


REVIEW

Open Access



Artificial intelligence in orthopedics: fundamentals, current applications, and future perspectives

Jian Song^{1†}, Guang-Chao Wang^{1†}, Si-Cheng Wang^{2†}, Chong-Ru He¹, Ying-Ze Zhang^{3*}, Xiao Chen^{1,4*} and Jia-Can Su^{1,4*} 

Abstract

Conventional diagnostic and therapeutic approaches in orthopedics are frequently time intensive and associated with elevated rates of diagnostic error, underscoring the urgent need for more efficient tools to improve the current situation. Recently, artificial intelligence (AI) has been increasingly integrated into orthopedic practice, providing data-driven approaches to support diagnostic and therapeutic processes. With the continuous advancement of AI technologies and their incorporation into routine orthopedic workflows, a comprehensive understanding of AI principles and their clinical applications has become increasingly essential. The review commences with a summary of the core concepts and historical evolution of AI, followed by an examination of machine learning and deep learning frameworks designed for orthopedic clinical and research applications. We then explore various AI-based applications in orthopedics, including image analysis, disease diagnosis, and treatment approaches such as surgical assistance, drug development, rehabilitation support, and personalized therapy. These applications are designed to help researchers and clinicians gain a deeper understanding of the current applications of AI in orthopedics. The review also highlights key challenges and limitations that affect the practical use of AI, such as data quality, model generalizability, and clinical validation. Finally, we discuss possible future directions for improving AI technologies and promoting their safe and effective integration into orthopedic care.

Keywords Artificial intelligence (AI), Orthopedics, Machine learning (ML), Deep learning (DL), Diagnostic, Therapeutics

[†]Jian Song, Guang-Chao Wang and Si-Cheng Wang contributed equally to this work.

*Correspondence:

Ying-Ze Zhang
yzzhang@hebm.edu.cn
Xiao Chen
sirchenxiao@sjtu.edu.cn
Jia-Can Su
drsujacan@163.com

¹ Department of Orthopedics, Trauma Orthopedics Center, Institute of Musculoskeletal Injury and Translational Medicine of Organoids, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China

² Department of Orthopedics, Shanghai Zhongye Hospital, Shanghai 200941, China

³ Department of Orthopedics, Orthopedic Research Institution of Hebei Province, NHC Key Laboratory of Intelligent Orthopedic Equipment, The Third Hospital of Hebei Medical University, Shijiazhuang 050051, China

⁴ Institute of Translational Medicine, Organoid Research Center, National Center for Translational Medicine (Shanghai), Shanghai University Branch, Shanghai 200444, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Artificial intelligence (AI) has experienced rapid progress in recent years, contributing to changes in multiple areas of medical practice through its ability to simulate human reasoning and aid in clinical judgments [1]. In 2016, preventable medical errors were reported to account for more than 250,000 deaths annually in the United States [2]. In health care, AI holds immense promise for enhancing the quality of care and minimizing such errors [3, 4]. Improvements in computing power and data access have led to rapid increases in the development of AI technologies and related medical devices, many of which have been approved by the United States Food and Drug Administration [5]. These innovations are gradually being integrated into various medical specialties. In particular, in orthopedics, a field long recognized for embracing technological progress, AI has begun to be applied to increase diagnostic accuracy and improve patient outcomes.

Although AI is increasingly used to support early diagnosis and precise treatment in orthopedics [6], the integration of AI into orthopedic practice presents unique challenges. Orthopedic conditions often involve complex biomechanical systems, patient-specific anatomical variations, and long-term recovery periods, complexities that demand robust and precise AI solutions. Challenges such as data scarcity, the complexity of unstructured clinical information, and integration issues with current workflows continue to restrict the practical use of AI in medicine. Moreover, ensuring the ethical use of patient data, regulatory compliance, and acceptance among healthcare providers further adds to the complexity.

Although challenges remain, various aspects of AI applications in orthopedics continue to show substantial potential. AI technologies enable more accurate diagnoses through advanced image recognition, optimize surgical interventions via predictive analytics, and personalize treatment plans using patient-specific data. Particularly in surgical applications, AI-based robotics and navigation systems are transforming the way orthopedic surgeries are planned and executed, improving precision and reducing recovery time [7, 8]. Collectively, these innovations align with the broader goals of precision medicine [9], offering tailored interventions to meet individual patient needs.

In recent years, interest in applying AI in orthopedic surgery has increased [6, 10–12]. This review examines current applications of AI in orthopedic diagnosis and treatment with a focus on its clinical benefits, technical limitations, and areas requiring further research and development (R&D). By identifying the current gaps in research and practice, we highlight key areas where AI has the potential to enhance orthopedic care and guide

future efforts toward more effective integration of AI technologies.

