accuracy and reli-

ability of AI algorithms beneficial for clinical practice. Understanding the underlying principles of ML models, such as neural networks and deep learning (DL) architectures, is essential for critical appraisal and informed decision-making. Radiologists must also grasp the limitations and potential biases inherent in AI systems, emphasizing the importance of human oversight in clinical decision-making. Furthermore, knowledge of regulatory frameworks and ethical considerations surrounding AI adoption in healthcare is imperative to navigate legal and ethical challenges.

Algorithms

Alongside advancements in computational power, computer algorithms, and data availability, AI has gained popularity as a rapidly developing tool that can transform industries. Broadly defined, AI refers to computer systems that can perform assigned tasks, such as learning, decision-making, and problem-solving, with satisfactory or better-than-expected performance within a given context. Subsets of AI include the following: artificial narrow intelligence, which can perform specific tasks well but cannot transfer knowledge; artificial general intelligence, which can transfer knowledge across systems or tasks; and artificial superintelligence, which functions beyond the capability of human beings and is currently mainly

conceptualized.[13] Commonly used AI concepts and descriptions are listed in Table 1.

ML essentially entails all techniques that can be employed to train a machine to mimic human performance. In the current context, it refers[14] to the development of algorithms that predict discrete labels (classification), continuous quantities (regression), data subgroups (clustering), or important features (dimensionality reduction) based on previous experiences using probability, statistics, and linear algebra. Traditional ML algorithms include linear classifiers, logistic regression, decision trees, and nearest-neighbor searches. Each of these algorithms seeks to learn a mapping between input and output variables by defining decision boundaries between labeled data or clustering of the data.

DL refers to a subset of ML that utilizes neural networks to learn new high-level feature representations of data for computer vision tasks, such as object segmentation, classification, and detection, with high efficiency.[15] Neural networks are composed of multiple layers of interconnected nodes with internal weights modeled after biological neural systems. The network learns to perform tasks by iteratively performing complex, non-linear transforms, involving

passing forward input data through the network to predict a desired output and then using the discrepancy between the predicted and expected output to update the internal weights of the nodes in the network to improve task performance.

Convolutional neural networks (CNNs) perform convolution operations over local regions using shared convolution weights such that networks achieve translational invariance (i.e., objects can be detected regardless of location). Additional pooling operations down-sample data representations, automatically extracting relevant spatial hierarchical features. Variational CNNs have modified the underlying network structure to improve versatility and effectiveness. The two-dimensional (2D) U-Net was a significant breakthrough for medical imaging tasks, particularly segmentation. In 2015, Ronneberger et al.[16] proposed a unique U-shaped architecture (Figure 2), which down-sampled and up-sampled input images of varying image modalities to predict regions of interest with "very good performance," even after training with a very limited amount of training data. Despite their successes, CNNs are prone to overfitting, meaning CNN-based models do not perform as well on new unseen data. They also suffer from a requirement for large amounts of data for training and a lack of in-

Figure 1. A schematic diagram of the usage of artificial intelligence (AI) in multiple levels of musculoskeletal radiology before, during, and after examination. *It is important to emphasize that continuous input from radiologists is crucial to minimize risks from AI in real-world clinical scenarios and to provide uncompromised patient safety at every step in the flowchart where AI-based solutions are being tested. The figure has been created with the help of the Biorender tool (https://www.biorender.com).

Figure 2: An introduction to the seminal U-Net architecture. Reproduced via Creative Commons License from.[16]

Table 1. A list of commonly used artificial intelligence concepts and descriptions

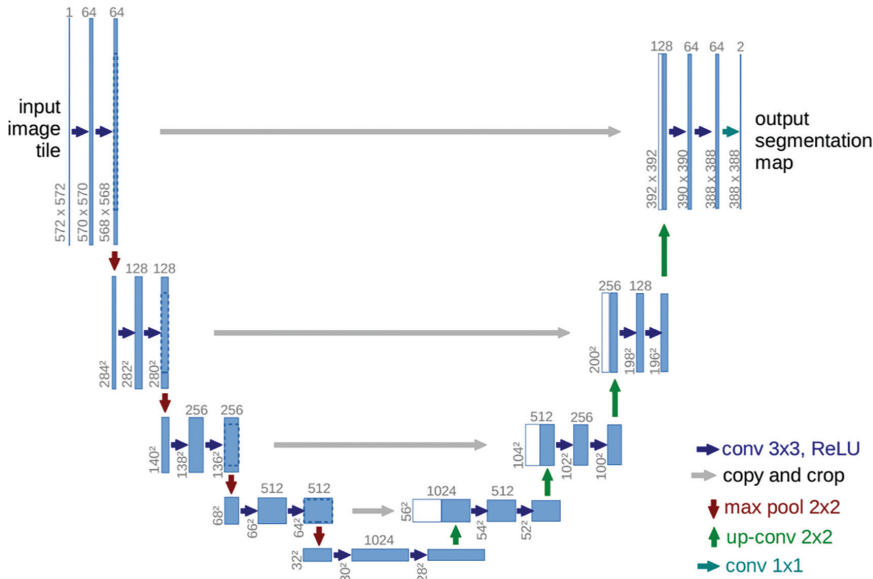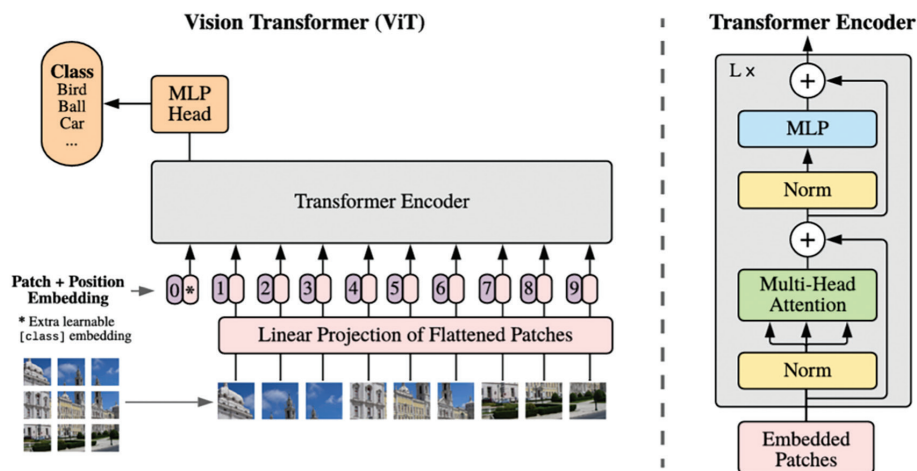| Concepts	Meanings in one line
Artificial intelligence (AI)	The simulation of human intelligence processes by machines, particularly computer systems.
Machine learning (ML)	A subset of AI that allows systems to learn from data and improve over time without being explicitly programmed.
Deep learning	A subset of ML where artificial neural networks mimic the structure and function of the human brain to process data.
Neural networks	A system of algorithms modeled after the human brain, used to recognize patterns.
Natural language processing	The ability of computers to understand, interpret, and generate human language.
Computer vision	The field of AI that enables computers to interpret and understand visual information from the real world.
Reinforcement learning	A type of ML where an agent learns to make decisions by trial and error, receiving feedback in the form of rewards or penalties.
Supervised learning	A type of ML where the model is trained on labeled data, with input–output pairs provided.
Unsupervised learning	A type of ML where the model is trained on unlabeled data and must find patterns and relationships on its own.
Semi-supervised learning	A hybrid approach where the model is trained on a small amount of labeled data and a large amount of unlabeled data.
Transfer learning	An ML technique where a model trained on one task is repurposed or fine-tuned for a similar task.
Generative adversarial networks	A class of algorithms used in unsupervised learning to generate new data instances similar to a given dataset.
Overfitting	When a model learns to perform well on the training data but fails to generalize to new, unseen data.
Bias and variance	Bias refers to the error introduced by approximating a real-world problem with a simplified model, while variance refers to the error introduced by sensitivity to fluctuations in the training set.
Feature engineering	The process of selecting and transforming variables or features to improve the performance of ML algorithms.
Hyperparameters	Parameters that are set prior to training and control the learning process of ML algorithms.
Ensemble learning	A technique that combines multiple models to improve the performance of the overall system.

terpretability due to the model's architectural complexity.

Federated learning proposes a framework to address challenges with model generalizability, with special benefits when using medical data. An aggerate model encapsulates shared model weights from multiple collaborators who trained the model on private datasets.[17]

Generative adversarial networks (GANs) are popular for image-to-image translation, consisting of two opposing networks: a generator and a discriminator.[18] The generator creates an image to fool the discriminator, while the discriminator attempts to discern real or synthetic images.[19] Due to the oppositional nature of the network, GANs can be challenging to train and often require careful consideration of hyperparameters. Mode collapse occurs when the generator produces similar images that may not capture the full distribution of the training data and the discriminator is unable to provide useful feedback to guide training.

Recently, large language models and vision transformers (ViTs)[20] have spurred a new wave of innovation. Both of these DL architectures are based on transformers, which consist of an encoder, which extracts meaning features from input data, and a decoder network, which uses the features to generate outputs. Transformers process data as a sequence of tokens, enabling the model to capture global relationships between the data (Figure 3). For ViTs, images are vectorized into tokens, which can be combined with text.[21]

A typical workflow to develop an ML algorithm involves several distinct stages. It begins with problem definition and data collection where a specific object is identified, and relevant data is gathered. Subsequently, data preprocessing involves cleaning, transforming, and processing the dataset for training. Common preprocessing techniques include image normalization and clipping to achieve favorable image intensity ranges and contrast for ML models. Before model development, data is split into training, validation, and testing subsets, often with balanced distributions of relevant metadata, such as age, for proper evaluation of model performance. During training, models may be prone to overfitting if highly sensitive to patterns in the training dataset. The validation dataset allows for the evaluation of model performance during training, while the test set is only used to assess performance on the final selected model for unbiased assessment.

Figure 3. An introduction to vision transformers. Reproduced via Creative Commons License from.[20]

Next, model selection and training occur, where various algorithms are evaluated, and a suitable model is chosen. Existing models may offer excellent zero-shot capabilities such that no modification of model weights is needed. On the other hand, models may be trained for a specific use case by fine-tuning, which involves further training of a pre-trained model on a smaller, targeted data set. After training, the model is evaluated on the test dataset using appropriate metrics specific to the objects of the model. Finally, the model is deployed and undergoes monitoring and maintenance to ensure optimal performance over time. This iterative process requires collaboration between domain experts, data scientists, and computer programmers to achieve successful outcomes. Some of the crucial technical terms and metrics used in everyday ML, and what they mean, are listed in Table 2. Although AI seems to be an omnipresent tool in current radiology practices, many users remain unfamiliar with the basic concepts, utilities, challenges, processes, and biases associated with it. We aim to provide comprehensive starting content that prepares the community of medical experts to become tuned to the vocabulary and its nuances and to get a sense of how AI can be integrated into their daily MSK radiology practice.

Applications in musculoskeletal radiology

Image acquisition

Imaging acceleration

Extensive research dedicated to reducing the time required to acquire medical images has led to the development of unique data sampling and reconstruction techniques in MSK radiology, primarily for computed tomography (CT) and magnetic resonance imaging (MRI). In particular, MRI is an important modality for radiologists to diagnose many MSK conditions, but it suffers from increased cost and increased time to acquire images compared with other modalities. AI-based image acceleration techniques aim to break those Nyquist limits, though this must be done while considering any in-domain and domain-shift artifacts. Reconstruction, therefore, is equally essential to ensure the quality of images is clinically preserved in rapidly acquired MRI. AI researchers have developed algorithms that achieve both high accelerations for faster imaging and excellent reconstitution with comparable or improved image resolution. Such methodologies have been developed using data-driven guidance, such as compressed sensing or dictionary learning, or physics-guided networks combined with artifact removal.[22] These techniques are often modified for solution-specific problems, including accelerating higher-dimensional 2D or 3D MRI scans, such as dynamic (temporal) MRI.[23] AI techniques for the joint optimization of a non-Cartesian *k*-space sampling trajectory and an image-reconstruction network have been rising in popularity. For example, one such framework, PROJECTOR,[24] proposed dubbed projection for jointly learning non-Cartesian trajectories while optimizing reconstructor trajectories. It also ensured that the learned trajectories were compatible with gradient-related hardware constraints. Previous techniques enforced these constraints via penalty terms, but PROJECTOR enforces them via embedded steps that project the learned trajectory on a feasible set.

Synthesis of images and parametric maps

Another exciting application of AI is to characterize meaningful tissue maps or images from raw data (Figures 4 and 5). Wu et al.[25] proposed CNNs for synthesizing water/fat images from only two echoes instead of multiple. The method achieved high-fidelity output images, a 10-fold acceleration in computation time, and also generalizability to unseen organ images and metal artifacts. Zou et al.[26] have also proposed reconstructing free-breathing cardiac MRI data and synthesizing cardiac cine movies from manifold learning networks. This enables a unique generation of synthetic breath-hold cine movies with data on demand: specifically, movies with different inversion contrasts. Additionally, it enables the estimation of T_1 maps with specific respiratory phases. So far, the derivation of tissue parameter maps has been achieved by repeating acquisition in steady-state conditions and longer scan times.[22] However, rapid extraction of such parameters is no longer a challenge due to AI-based solutions, such as synthetic mapping of T_1, T_{1p}, R_2^*, and T_2 relaxation, chemical exchange saturation transfer proton volume fraction and exchange rate, magnetization transfer, and susceptibility. Conventional magnetic resonance fingerprinting (MRF) is regularly used for quantitative parameter estimation. However, it suffers from the computational burden of dictionary generation and pattern matching. The burden further grows exponentially with the number of fitting parameters considered. ML has also been utilized to accelerate both the acquisition and reconstruction and thus optimize MRF sequences.[22]

End-to-end design

End-to-end design of reconstruction and segmentation techniques have recently been a heavy focus in the medical imaging community. Often addressed separately, these two tasks could benefit from being handled in tandem. Tolpadi et al.[27] recently hosted and summarized a challenge entitled "K2S," hosted at the 25th International Conference on Medical Image Computing and Computer-Assisted Intervention (Singapore, 2022). Eight-times under-sampled raw MRI measurements were provided as training data with their fully sampled counterparts and segmentation masks (i.e., a unique dataset consisting of 300 knee MRI scans accompanied by radiologist-approved tissue segmentation labels). In the testing phase, the challenge participants submitted DL models that generated high-efficiency segmenta-

Table 2. Technical terms and metrics used in everyday machine learning: what do they mean?

Technical terms

Terms	Explanations
Feature	An individual measurable property or characteristic of a phenomenon being observed, often represented as a variable in a dataset.
Label	The output or target variable in supervised learning, representing the prediction or classification to be made.
Instance	A single example or data point in a dataset, typically represented as a row in a table.
Model	A mathematical representation or algorithm that learns patterns and relationships from data to make predictions or decisions.
Training data	The data used to train a machine learning (ML) model, consisting of input features and corresponding labels.
Test data	Data used to evaluate the performance of a trained ML model, separate from the training data.
Validation data	Data used to fine-tune hyperparameters and assess model performance during the training process.
Loss function	A function that measures the difference between predicted and actual values, used to train and optimize ML models.
Optimization algorithm	An algorithm used to adjust the parameters of a model during training to minimize the loss function.
Gradient descent	An optimization algorithm that iteratively updates the parameters of a model by moving in the direction of steepest descent of the loss function.
Epoch	One complete pass through the entire training dataset during the training of an ML model.
Batch	A subset of the training data used in one iteration of training, typically chosen to improve efficiency.
Batch size	The number of training examples utilized in one iteration of training during the gradient descent process.
Learning rate	A hyperparameter that controls the step size during the optimization process, determining the rate at which the model parameters are updated.
Stop criteria	Criteria by which model stop training, such as for "x" number of epochs or until the loss stops decreasing by "x"%. Clear stop criteria and assessment of training loss allow a fairer comparison of model weights.
Regularization	Techniques used to prevent overfitting by adding a penalty term to the loss function, discouraging complex models.
Dropout	A regularization technique used in neural networks to randomly deactivate neurons during training to prevent overfitting.
Activation function	A mathematical function applied to the output of each neuron in a neural network, determining its output.
Backpropagation	An algorithm used to train neural networks by iteratively adjusting the weights of connections based on the error calculated during forward pass.
Convolutional neural network	A type of neural network designed for processing structured grids of data, commonly used in image recognition tasks.
Recurrent neural network (RNN)	A type of neural network designed to process sequences of data, with connections between units forming directed cycles, commonly used in natural language processing tasks.
Long short-term memory	A type of RNN unit capable of learning long-term dependencies, commonly used in sequence prediction tasks.