Definitions of AI

The development of AI has been a gradual process [13] (Fig. 1). The foundation of AI traces back to McCarthy's work in 1956, which marked the beginning of efforts to build machines that could emulate human thinking, imagined then as "thinking, feeling machines" [14–17] (Fig. 2). However, it was not until the 1980s that computerization and the automation of processes enabled AI to develop more widely [7]. Since the early 2000s, AI has rapidly advanced. By leveraging large datasets, deep learning (DL) has enabled machines to perform advanced tasks that require higher levels of pattern recognition and abstraction [18], with remarkable results in natural language processing, image recognition, and speech recognition. The past two decades have witnessed a steady increase in AI applications across various industries, including health care [19–21]. In orthopedics, the application of AI enhances diagnostic accuracy, enables personalized treatment planning, and supports predictive modeling, collectively improving patient outcomes and experiences.

Machine learning (ML)

ML is a crucial subset of AI that empowers computer systems to learn from experience and acquire knowledge from datasets (Fig. 2). ML leverages algorithms to predict outcomes by identifying patterns in input data, and these predictions are compared with known outcomes to quantify and refine the accuracy of the algorithm [22]. ML manifests in two primary forms: supervised and unsupervised learning [22–24]. In supervised learning, algorithms are trained via labeled datasets [25]. The model learns to make predictions based on the provided examples, thus enabling autonomous decision-making. Supervised learning focuses on classification and prediction [26]. Conversely, unsupervised learning involves training algorithms on unlabeled data, allowing them to uncover hidden patterns or structures independently [27]. Unsupervised ML techniques are commonly employed for clustering similar data points or identifying relationships between variables within a dataset.

In health care, both supervised and unsupervised learning strategies have been extensively applied [26, 28]. Supervised ML facilitates the development of predictive models, enhancing capabilities in areas such as disease detection and therapeutic outcome forecasting [28, 29]. Conversely, unsupervised ML techniques enable the discovery of latent patterns within health care data, leading to insights such as disease subtyping and patient stratification [26, 28]. Recent studies have effectively employed

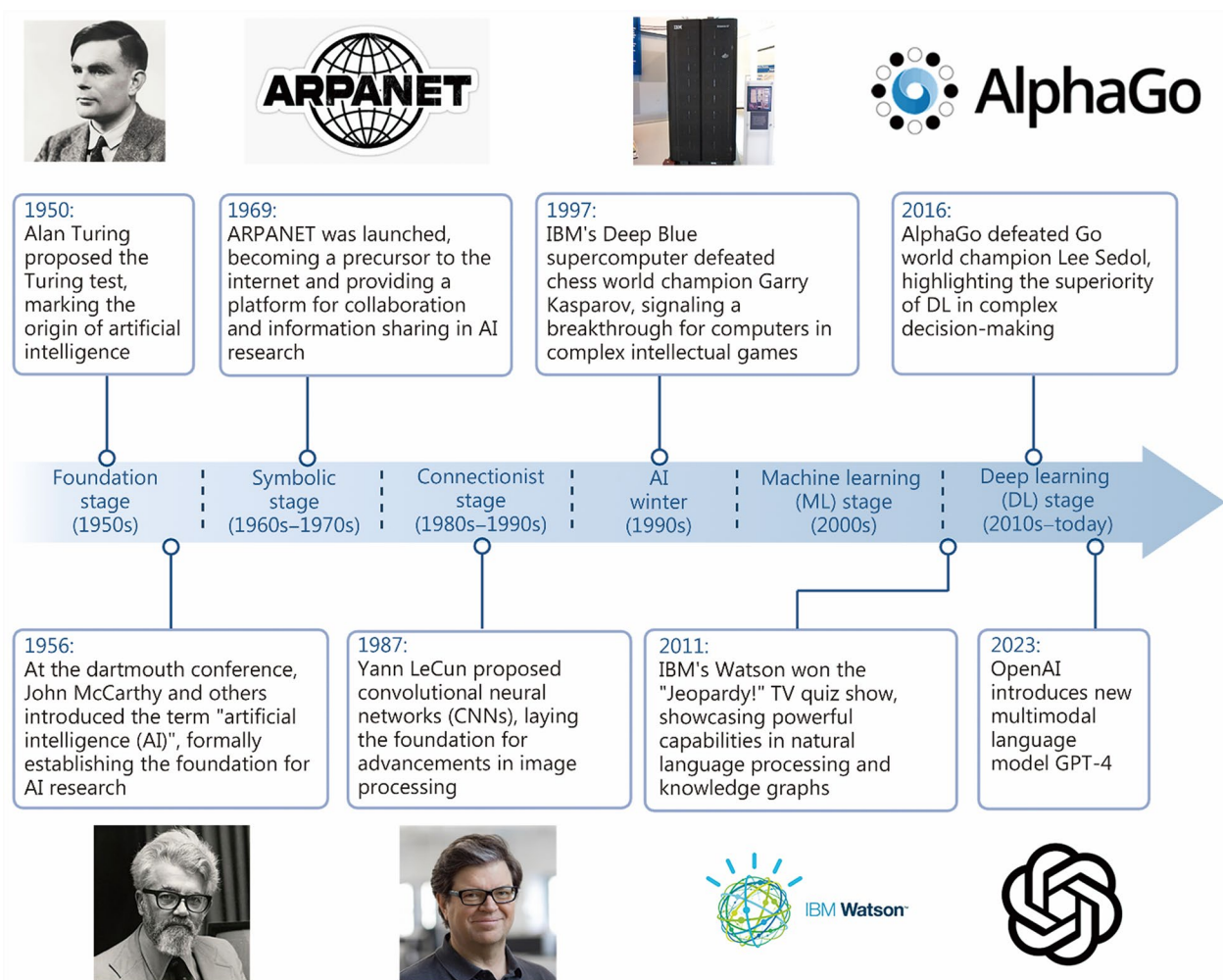


Fig. 1 A brief history of AI development. GPT-4 generative pre-trained transformer 4