Common metrics

Accuracy	The proportion of correctly classified instances (both true positives and true negatives) out of the total instances.
Precision	The proportion of true positive predictions out of all positive predictions made by the model.
Recall (sensitivity)	The proportion of true positive predictions out of all actual positive instances in the dataset.
F1 Score	The harmonic mean of precision and recall, providing a balance between the two metrics.
Specificity	The proportion of true negative predictions out of all actual negative instances in the dataset.
ROC area under the curve score	The area under the receiver operating characteristic (ROC) curve, representing the model's ability to discriminate between positive and negative classes across different thresholds.
Confusion matrix	A table used to evaluate the performance of a classification model, showing the counts of true positive, true negative, false positive, and false negative predictions.
Mean squared error (MSE)	The average of the squared differences between predicted and actual values, commonly used for regression tasks.
Root mean squared error	The square root of the MSE, providing a measure of the average magnitude of error in the predicted values.
Mean absolute error	The average of the absolute differences between predicted and actual values, providing a measure of average error magnitude.
Peak signal-to-noise ratio	A measure of image quality and fidelity, calculated as the ratio between the maximum power of a signal verses noise. Commonly used for reconstruction tasks.
Structural similarity index metric	A measure of preceptive similarity between two images whose formula is based on comparison of image structure, contrast, and brightness. Commonly used for reconstruction tasks.
R-squared	A statistical measure of the proportion of variance in the dependent variable that is explained by the independent variables in a regression model.
Mean average precision	A metric used to evaluate object detection models, representing the average precision over all classes at various intersection over union thresholds.
Cohen's kappa	A statistic that measures inter-rater agreement for categorical items, considering how much agreement would be expected by chance.
Mean intersection over union	A metric commonly used to evaluate semantic segmentation models, measuring the ratio of intersection to union of predicted and ground truth masks. Values range from 0 to 1, indicating no to perfect overlap, respectively.
Dice coefficient	A metric for segmentation assessment calculated by the ratio of 2 × the intersection divided by the total area of predicted and ground truth masks. This metric has good utility for small regions of interest because there is no bias from background labels. Background is often more prevalent so inclusion of these labels leading to unfavorable class imbalance.
Log loss (binary cross-entropy)	A loss function used in binary classification tasks, measuring the difference between predicted probabilities and actual binary outcomes.
Silhouette score	A measure of how similar an object is to its own cluster compared with other clusters, used to assess the quality of clustering algorithms.
Explained variance score	A metric used to evaluate the performance of regression models, measuring the proportion of variance in the target variable explained by the model.

tion maps directly from the under-sampled raw data. No correlations were found between the reconstruction and segmentation metrics (Figure 6). Some researchers suggest pre-training segmentation models on "pretext tasks". In these tasks, the model is trained to restore distorted images. Context prediction and context restoration challenges demonstrate that segmentation models can be made robust with pre-training, particularly if labeled data availability is limited.[22]
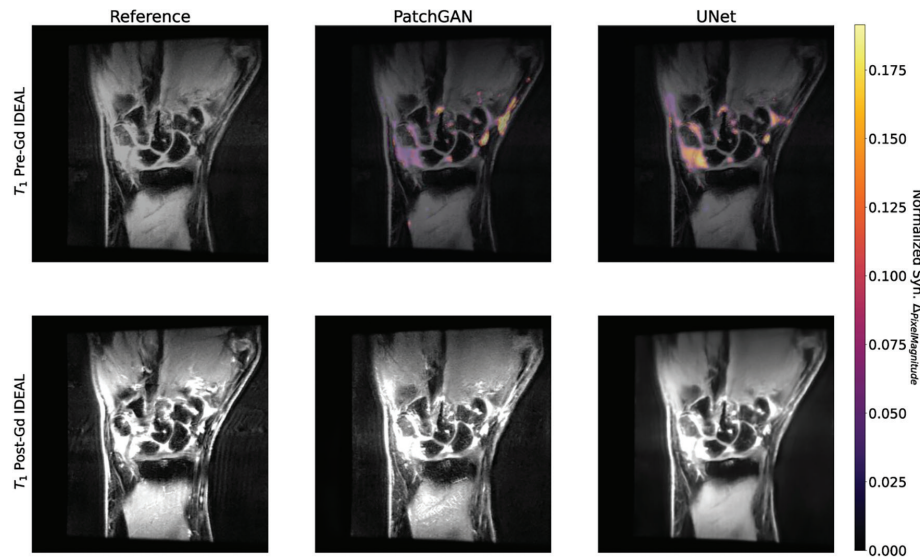

Figure 4. Occlusion maps for PatchGAN and U-Net pipelines. For U-Net and PatchGAN, hotspots primarily included intercarpal joint regions. Particularly for the U-Net, the maps also emphasized the forearm muscles. Given that the synovial joints are where an inflammatory imaging algorithm would see the most utility, the fact that both algorithms placed heavy emphasis on the intercarpal regions was promising, indicating that both focused on synovitis-relevant regions to make predictions. Reproduced via Creative Commons License from.[72]
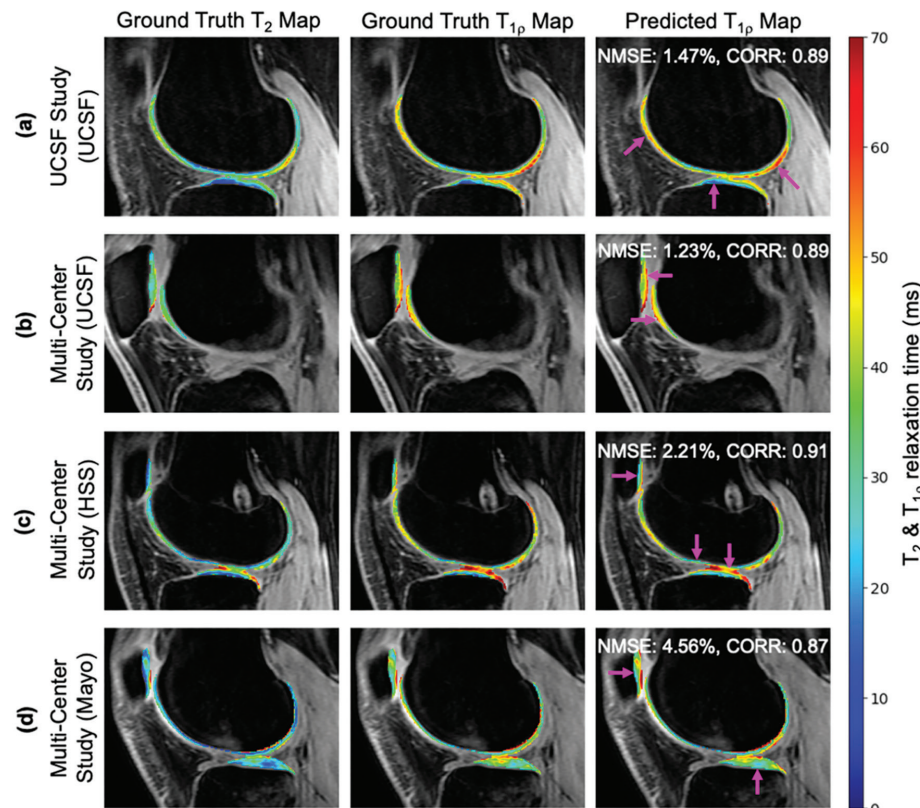

Figure 5. Four knees from patients who participated in one of two studies: **(a)** the UCSF (cohort A) study or **(b-d)** the multi-center (cohort B) study at one of three centers. Input ground truth T2 maps exhibit distinct intensity elevation and textural patterns compared with ground truth $T_{1\rho}$ maps. Nevertheless, predicted $T_{1\rho}$ maps generated by the convolutional neural network preserve these differences, as indicated by the regions marked by the arrows. Reproduced via Creative Commons License from.[74]

Image post-processing

Registration

Image registration is a critical process in imaging that focuses on the accurate alignment of images, which is necessary for the diagnosis, treatment planning, and monitoring of diseases. However, it is difficult to develop robust algorithms to register images of varying resolution and from different modalities efficiently and accurately. This is particularly challenging in the presence of significant anatomical variation in the case of MSK disease. Conventional registration methods often rely on solving pairwise optimization problems, which can be time-consuming and computationally expensive.[28] Recent literature has demonstrated the growing application of AI, in particular DL models, in image registration. CNNs, for instance, have been employed to predict the transformation required to align images. For example, a study by Sokooti et al.[29] proposed a CNN-based method for non-rigid registration on 3D chest CT follow-up data. Another novel approach involves using spatial transformer networks (STNs), a DL model that can learn spatial transformations to align images. In a study by Sokooti et al.[29] an STN was used for image registration, showing that the model could learn complex transformations from training data.[30] Models such as VoxelMorph, a CNN-based unsupervised framework for image registration,[31] have also shown promising results. Although VoxelMorph was trained on 3D brain MRI, the architecture of the models can be used to train on specific MSK datasets due to the unsupervised and generalizable nature of the models.