supervised learning methods across various applications, including injury risk prediction among professional athletes, evaluation of knee osteoarthritis (OA) through kinematic analysis [30, 31], and prediction of clinically relevant outcomes [32–36]. In contrast, unsupervised learning has been used to identify distinct patient subgroups, aiding in risk stratification and personalized treatment. A recent study applied unsupervised clustering to knee ligament registry data and identified 5 patient subgroups with distinct anterior cruciate ligament (ACL) revision rates. The classification was based on age, graft type, and preoperative function scores [37].

DL

DL is a multilayer representation learning method that transforms raw data into abstract features through structures that mimic human brain neurons, enabling automatic pattern recognition and classification (Fig. 2). DL utilizes artificial neural networks to extract complex

features from large, high-dimensional data [18]. These networks consist of millions of interconnected neurons arranged in layers. Each neuron receives input from the previous layer and passes processed information to the next. During training, data with known outcomes are fed into the input layer and processed through hidden layers before a prediction at the output layer is produced [38].

There are several foundational types of neural networks, each suited to different data types and tasks. The multilayer perceptron forms the basis for many DL models through its fully connected structure [20]. Recurrent neural networks are designed to handle sequential data by processing one element at a time while retaining information from previous inputs through hidden states [39]. In contrast, convolutional neural networks (CNNs) are particularly effective for image analysis because of their ability to capture local spatial features [38]. These models are often used independently or combined as components in more advanced DL architectures. These neural

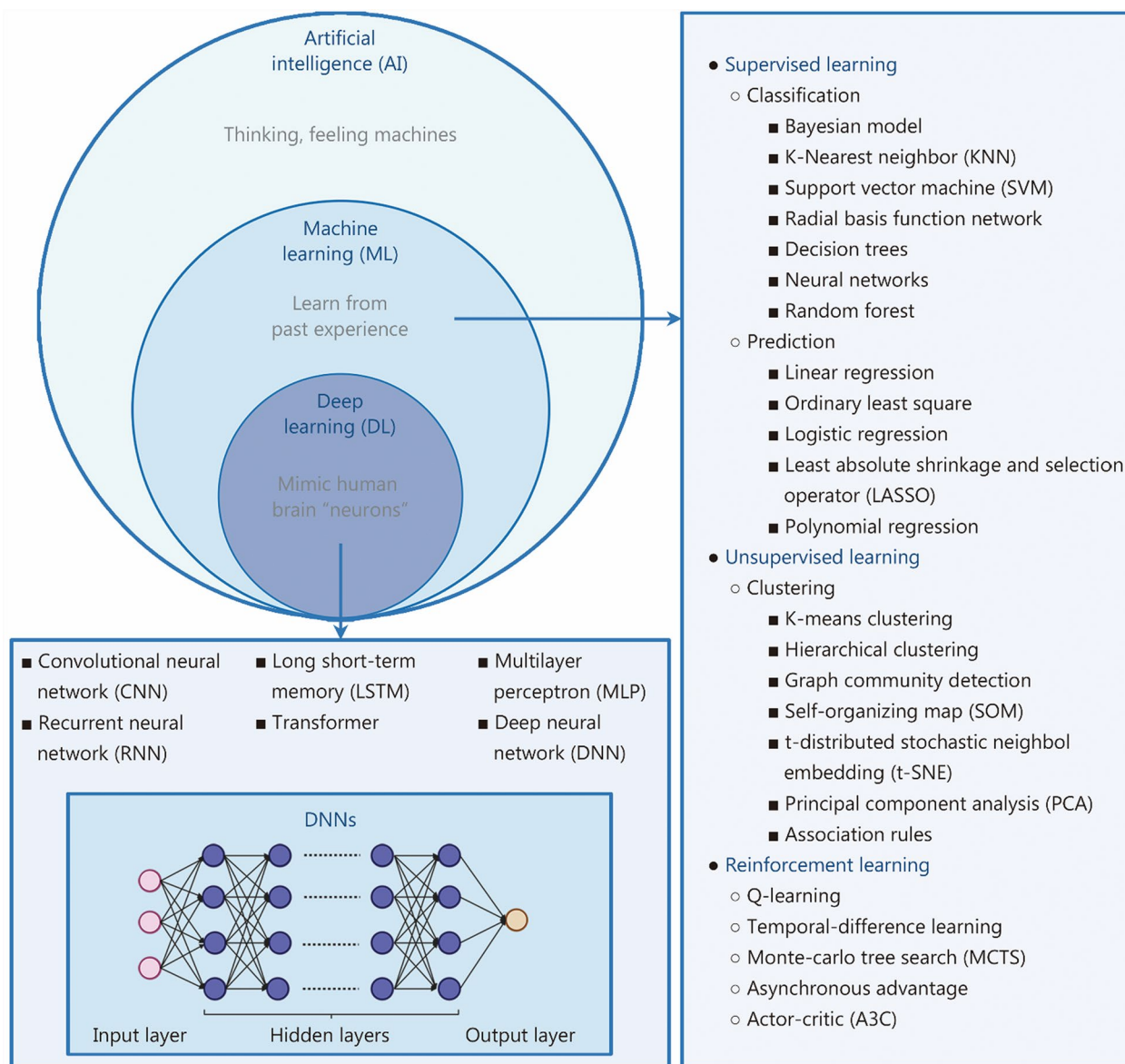


Fig. 2 The relationship between AI, ML, and DL and commonly used algorithms as examples

network architectures have enabled significant advances in various fields, including orthopedics, where DL has been widely applied to tasks such as image analysis [40, 41], diagnosis [42–44], surgery [45], drug development [46–48], and predictive analytics [49].

Algorithm development for orthopedic image analysis

Recent developments in AI have facilitated progress in orthopedic imaging, particularly in the automated segmentation and localization of anatomical structures (Table 1) [40, 41, 50–58]. AI offers practical tools for

streamlining these complex tasks and supports the application of ML and DL techniques in the diagnosis and treatment of orthopedic conditions.

Progress in segmentation and classification algorithms

Orthopedic image analysis has benefited from the refinement of segmentation algorithms that enable the delineation of structures such as cartilage, bone, muscle, and neural elements. These advances are grounded in both traditional ML models and modern DL architectures.

For articular cartilage segmentation, Shah et al. [58] utilized a validated ML model to quantify cartilage

Table 1 Application of AI in the analysis of orthopedic images

Algorithm	Application	References
ML model	Measuring and segmenting articular cartilage thickness in healthy knees	[58]
DL model	Segmenting cartilage and meniscus in knee MR images	[41]
DCNN	Segmenting pelvic muscles, fat, and bone	[40]
U-Net and SegResNet architectures	Segmenting femur	[53]
ML model	Identifying and labeling vertebrae and intervertebral discs on MR images	[57]
Algorithms based on regression forests and probabilistic graphical models	Locating and identifying vertebrae in CT scans	[54]
ML model	Segmenting the spine and identifying the pedicles	[50]
DL model	Segmenting lumbosacral nerves	[52]
CNN	Identifying and labeling intervertebral discs and predicting multiple pathology grades	[55]
ResNet-based neural network	Identifying and classifying various nested fracture categories	[56]
DL model	Classifying foot types	[51]

ML machine learning, DL deep learning, MR magnetic resonance, DCNN deep convolutional neural networks, CNN convolutional neural networks, CT computed tomography

thickness from 3910 magnetic resonance imaging (MRI) datasets of healthy knees. This approach enabled differentiation between tissue layers, supporting longitudinal observations of joint degeneration. However, ML-based models often lack adaptability, and their dependence on manual feature extraction limits scalability. In contrast, Norman et al. [41] implemented a U-Net-based DL model for segmenting cartilage and menisci. The model achieved Dice coefficients between 0.770 and 0.878 for cartilage and 0.753 and 0.809 for menisci, with segmentation times averaging 5 s per scan. It also showed a strong correlation with manual relaxometry and morphology measurements, demonstrating its potential for reliable OA assessment in clinical workflows (Table 1).

For bone structure segmentation, Ghidotti et al. [53] compared the U-Net and SegResNet architectures and achieved high spatial accuracy. SegResNet, with its enhanced skip connections, demonstrated superior adaptability under specific training conditions, underscoring its potential for precise three-dimensional (3D) anatomical modeling in orthopedic applications. Similarly, in pelvic segmentation, Hemke et al. [40] developed a deep convolutional neural network (DCNN) model capable of segmenting pelvic computed tomography (CT) images into multiple tissue classes with high accuracy, completing segmentation of each image in under 0.1 s on a graphics processing unit. These results highlight the robustness and efficiency of the model, reinforcing the value of DL in comprehensive body composition analysis (Table 1).

Spinal segmentation, particularly under pathological and incomplete conditions, remains challenging. Early approaches, such as Oktay and Akgul's Markov-chain-like graphical model with handcrafted pyramidal

histogram of oriented gradient (PHOG) and image projection descriptor features, addressed disc and vertebra localization in abnormal magnetic resonance datasets [57]. Building on this, Glocker et al. [54] and Burström et al. [50] introduced automated methods for CT and cone-beam computed tomography (CBCT) that facilitated pedicle screw navigation. To overcome the limitations of handcrafted features, Jakubicek et al. [59] proposed an innovative multistage DL framework integrating 3 CNNs with novel spatial filtering and global optimization methods, achieving a mean intervertebral disc localization error of 4.4 mm and 87.1% vertebra labeling accuracy on challenging pathological and incomplete 3D CT spine scans. Complementing these efforts, Fan et al. [52] demonstrated high-accuracy segmentation of lumbosacral nerves and bone, with performance statistically comparable to that of manual segmentation, reinforcing the clinical potential of DL-based models. Across studies, multistage CNN frameworks exhibit better resilience than other frameworks in handling abnormal spine morphology (Table 1).