Segmentation

Image segmentation is a well-defined problem that involves the delineation of specific regions of interest. As manual image segmentation is both time-consuming and repetitive, the research community has explored AI to improve medical image segmentation workflows with great interest.[16] Over the years, various network architectures have been developed to segment MSK structures. One of the most popular CNN models is the U-Net, discussed earlier. It is often utilized to solve 2D or 3D segmentation tasks, such as identifying muscles, bones, cartilages, menisci, femoral and acetabular regions,

and shoulder structures in knee, spine, hip, thigh, and wrist anatomy.[32,33] Usually, the performance of existing segmentation algorithms can only be fairly compared on a specific case basis, such as anatomical region, medical imaging acquisition setting, or study population.[34]

DL can establish a useful representation of any object without prior super-imposition of user-designed features. This is why the performance of a vertebral body segmentation algorithm relies on the integrity of intervertebral discs and is compromised when disc pathologies are present if not trained with enough variety of data. Identification of a thoracic vertebral body is achieved using intrinsic features and its proximity to a disc. The disc serves as an extrinsic feature for the vertebral body. In other words, it becomes the landmark that the network learns in the context of spine segmentation (Figure 7a-c). This is also the reason for failures in patch-based approaches. Only limited contextual information is passed, which limits the outcome efficiency.

On the positive side, network learning from diverse data may often learn how the images, anatomies, and pathologies are integrated beyond visual perception, suggest new biomarkers as predictors of MSK diseases through image analysis, and potentially over-come the limitations of human perception.

Anomaly detection

Anomaly detection involves identifying abnormal structures or pathologies, such as fractures, tumors, or degenerative diseases, amidst a wide range of normal anatomical variations. To accurately distinguish between benign variants and clinically significant abnormalities, DL models-particularly CNNs-have been implemented due to their ability to learn hierarchical feature representations.[35,36] Autoencoders have also been used for unsupervised anomaly detection, whereby during the training process for reconstructing input data, they learn to encode "normal" data patterns and can thus highlight deviations from the norm when encountering an anomalous data point and produce a significantly different output.[37] These models can assist in identifying subtle or complex anomalies that may be missed by the human eye while providing consistent performance, thus reducing variability between different radiologists' interpretations. Workflow efficiency can be improved by prioritizing cases with potential anomalies identified by AI. However, there is a risk of generating false positives, false negatives, or model hallucinations, leading to unnecessary interventions or missed diagnoses. Radiol-ogists should seek AI tools that balance sensitivity and specificity to minimize false positive and negative rates.

Shape modeling

Shape modeling focuses on the accurate representation and analysis of the anatomical structures of the MSK system, with the challenge of capturing the complex geometry and variability of bones and soft tissues; this is essential for surgical pre-operative planning, prosthesis design, and the study of biomechanical properties. Active shape models and statistical shape modeling are common statistical methods to capture the variability of shape across a population and can be used for tasks such as segmentation.[38] However, they require a large amount of representative data for accurate modeling and can be sensitive to outliers with large shape deviations (Figure 8).

DL-based methods have been increasingly utilized for shape modeling due to their ability to learn complex, non-linear relationships. CNNs are commonly used due to their ability to process hierarchical features from image data directly. For instance, the U-Net architecture[16] and its variants have been extensively used for biomedical image segmentation tasks, providing detailed shape models of various anatomical structures. U-Net's strength lies in its symmetric expanding path, which allows precise localization, a key factor in accurate shape modeling. Another DL model, V-Net,[39] is a 3D variant of U-Net and is used for volumetric medical image segmentation, providing 3D shape models. Both U-Net and V-Net have shown competitive performance compared with traditional methods, with the added advantage of handling large datasets and capturing fine-grained details. DL models have recently been used for shape prediction and generation. For instance, GANs have been employed to generate realistic 3D shapes to synthesize anatomical structures for augmentation and analysis.[40] One hidden benefit of an AI-based shape model is the ability to predict changes in MSK structures over time, aiding in prognostic assessments.[35]

Radiomics

Radiomics, merging the word "radiology" with "-omics" to describe the high-through-put, data-driven approach to characterizing radiological images, involves computer-assisted image analysis where many quantitative "features" are extracted from images that are not readily appreciable to the human
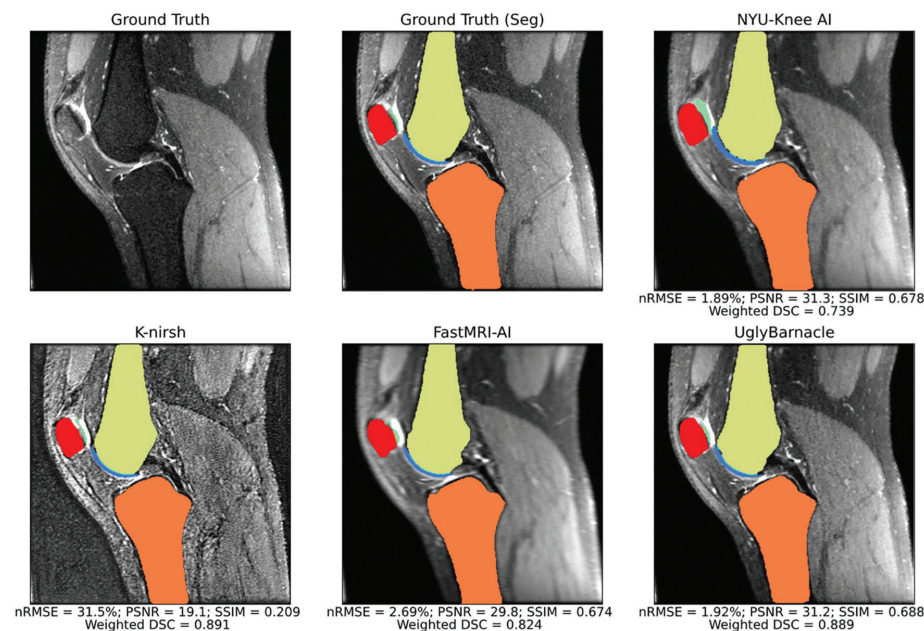

Figure 6. Miccai 2022 challenge results and submissions from the top teams. Sagittal slice segmentations are overlaid on intermediate pipeline reconstructions, displaying reconstruction and segmentation metrics for the segmented slice. Background anatomy slices were thus blurrier for some teams than for others, as different teams had different qualities of intermediate pipeline reconstruction outputs. In this example, segmentation quality was strong for all top submissions, with only some overestimation of cartilage thickness from the NYU knee artificial intelligence pipeline being apparent. K-nirsh maintains a slight edge over UglyBarnacle in reconstruction metrics for this volume. Reproduced via Creative Commons License from.[27]

eye. Radiomic features have historically involved mathematical operations on the voxels of an image, converting morphological information about anatomical structure into quantitative values. Over time, the number of features has grown exponentially as more features have been identified, making the application of ML techniques, or classifiers, to identify radiomic features increasingly popular over the past few years.[41] Support vector machines, random forests, and neural networks have been used to identify and analyze features that are most predictive of disease presence, severity, progression, and response to treatment. CNNs are also increasingly being applied to automate feature extraction. However, the clinical utility of radiomics is still being established, and integration into clinical workflows remains a challenge.

Metal artifact reduction

AI, particularly DL algorithms, is increasingly applied to mitigate metal artifacts in MSK imaging. Metal implants or instruments introduce significant artifacts, particularly in MRI, which can impair diagnostic accuracy and limit the utility of these scans. Current literature points to the use of AI in CT and radiography, but its application in MRI is less explored.[42] In the context of MRI, the integration of AI for metal artifact reduction is still in its infancy. Existing techniques without the use of AI, such as multi-acquisition variable-resonance image combination and slice encoding for metal artifact correction (SEMAC), have limitations in their application and efficacy. Studies have used neural networks to accelerate SEMAC MRI while maintaining comparable metal artifact suppression,[43] as well as using unsupervised learning or attention maps from deep neural networks to guide correction.[44] However, most of these studies rely on phantom data or MRIs of other organs of interest. There is a need for more research and development, including robust validation studies, to explore the full potential of AI in MSK MRI specifically.

Report generation

Generating accurate and informative reports is a crucial task for radiologists to convey their findings and interpretations to the referring physician in a clear, concise, and clinically relevant manner. To reduce the reporting burden on radiologists, natural language processing (NLP) techniques, such as recurrent neural networks, long short-term memory networks, and more recently, transformer-based models, such as bidirectional encoder representations from transformers and generative pre-trained transformer, can be utilized for generating radiological reports. These are trained on a large body of annotated radiological reports to learn the language and structure of report writing, as well as the relationships between imaging findings and clinical diagnoses. An additional speech recognition step can also add to the automation of the report generation process,[45] creating a text output that can be considered a "preliminary report." As radiology reports traditionally lack standardized structure and content, NLP can then be used for the extraction of meaningful or contextual information[46] from the preliminary radiology report, whether traditional text or text from speech recognition. Applications range from the extraction of specific MSK data or follow-up recommendations[47] to the generation of a final report of classification, diagnostic criteria, disease probability, or follow-up recommendations. However, AI may not capture the subtleties of human
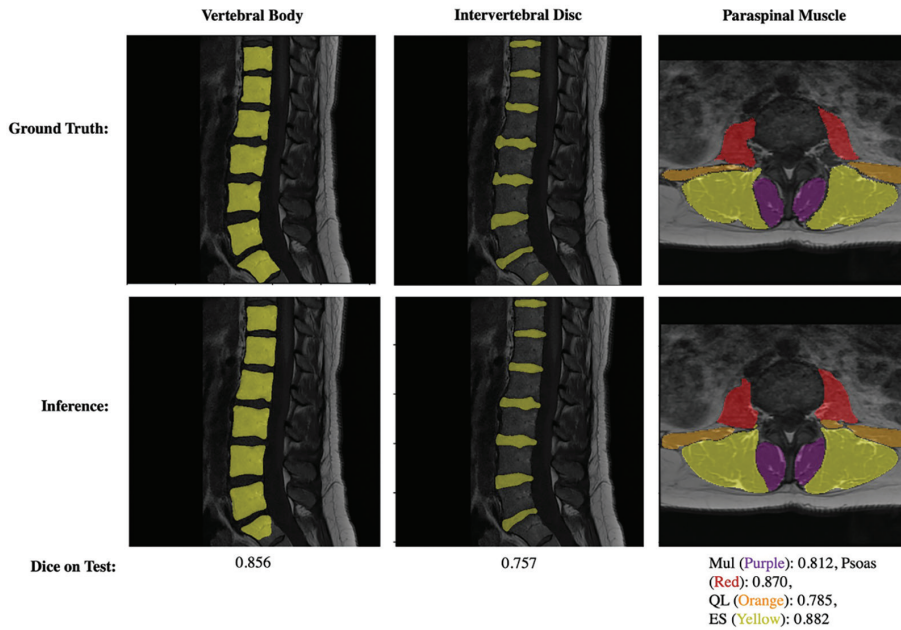

Figure 7. (a) Visualization of segmentation results from each network. The first, second, and third columns show examples of the vertebral body, intervertebral disc, and paraspinal muscle segmentation results, respectively, along with a three-dimensional Dice coefficient of each network's performance. The Dice coefficient measures the similarity between segmentation masks, where 1 indicates perfect overlap and 0 indicates no overlap. Reproduced via Creative Commons License from.[32]
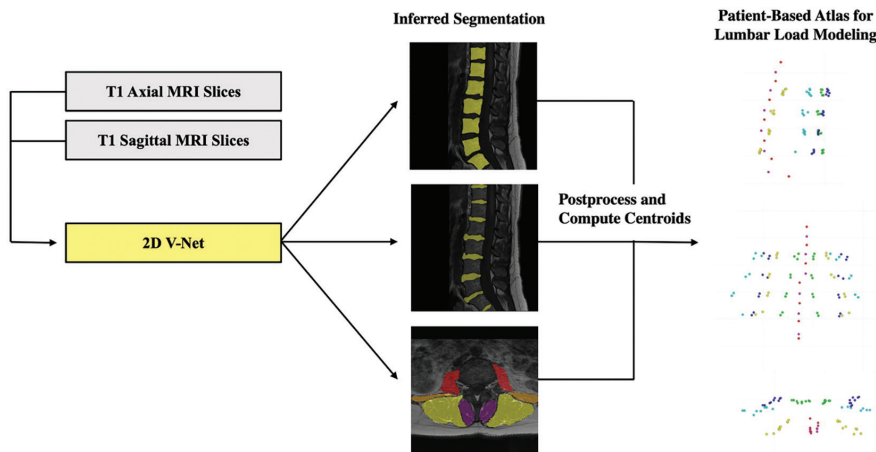

Figure 7. (b) Visualization of centroid construction. T_1 axial and T_1 sagittal MRI slices were input into their respective V-Net to generate inferred segmentation masks of the vertebral bodies, intervertebral discs, and paraspinal muscles. After postprocessing, centers of mass were computed on each segmentation mask to calculate the position of volume-wise centroids for each vertebral body and intervertebral disc and slice-wise centroids for each paraspinal muscle. These centroids were then converted to patient-based space, yielding a three-dimensional atlas of the lumbar spine for further biomechanical modeling. Reproduced via Creative Commons License from.[32] MRI, magnetic resonance imaging.
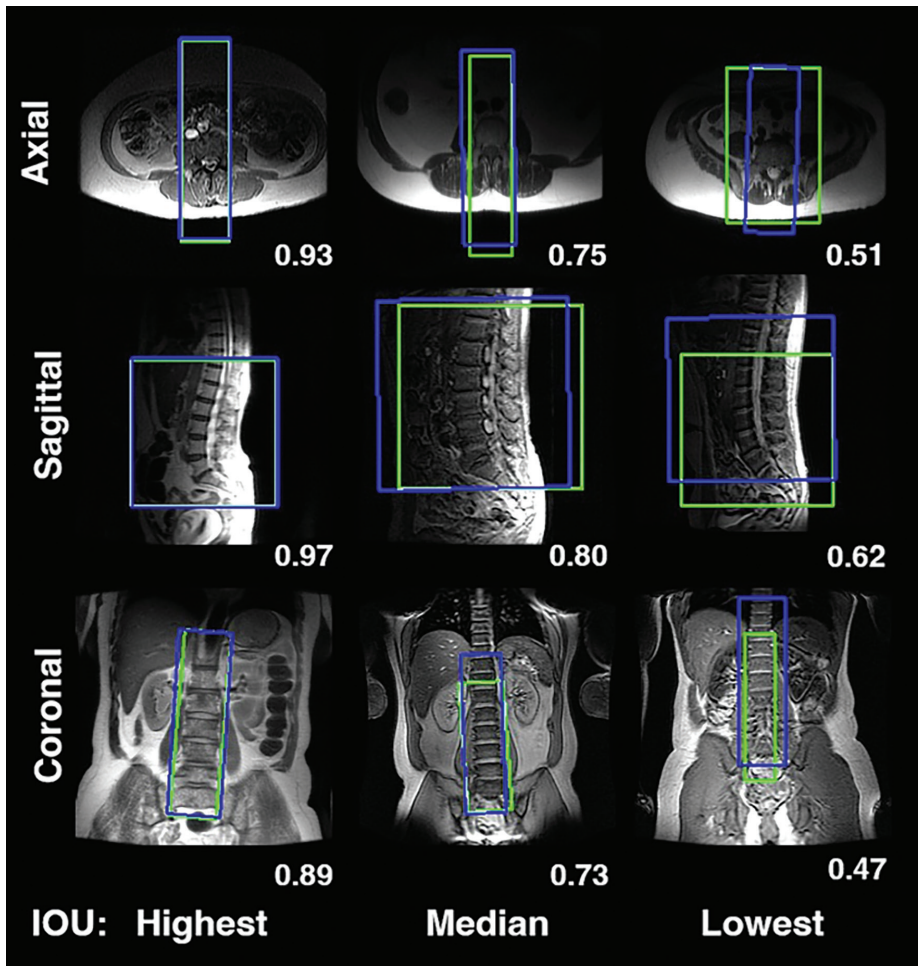
Figure 7. (c) Visualization of centroid construction. T_1 axial and T_1 sagittal magnetic resonance imaging slices were input into their respective V-Net to generate inferred segmentation masks of the vertebral bodies, intervertebral discs, and paraspinal muscles. After postprocessing, centers of mass were computed on each segmentation mask to calculate the position of volume-wise centroids for each vertebral body and intervertebral disc and slice-wise centroids for each paraspinous muscle. These centroids were then converted to patient-based space, yielding a three-dimensional atlas of the lumbar spine for further biomechanical modeling. Reproduced via Creative Commons License from.[71]
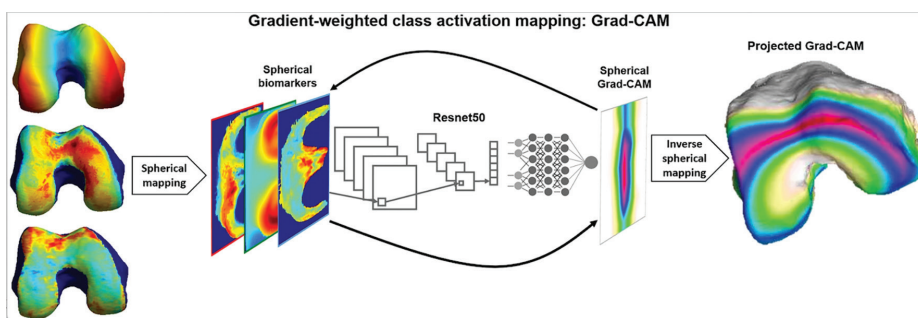