Beyond anatomical segmentation and localization, AI has also been increasingly applied to automated detection and classification tasks, expanding its utility in clinical decision-making. Jamaludin et al. [55] developed a CNN-based system that achieved 95.6% accuracy in disc detection and labeling from MR images, and its pathological grading predictions closely matched radiologists' evaluations, offering consistent voxel-level "evidence hotspots" to support clinical interpretation. Similarly, Lind et al. [56] created a ResNet-based neural network that classified 49 nested fracture categories around the knee, achieving weighted mean area under the curves (AUCs) of 0.87 for proximal tibia fractures, 0.89 for patella

fractures, and 0.89 for distal femur fractures, with nearly 75% of the AUC estimates exceeding 0.8. These results demonstrate the model's capacity for detailed and reliable fracture classification based on radiographic data. Expanding into multimodal data, Chae et al. [51] combined imaging and digital foot pressure data to classify foot types, demonstrating the adaptability of AI beyond conventional imaging tasks (Table 1).

These studies illustrate the growing potential and complexity of AI in orthopedic diagnostics, particularly when diverse data types are integrated and atypical pathologies are addressed. Despite these advancements, certain limitations and challenges remain to be addressed to ensure consistent and accurate performance across diverse clinical scenarios.

Technical limitations in image analysis

Despite recent advancements in image segmentation and its increasing accuracy and efficiency in identifying anatomical structures and assisting in clinical decisions, several persistent limitations remain across current studies. These limitations are related to data dependency, generalizability, model design, and clinical applicability. To begin with, a number of investigations have faced difficulties related to annotation accuracy [41, 56]. Certain works have used manual segmentation as the reference standard, which is inherently subjective and lacks consistency across annotators, thus limiting its validity as a ground truth [41]. Reliance solely on radiology reports for annotation, without corroborating evidence from CT, MRI, or surgical findings, raises the risk of misclassification, particularly when dealing with complex anatomical structures or fine-grained subclassifications [56]. Furthermore, the representativeness and pathological diversity of training datasets are often insufficient. Studies based on data from a single institution may be influenced by specific imaging protocols, patient populations, and equipment settings, limiting their broader applicability [40, 56]. Additionally, certain works have limited the scope of segmentation to single anatomical planes or specific structures, such as assessing muscle quality using only a standard axial hip slice [40] or segmenting exclusively the L5/S1 spinal level, which prevents the integration of multiview or 3D information essential for comprehensive analysis [41, 55]. For example, Burström et al. [50] trained a model on cadaveric images without overt pathology, which, although effective for general anatomical segmentation, may underperform in cases of disease such as spinal deformities or degenerative lesions.

Equally important is the lack of standardization in evaluation metrics, which limits comparability across studies. Some studies have prioritized classification accuracy but neglected consistency metrics such as the kappa

coefficient, potentially obscuring performance differences across categories and limiting transparency regarding clinical reliability [55]. Moreover, the clinical utility of many existing models has only been preliminarily evaluated and still needs confirmation through future large-scale, prospective validation. For example, Lind et al. [56] proposed DL-assisted plane localization for planning epidural steroid injections, but its feasibility remains to be established through extensive trials. Similarly, Hemke et al. [40] acknowledged the need to validate the prognostic utility of muscle quality assessment models in patients with cancer or chronic diseases.

AI-powered innovations in orthopedic diagnostics

Diagnostic errors occur across all phases of surgical care and specialties, with clinical decision-making and communication breakdown as leading contributors, resulting in at least moderate harm in over half of cases and death in 1 in 7 patients [49, 60]. The accelerated advancement of AI technologies, coupled with the vast expansion of medical data [61], has driven increasing integration of AI into diagnostic research and clinical applications [62].

Facilitating precise and efficient fracture diagnosis

Fracture diagnosis remains a cornerstone of orthopedic practice, yet conventional radiographic interpretation is prone to misdiagnosis due to interobserver variability, complex anatomical structures, and heavy clinical workloads [63, 64]. The global age-standardized incidence rate of fractures in 2019 was estimated at 2296.2 cases per 100,000 people [65], and diagnostic errors can lead to delayed treatment and poor recovery of function [44, 66]. The incorporation of AI into fracture diagnosis represents a promising advancement aimed at enhancing diagnostic precision, promoting consistency in interpretation, and streamlining clinical processes [67, 68]. Many studies have shown that AI-based models perform strongly in detecting acute fractures, achieving accuracies comparable to those of expert radiologists (Table 2) [42–44, 69–78].