Figure 8. The authors used the Grad-CAM model interpretation technique to obtain a class discriminative localization map for each prediction. They first computed the gradient of the class of interest (before the "softmax" function) regarding feature maps of the last convolutional layer in the Resnet. These gradients flowing back were globally average-pooled to obtain the neuronal importance weights for the target class. A heat map of location importance was then up-sampled to match the image size and overlaid on the input image. The authors then leveraged the invertible property of their spherical transformation method to generate articular surface importance heat maps for model interpretation for each bone and each single biomarker. This process was performed on the first timepoint of every unique patient in the hold-out test set (n = 875) and is illustrated for the femur. Reproduced via Creative Commons License from.[73]

language, leading to reports that lack the nuanced communication often necessary between radiologists and referring physicians. Radiologists should view AI in report generation as a complementary tool that can assist with the reporting process but not as a replacement for the expert interpretation provided by a trained radiologist.

Considerations

Challenges defining ground truth data, benchmarks, and radiologists' availabilities

To achieve the highest yield from AI technologies, it is imperative to have large and reliable ground truth datasets for training, validation, and testing. Ideally, these should be from several different sources and representative of diverse communities accessible by non-radiologists, such as AI researchers, engineers, and data scientists.[48] The recent increase in the availability of such publicly available medical image banks and large-scale international AI challenges have catalyzed progress in the field, leading to the development of AI algorithms capable of handling different tasks, such as classification, detection, or segmentation, in different modalities.[49-51] The ground truth required for the current supervised AI models requires a labor- and time-intensive curation process for ideal workflow and to ensure the generalizability of a model. Moreover, this process is subject to regulatory constraints, commercial and operational pressures, as well as epistemic differences and limits of labeling.[52,53] Annotated images and their respective radiology reports are available in hospital databases but due to ethical reasons are not readily available to developers. It is important to follow the regulatory procedures and obtain approval from responsible committees to ensure an ethical approach when accessing and sharing this data between developers.[52]

Radiologists rely on visual detection, pattern recognition, memory, and cognitive reasoning to consolidate a final interpretation while making decisions.[4] Radiologists' errors have a vast impact on medical errors, which constitute the third most common cause of death in the USA, following cancer and heart disease.[54,55] The error rate is approximately 4% in clinical radiology practices, which translates into 40 million errors out of 1 billion worldwide radiographs annually.[4] Of particular importance, the distinction between an "error" and "observation variation" is highly relevant when creating such datasets. Imaging findings alone, without clinical information, are frequently

not enough to definitively indicate a specific diagnosis. Consequently, interpreting radiologic studies is typically not a straightforward binary process of discriminating normal from pathologic entities. Professional acceptability lies on an arbitrary scale, between an obvious error and the unavoidable difference of opinion in interpretation.[56] This is particularly of concern given that most clinical AI applications are developed using data generated by "expert radiologists." Thus, these models are subjected to many kinds of human errors and biases and it falls on us humans to be cognizant of inequality, data availability, and privacy, ethical and medicolegal concerns with these rapidly evolving technologies.[57,58]

The top five most influential radiology societies from the USA, Canada, Europe, Australia, and New Zealand recently released a joint statement on potential practical and ethical concerns in deploying and integrating AI in radiology practices. The key take-home statements, which also apply specifically to MSK radiology, include a strong recommendation for rigorous monitoring of its uses and safety in clinical practice, close collaboration between developers, end-users, and regulators, and strict adherence to all the regulatory steps from the development to deployment and integration in the clinical workflow.[59] Radiologists in particular should be aware of automation bias as a potential source of error when working with AI tools in decision making.[60]

Model deployment

Deploying and maintaining AI models requires a robust infrastructure that addresses computational needs for both initial deployments using off-the-shelf pre-trained models and more advanced adaptations through fine-tuning. Most radiologists and clinical departments start with off-the-shelf pre-trained AI models. These models are developed on large, general datasets and can be used directly for common imaging tasks with minimal setup and without extensive customization. Standard computing hardware, including central processing units or modest graphics processing units (GPUs), can be used to run these models, making them accessible to most clinical environments.

Fine-tuning is necessary when adapting a pre-trained model to specific datasets or unique clinical scenarios in MSK radiology. This involves modifying the pre-trained model's parameters to better fit the particular characteristics of the new data,

such as custom protocols for rare conditions, integrating specific patient demographics, or adapting models to unique imaging modalities or contrasts, improving the performance and relevance of the model. From a computational perspective, fine-tuning is less resource-intensive than training a model from scratch, as the model has already learned useful features from the initial large-scale dataset. This can be particularly beneficial in medical imaging, where annotated datasets are often limited and expensive to acquire. For instance, a model initially trained on a large dataset of general MRI images can be fine-tuned on a smaller dataset of specific MSK conditions. Studies using this approach have been reviewed by Cheplygina et al.[61], demonstrating improved performance on the tasks of interest. However, higher computational resources than those used for deployment are still needed for the fine-tuning process to handle the training workload. High-performance GPUs or tensor processing units are resources that can accelerate the processing of large datasets and complex model architectures during the training phase of fine-tuning. Cloud-based solutions with an environment that is secure and compliant with the Health Insurance Portability and Accountability Act also offer scalable resources that can be dynamically adjusted based on the computational load, making them ideal for training and deploying models without the need for local high-performance hardware.

Successful deployment of AI tools requires seamless integration into clinical workflows, which may involve Digital Imaging and Communications in Medicine (DICOM) standards and interoperability with various Picture Archiving and Communication System software, supported by robust infrastructure capable of handling ongoing model monitoring and updates to ensure sustained performance over time, adjust for any data shifts or incorporate new data, and maintain model relevance and performance.