AI models trained on X-ray and CT datasets have successfully identified fractures of varying sizes and complexities. In a study by Chung et al. [42], a DCNN demonstrated proficiency in detecting proximal humerus fractures. Similarly, Lindsey et al. [44] trained a deep neural network (DNN) capable of detecting and localizing wrist fractures in radiographs with diagnostic accuracy comparable to that of advanced subspecialty plastic surgeons. This study also demonstrated that emergency medicine clinicians aided by the trained model achieved a sensitivity of 92.5% and a specificity of 94.1%. Another study by Kim et al. [43] demonstrated that transfer learning, where a CNN

Table 2 Application of AI in the diagnosis of fractures

Fracture site	Algorithm	References
Fractures of upper limb bones	CNN, DNN	[42–44, 71]
Rib fractures	DL model	[75, 78]
Thoracolumbar fractures	ML model	[70, 74]
Hip fractures	CNN	[69, 77]
Scaphoid fractures	DL model	[73]
Calcaneal fractures	CNN	[76]
Fractures across the musculoskeletal system	DL model	[72]

CNN convolutional neural network, DNN deep neural network, DL deep learning, ML machine learning

pretrained on nonmedical images was fine-tuned on medical datasets, could produce highly accurate fracture diagnoses, even with limited labeled radiographs. Furthermore, Guan et al. [71] proposed a novel DL approach for the X-ray-based detection of arm fractures and achieved a state-of-the-art average accuracy of 62.04% for arm fracture detection, even for a musculoskeletal radiograph (MURA) dataset with images of low quality. In contrast to studies that reported higher sensitivity and specificity based on high-resolution data or more distinct anatomical targets, the study by Guan et al. [71] focused on overcoming real-world imaging constraints. Following the application of AI models in upper limb fracture detection, AI models have also been applied for detecting rib fractures, an especially challenging task owing to the high miss rate of up to 50% on conventional chest radiographs, which are limited by overlapping structures, reader variability, and image quality [79, 80]. Niiya et al. [75] introduced a system that achieved a sensitivity of 93.5%, significantly improving fracture detection efficiency in trauma settings. Similarly, Yao et al. [78] constructed a DL-based rib fracture detection system using a 3-step algorithm.

Moreover, AI has shown promise in the diagnosis of thoracolumbar vertebral fractures, which are frequently missed or misdiagnosed on plain radiographs because of their suboptimal sensitivity, potentially leading to delayed treatment and a worse prognosis [81, 82]. Furthermore, vertebral compression fractures, particularly those involving significant anterior height restoration and low bone mineral density, not only increase the risk of subsequent adjacent vertebral fractures but also significantly increase the risk of future hip fractures [83, 84]. Therefore, early diagnosis of thoracolumbar compression fractures is crucial. Burns et al. [70] utilized an automated ML computer system to detect spinal fractures with 95.7% sensitivity. Additionally, Li et al. [74] developed an AI-based lumbar vertebral fracture

detection system that demonstrated 92% accuracy, 91% sensitivity, and 94% specificity.

DL algorithms have been widely applied to multisite fracture detection. Studies have shown that CNN and ResNet models achieve high accuracy in diagnosing fractures of the femoral neck, hip, scaphoid, and calcaneus [63, 73, 76, 77]. Jones et al. [72] reported an AUC of 0.974 for a DL model targeting 16 anatomical fracture sites. The sensitivity and specificity were 95.2% and 81.3%, respectively. In a secondary analysis of radiographs with complete expert agreement, the AUC further increased to 0.993, which closely approximates the diagnostic accuracy of experienced radiologists and orthopedic surgeons.

Nonetheless, AI-based fracture diagnosis still faces several unresolved challenges [85]. First, real-world, multi-center validation is essential to ensure that AI systems perform reliably in clinical settings, as unexpected behaviors may not surface during standard evaluations [86]. Second, cost concerns and lack of transparency remain key barriers to AI adoption in radiology. Real-world improvements in clinical efficiency and diagnostic quality are essential to justify investment and promote broader implementation [87]. Furthermore, current AI systems struggle to detect subtle or atypical fractures, such as stress fractures or those in pathologic bone, owing to limited training exposure. Moreover, most AI tools are narrowly designed for specific tasks such as image classification and fall short of replicating the full scope of radiologists' responsibilities, which include clinical correlation and complex decision-making [85].

Assisting in the early diagnosis of developmental dysplasia of the hip (DDH)

DDH, the most common pediatric musculoskeletal disorder, can lead to severe disability if not diagnosed and treated early; timely and accurate detection is therefore critical for preserving hip joint function and preventing long-term complications [88, 89]. Traditional diagnostic methods, such as hip ultrasound for infants and pelvic radiographs for older children, rely heavily on clinician expertise, making them time-consuming, labor-intensive, and prone to variability [90]. The application of AI in imaging-based DDH screening contributes to greater diagnostic accuracy and improves workflow efficiency. Huang et al. [91] introduced DDHnet, a fully automated AI system that measures 4 key hip parameters (α -angle, β -angle, femoral head coverage, and pubofemoral distance) from ultrasound images. DDHnet reached a diagnostic accuracy of 98.64%, with 100.00% specificity and 90.56% sensitivity. These results demonstrate the ability of the system to deliver precise and consistent measurements while significantly reducing the time required for

analysis. Xu et al. [92] further developed an AI model that measures additional parameters such as the acetabular index, center–edge angle, Tönnis grade, and International Hip Dysplasia Institute (IHDI) grade, demonstrating diagnostic performance comparable to that of orthopedic specialists but with a significantly reduced processing time, the model required only 1.21 s per case, compared with 150.36 to 200.71 s for surgeons of varying experience levels ($P < 0.001$).