Equitable medical artificial intelligence

The development and deployment of AI technologies in MSK radiology must be prioritized for fairness and justice. Algorithms should aim to mitigate biases, ensure accessibility to all demographic groups, and deliver personalized care tailored to individual needs, irrespective of socio-economic status or background. Doo and McGinty[62] argue that bias in radiology AI stems from various stages of model design encompassing the

selection of training data, algorithm development, deployment, and performance assessment. These biases, in turn, have repercussions on patient care and health outcomes. Notably, there is a lack of standardized protocols for demographic labeling in AI. Existing datasets often blur distinctions between crucial identifiers, such as sex and gender, or oversimplify complex racial categories, leading to distorted outcomes and predictive inaccuracies. Consequently, AI models trained on such biased datasets tend to reinforce preexisting biases, contributing to unintended consequences.

When contemplating advanced healthcare imaging within the AI landscape, a fundamental query arises: Is it possible to completely anonymize (deidentifying without any possibility of reidentification) data?[63] At first glance, the task appears simple: selectively erase or encode identifiers within the metadata headers of images. Despite the widespread use of the DICOM standard for radiologic data, an increasing number of exceptions complicate efforts to establish standardized procedures. Recently, the progress in facial recognition technology has raised concerns about the potential for matching images from CT or MRI scans with individuals' photographs. Consequently, it has become standard practice in medical imaging research to alter images using defacing or skull-stripping algorithms to eliminate facial features. Unfortunately, such alterations can undermine the generalizability of ML models developed using such data.[64] The topic is extremely complicated in terms of types of biases and there are several remedies, which are almost impossible to comprehensively cover in the scope of the article. However, it is important to introduce the concepts of bias and equitable medical AI in MSK radiology and something to be conscious of while utilizing the AI tools.[64] Some of the most common issues with MSK imaging in AI and potential solutions to those are listed in Table 3.

Conclusion: current trends and future directions

Integration of AI with other emerging technologies, such as augmented reality and virtual reality is enabling more immersive and interactive visualization of medical images. New tools may facilitate better surgical planning, training, and intraoperative guidance. Additionally, AI-assisted tools have a niche role in aiding radiologists who are training and provide an avenue for additional diagnostic opinion where multiple

Table 3. Common problems and potential solutions

Common problems	Potential solutions
Data quality issues	- Preprocessing techniques, such as denoising and image enhancement. - Augmentation methods to increase dataset diversity and robustness.
Limited annotated data	- Semi-supervised or weakly supervised learning approaches. - Active learning strategies to prioritize data labeling efforts. - Transfer learning from pre-trained models on larger datasets.
Class imbalance	- Data resampling techniques, such as oversampling or undersampling. - Class-weighted loss functions to penalize errors on minority classes. - Synthetic data generation to balance class distribution.
Interpretability and explainability	- Model visualization techniques, such as saliency maps and activation maximization. - Explainable artificial intelligence methods, such as local interpretable model-agnostic explanations or Shapley additive explanations. - Incorporating attention mechanisms to highlight important image regions.
Overfitting and generalization	- Regularization techniques, such as dropout and weight decay. - Cross-validation and validation set monitoring to detect overfitting. - Domain adaptation methods to improve model robustness across different datasets.
Computational resource constraints	- Model compression techniques, such as pruning and quantization. - Distributed training frameworks for parallel processing. - Cloud-based solutions for scalable compute resources.
Ethical and legal considerations	- Adherence to data protection regulations, such as the Health Insurance Portability and Accountability Act. - Bias detection and mitigation strategies during model development. - Transparent reporting of model performance and limitations.

radiologists reading images is not feasible. Protocolling, which involves choosing the right imaging protocol to obtain the most diagnostic images for each patient, is supervised by a radiologist and is particularly important in MSK MRI applications where imaging protocols frequently require patient-specific tailoring. The limited number of research reports, using CNN and natural language classifier-based algorithms, have demonstrated encouraging outcomes.[65-67] Nevertheless, it is important to acknowledge the diversity of MSK imaging protocols for a wide spectrum of clinical scenarios, where these tools should be fine-tuned and advanced by taking medical history, prior imaging studies, scanner-specific data, contrast information, and radiation exposure dose into account.[68] AI can also offer dual working solutions for scheduling, by reducing both MRI times and waiting times by identifying no-shows or canceled appointments ahead of time.[69] Finally, radiology reports are the final product of radiologists and are the means of communication of findings between physicians. ML can help generate decision-making algorithms as a support system based on the available information on the patient's medical background.[68,70] Conversely, ML-based NLP can be a powerful tool to harness data from radiology reports and is currently being investigated.[9]

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. World Health Organization. Musculoskeletal health. [Crossref]

2. Harkey P, Duszak R Jr, Gyftopoulos S, Rosenkrantz AB. Who refers musculoskeletal extremity imaging examinations to radiologists? *AJR Am J Roentgenol.* 2018;210:834-841. [Crossref]

3. Reiner BI, Krupinski E. The insidious problem of fatigue in medical imaging practice. *J Digit Imaging.* 2012;25:3-6. [Crossref]

4. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive error in radiology. *Am J Roentgenol.* 2017;208:739-749. [Crossref]

5. Turing AM. Computing machinery and intelligence. *Mind New Ser.* 1950;59:433-460. [Crossref]

6. Driver CN, Bowles BS, Bartholmai BJ, Greenberg-Worisek AJ. Artificial intelligence in radiology: a call for thoughtful application. *Clin Transl Sci.* 2020;13:216-218. [Crossref]

7. Health C for D and R: Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. FDA 2023. [Crossref]

8. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp.* 2018;2:35. [Crossref]

9. Debs P, Fayad LM. The promise and limitations of artificial intelligence in musculoskeletal imaging. *Front Radiol.* 2023;3:1242902. [Crossref]

10. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18:500-510. [Crossref]

11. Richardson ML, Garwood ER, Lee Y, et al. Noninterpretive uses of artificial intelligence in radiology. *Acad Radiol.* 2021;28:1225-1235. [Crossref]

12. Glover M, Almeida RR, Schaefer PW, Lev MH, Mehan WA Jr. Quantifying the impact of noninterpretive tasks on radiology report turn-around times. *J Am Coll Radiol.* 2017;14:1498-1503. [Crossref]

13. Kelly S, Kaye SA, Oviedo-Trespalacios O. What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telemat Inform.* 2023;77:101925. [Crossref]

14. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics.* 1970;12:55-67. [Crossref]

15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436-444. [Crossref]

16. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. 2015;234-241. [Crossref]

17. Sheller MJ, Edwards B, Reina GA, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep.* 2020;10:12598. [Crossref]

18. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In *Adv Neural Inf Process Syst.* 2014;27:1-9. [Crossref]

19. Nie D, Trullo R, Lian J, et al. Medical image synthesis with deep convolutional adversarial networks. *IEEE Trans Biomed Eng.* 2018;65:2720-2730. Erratum in: *IEEE Trans Biomed Eng.* 2020;67(9):2706. [Crossref]

20. Dosovitskiy A, Beyer L, Kolesnikov A, et al. an image is worth 16x16 words: transformers for image recognition at scale. 2021. [Crossref]

21. Vision Transformer and Language Model Based Radiology Report Generation | IEEE Journals & Magazine | IEEE Xplore. [Crossref]

22. Shimron E, Perlman O. AI in MRI: computational frameworks for a faster, optimized, and automated imaging workflow. *Bioengineering (Basel)*. 2023;10:492. [Crossref]

23. Oscanoa JA, Middione MJ, Alkan C, et al. Deep Learning-based reconstruction for cardiac MRI: a review. *Bioengineering (Basel)*. 2023;10:334. [Crossref]

24. Improving spreading projection algorithm for rapid k-space sampling trajectories through minimized off-resonance effects and gridding of low frequencies - Giliyar Radhakrishna - 2023 - Magnetic Resonance in Medicine - Wiley Online Library. [Crossref]

25. Wu Y, Alley M, Li Z, et al. Deep learning-based water-fat separation from dual-echo chemical shift-encoded imaging. *Bioengineering (Basel)*. 2022;9:579. [Crossref]

26. Zou Q, Priya S, Nagpal P, Jacob M. Joint cardiac T1 mapping and cardiac cine using manifold modeling. *Bioengineering (Basel)*. 2023;10:345. [Crossref]

27. Tolpadi AA, Bharadwaj U, Gao KT, et al. K2S Challenge: From Undersampled K-space to automatic segmentation. *Bioengineering (Basel)*. 2023;10:267. [Crossref]

28. Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: a survey. *IEEE Trans Med Imaging*. 2013;32:1153-1190. [Crossref]