AI also has the potential to increase access to DDH screening in diverse clinical settings. Jaremko et al. [93] published a study showing that MEDO-Hip (MEDO.ai, Edmonton, Canada, 2021), an AI-based ultrasound evaluation software program, enables primary care workers with simple training to screen infants for DDH. This study demonstrated that the AI-based portable ultrasound screening can achieve follow-up and case detection rates similar to those of specialized ultrasound screening. MEDO-Hip facilitates population-level DDH screening, which is expected to reduce screening costs and enable widespread implementation.

AI performs well in objective measurements but struggles with subjective assessments [92]. More training data is needed to improve its ability in complex cases. Image quality remains a critical factor, as poor-quality scans can lead to inaccurate results. AI systems are highly dependent on the presence of all key anatomical landmarks. Moreover, most studies are on a small scale and rely on specific software–hardware combinations, limiting their generalizability to other clinical settings and devices [93–95].

Enhancing soft tissue injury diagnosis

Meniscal tears resulting from trauma and degeneration frequently cause knee pain in patients [96]. Accurate diagnosis and appropriate treatment are essential for improving patients' quality of life [97]. The CNN model proposed by Bien et al. [98] for ACL and meniscal tear detection achieved AUCs of 0.965 and 0.847, respectively, on the internal validation dataset. However, the model was evaluated with only one MR image, limiting its clinical application. To address these limitations, Fritz et al. [99] conducted a clinical validation study of a fully automated DCNN for detecting surgically confirmed meniscal tears. Compared with surgical findings and assessments by musculoskeletal radiologists, the DCNN model demonstrated fully automated detection of meniscal tears, achieving similar specificity but reduced sensitivity. Pedoia et al. [100] used a CNN for the diagnosis of meniscal injuries. They achieved 80% accuracy and reported a tendency for the misdiagnosis of posterior horn meniscal tears.

Among sport-related musculoskeletal injuries, ACL tears are common but challenging to diagnose [101, 102]. Researchers have developed various diagnostic models to aid in clinical analysis. The groundbreaking study by Štajduhar et al. employed a semiautomated method using MRI data to detect ACL injuries [103]. The research utilized histograms of oriented gradients (HOGs) and global image statistics methods for feature extraction, combined with support vector machines (SVMs) and random forests, and achieved an AUC of 0.943 in detecting ACL injuries, including mild and complete ruptures. Richardson et al. [104] demonstrated that CNNs can effectively replace human readers in MRI protocol optimization, particularly in detecting ACL tears, where fat-saturated sequences outperform non-fat-saturated sequences in terms of sensitivity. To enhance generalizability, Tran et al. [105] trained the algorithm using a multicenter dataset consisting of 19,765 knee MRI scans from 12 centers, incorporating diverse scanner types (1/1.5/3 T) and imaging protocols, demonstrating high performance across external datasets.

In addition to MR image-based diagnosis, many studies have established diagnostic models based on other criteria. Liu et al. [106] utilized arthroscopy as a reference standard, Zeng et al. [107] explored gait analysis to train their model, and Li et al. [108] based their approach on plantar pressure monitoring. These alternative methods expand the potential applications of AI beyond traditional imaging and contribute to the development of comprehensive, multimodal diagnostic strategies.

Although some models are trained on large and diverse datasets, their performance is often evaluated on the basis of data from single institutions, raising concerns about generalizability. Notably, improved diagnostic accuracy following retraining on external datasets suggests that model adaptation to local imaging protocols and population characteristics may be necessary [98]. These limitations underscore the need for larger, multi-institutional studies with standardized reference standards and broader clinical validation to ensure reliable real-world performance of AI models in musculoskeletal imaging.

Improving OA diagnosis and grading

In 2020, OA accounted for 4.3% of the number of years lived with disability worldwide, reflecting a 9.5% increase since 1990 and underscoring its growing impact on public health worldwide [109]. Therefore, it is particularly important to automate the detection of OA via AI. Conrozier et al. [110] introduced a method to diagnose early OA by measuring the joint gap width. The disadvantage of this method is that it requires frequent interventions by the observer to select the region of interest and adjust

the bone edge detection. Unlike manual methods, which heavily depend on observer input to select regions of interest and adjust bone edge detection, this AI-driven algorithm provides an automated and objective measurement of joint space width, reducing user dependency and improving reproducibility [111]. Üreten et al. [112] developed a CNN-based computer-aided diagnostic method using transfer learning for hip OA detection on plain pelvic radiographs. Their model achieved 90.2% accuracy, 97.6% sensitivity, and 83.0% specificity, demonstrating the potential to assist clinicians with objective interpretation and reduce the need for advanced imaging.