29. Sokooti H, de Vos B, Berendsen F, et al. 3D convolutional neural networks image registration based on efficient supervised learning from artificial deformations. 2019. [Crossref]

30. Lee MCH, Oktay O, Schuh A, Schaap M, Glocker B. Image-and-spatial transformer networks for structure-guided image registration. 2019. [Crossref]

31. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging*. 2019;38:1788-1800. [Crossref]

32. Hess M, Allaire B, Gao KT, et al. Deep learning for multi-tissue segmentation and fully automatic personalized biomechanical models from BACPAC clinical lumbar spine MRI. *Pain Med*. 2023;24(Suppl 1):139-148. [Crossref]

33. Caliva F, Iriondo C, Martinez AM, Majumdar S, Pedoia V. Distance Map Loss Penalty Term for Semantic Segmentation. 2019. [Crossref]

34. Bonaldi L, Pretto A, Pirri C, Uccheddu F, Fontanella CG, Stecco C. Deep learning-based medical images segmentation of musculoskeletal anatomical structures: a survey of bottlenecks and strategies. *Bioengineering (Basel)*. 2023;10:137. [Crossref]

35. Hirvasniemi J, Runhaar J, van der Heijden RA, et al. The KNee OsteoArthritis Prediction (KNOAP2020) challenge: an image analysis challenge to predict incident symptomatic radiographic knee osteoarthritis from MRI and X-ray images. Osteoarthritis Cartilage. 2023;31:115-125. [Crossref]

36. Kijowski R, Liu F, Caliva F, Pedoia V. Deep learning for lesion detection, progression, and prediction of musculoskeletal disease. *J Magn Reson Imaging*. 2020;52:1607-1619. [Crossref]

37. Jamaludin A, Lootus M, Kadir T, et al. ISSLS prize in bioengineering science 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J*. 2017;26:1374-1383. [Crossref]

38. Heimann T, Meinzer HP. Statistical shape models for 3D medical image segmentation: a review. *Med Image Anal*. 2009;13:543-563. [Crossref]

39. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional neural networks for volumetric medical image segmentation. 2016. [Crossref]

40. Wu J, Zhang C, Xue T, Freeman WT, Tenenbaum JB. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. 2017. [Crossref]

41. Fritz B, Yi PH, Kijowski R, Fritz J. Radiomics and deep learning for disease detection in musculoskeletal radiology: an overview of novel MRI- and CT-based approaches. *Invest Radiol*. 2023;58:3-13. [Crossref]

42. Selles M, Wellenberg RHH, Slotman DJ, et al. Image quality and metal artifact reduction in total hip arthroplasty CT: deep learning-based algorithm versus virtual monoenergetic imaging and orthopedic metal artifact reduction. *Eur Radiol Exp*. 2024;8:31. [Crossref]

43. Seo S, Do WJ, Luu HM, Kim KH, Choi SH, Park SH. Artificial neural network for slice encoding for metal artifact correction (SEMAC) MRI. *Magn Reson Med*. 2020;84:263-276. [Crossref]

44. Kim JW, Kwon K, Kim B, Park H. Attention Guided Metal Artifact Correction in MRI using deep neural networks. 2019. [Crossref]

45. Hannun A, Case C, Casper J, et al. Deep speech: scaling up end-to-end speech recognition. 2014. [Crossref]

46. Do BH, Wu AS, Maley J, Biswal S. Automatic retrieval of bone fracture knowledge using natural language processing. *J Digit Imaging*. 2013;26:709-713. [Crossref]

47. Carrodeguas E, Lacson R, Swanson W, Khorasani R. Use of machine learning to identify follow-up recommendations in radiology reports. *J Am Coll Radiol*. 2019;16:336-343. [Crossref]

48. Liew C. The future of radiology augmented with artificial intelligence: a strategy for success. *Eur J Radiol*. 2018;102:152-156. [Crossref]

49. Syeda-Mahmood T. Role of big data and machine learning in diagnostic decision support in radiology. *J Am Coll Radiol*. 2018;15:569-576. [Crossref]

50. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88. [Crossref]

51. Lassau N, Bousaid I, Chouzenoux E, et al. Three artificial intelligence data challenges based on CT and MRI. *Diagn Interv Imaging*. 2020;101:783-788. [Crossref]

52. Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020;295:4-15. [Crossref]

53. Zając HD, Avlona NR, Kensing F, Andersen TO, Shklovski I. Ground Truth Or Dare: factors affecting the creation of medical datasets for training AI. In *Proc 2023 AAAIACM Conf AI Ethics Soc*. 2023:351-362. [Crossref]

54. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ*. 2016;353:i2139. [Crossref]

55. Sabih DE, Sabih A, Sabih Q, Khan AN. Image perception and interpretation of abnormalities; can we believe our eyes? Can we do something about it? *Insights Imaging*. 2011;2:47-55. [Crossref]

56. Brady AP, Neri E. Artificial intelligence in radiology-ethical considerations. *Diagnostics (Basel)*. 2020;10:231. [Crossref]

57. Link TM, Pedoia V. Using AI to improve radiographic fracture detection. *Radiology*. 2022;302:637-638. [Crossref]

58. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagn Interv Radiol*. 2024;30:80-90. [Crossref]

59. Brady AP, Allen B, Chong J, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: practical considerations. A multi-society statement from the ACR, CAR, ESR, RANZCR & RSNA. *Can Assoc Radiol J*. 2024;75:226-244. [Crossref]

60. Geis JR, Brady AP, Wu CC, et al.: Ethics of artificial intelligence in radiology: summary of the Joint European and North American Multisociety Statement. *Radiology*. 2019;293:436-440. [Crossref]

61. Cheplygina V, de Bruijne M, Pluim JPW. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med Image Anal*. 2019;54:280-296. [Crossref]

62. Doo FX, McGinty GB. Building Diversity, Equity, and Inclusion within radiology artificial intelligence: representation matters, from data to the workforce. *J Am Coll Radiol*. 2023;20:852-856. [Crossref]

63. Lotan E, Tschider C, Sodickson DK, et al. Medical imaging and privacy in the era of artificial intelligence: myth, fallacy, and the

future. *J Am Coll Radiol.* 2020;17:1159-1162. [Crossref]

64. Cestonaro C, Delicati A, Marcante B, Caenazzo L, Tozzo P. Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review. *Front Med (Lausanne).* 2023;10:1305756.[Crossref]

65. Lee YH. Efficiency improvement in a busy radiology practice: determination of musculoskeletal magnetic resonance imaging protocol using deep-learning convolutional neural networks. *J Digit Imaging.* 2018;31:604-610. [Crossref]

66. Trivedi H, Mesterhazy J, Laguna B, Vu T, Sohn JH. Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM Watson's Natural Language Processing Algorithm. *J Digit Imaging.* 2018;31:245-251. [Crossref]

67. Richardson ML. MR protocol optimization with deep learning: a proof of concept. *Curr Probl Diagn Radiol.* 2021;50:168-174. [Crossref]

68. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *Am J Roentgenol.* 2019;213:506-513. [Crossref]

69. Kurasawa H, Hayashi K, Fujino A, et al. Machine-learning-based prediction of a missed scheduled clinical appointment by patients with diabetes. *J Diabetes Sci Technol.* 2016;10:730-736. [Crossref]

70. Gorelik N, Gyftopoulos S. Applications of artificial intelligence in musculoskeletal imaging: from the request to the report. *Can Assoc Radiol J.* 2021;72:45-59. [Crossref]

71. Ozhinsky E, Liu F, Pedoia V, Majumdar S. Machine learning-based automated scan prescription of lumbar spine MRI acquisitions. *Magn Reson Imaging.* 2024;110:29-34. [Crossref]

72. Tolpadi AA, Luitjens J, Gassert FG, et al. Synthetic inflammation imaging with PatchGAN deep learning networks. *Bioenginering (Basel).* 2023;10:516. [Crossref]

73. Morales AG, Lee JJ, Caliva F, et al. Uncovering associations between data-driven learned qMRI biomarkers and chronic pain. *Sci Rep.* 2021;11:21989. [Crossref]

74. Bioengineering | Free Full-Text | Synthetic Knee MRI T1p Maps as an Avenue for Clinical Translation of Quantitative Osteoarthritis Biomarkers. [Crossref]