Recent developments have extended AI applications to OA severity grading. To automate the evaluation of knee OA severity, Tiulpin et al. [113] introduced a DL framework utilizing CNN aligned with the Kellgren-Lawrence classification system. Their approach yielded a multiclass accuracy of 66.71%, a Cohen’s kappa (quadratic) value of 0.83, and an AUC of 0.93, indicating substantial concordance with specialist-provided labels. Similarly, another automatic Kellgren-Lawrence grading model using the DenseNet neural network architecture achieved sensitivities from 68.9 to 86.0% and specificities ranging from 83.8 to 99.1% across different severity levels, indicating strong classwise discriminatory ability [114]. A DenseNet-based model was proposed by Pedoia et al. [115] for early-stage knee OA detection, employing T2-weighted MRI imaging to facilitate diagnosis before conventional radiographic signs are evident. When patient demographic data were integrated, their model exhibited 76.99% sensitivity and 77.94% specificity across a vast patient cohort from the OA Initiative baseline dataset. These studies highlight the utility of ML techniques, particularly DL-based approaches, in automating the detection and severity grading of OA from medical imaging data. These advancements hold promise

for facilitating early diagnosis, personalized treatment strategies, and improved clinical management of patients with OA.

AI-based advances in orthopedic treatment

Enhancing orthopedic surgery quality

The application of AI in orthopedic surgery has transformed the quality of care by improving precision, enhancing surgical planning, and streamlining intraoperative workflows [116–120]. Robotic platforms, navigation technologies, and AI-powered imaging analysis have become integral to optimize patient-specific interventions (Table 3) [45, 121–128].

Lower limb arthroplasty

AI has significantly advanced total hip arthroplasty (THA) and total knee arthroplasty (TKA) by improving preoperative planning, implant selection, and surgical precision. Chen et al. [121] investigated the relationship between the typology of DDH cases and the internal and external diameters of the socket via a 3D hip surgery simulation system, which demonstrated superior accuracy in prosthesis implantation compared with conventional two-dimensional (2D) templates. This transition from 2D to AI-enhanced 3D simulations allows for patient-specific solutions in complex developmental DDH cases, reducing errors in prosthesis placement.

For THA, AI-based tools have optimized implant recognition and alignment. Borjali et al. [45] developed a CNN-based algorithm which was evaluated on an independent test set of 252 anteroposterior hip radiographs and achieved 100% accuracy in identifying 3 common THR implant designs. These controlled test conditions demonstrate the model’s strong performance of the model, although further validation in larger clinical datasets is needed. Robot-assisted techniques have led

Table 3 Application of AI in orthopedic surgery

Surgery	Algorithm/robot	Application	References
THA	3D hip surgery simulation system	3D simulation for acetabular reconstruction in dysplastic hips	[121]
THA	CNN	Identifying the design of a failed THA implant within seconds before surgery	[45]
TKA	Robotic arm	Providing good positioning, improving prognosis, and reducing bone and soft tissue damage	[125]
TKA	Robotic arm	Improving the mechanical alignment, range of motion, and functional outcomes	[126]
RSA	Navigation technology	Improving the accuracy of glenoid component placement	[124]
SA	Robotic platforms	Real-time manufacturing of patient-specific instruments	[122]
Ankle joint reset	Image-guided robotic assistant	Assisting ankle reduction with the contralateral ankle as a reference	[123]
Tendon graft	Robotic biomechanical testing system	Simulating 3D kinematics and contact pressure	[127]
ACL reconstruction	Robot-assisted systems	Providing real-time feedback and automatic adjustments during surgery	[128]

THA total hip arthroplasty, 3D three-dimensional, CNN convolutional neural network, TKA total knee arthroplasty, RSA reverse shoulder arthroplasty, SA shoulder arthroplasty, ACL anterior cruciate ligament

to notable improvements in TKA, particularly in achieving more accurate mechanical alignment, minimizing intraoperative trauma, and enhancing early postoperative recovery. Systems such as Zimmer Biomet's Robot of Stereotactic Assistant (ROSA)[®] and Stryker's Mako[®] systems provide real-time haptic feedback during bone cutting and soft tissue balancing, enhancing intraoperative decision-making. Kayani et al. [125] confirmed that robotic arm-assisted TKA leads to significantly reduced surgical trauma compared with conventional jig-based TKA. Specifically, patients in the robotic group experienced lower postoperative pain and analgesia requirements and smaller reductions in postoperative hemoglobin levels, suggesting less intraoperative blood loss and faster functional recovery, including a shorter time to straight leg raise and earlier discharge. Rossi et al. [126] confirmed that robotic-assisted TKA improved mechanical alignment and range of motion in patients with severe varus and valgus deformities, achieving a 100% short-term survival rate with no major complications at a minimum follow-up of 6 months. Collectively, these advancements highlight the critical roles of AI and robotics in THA and TKA. AI tools aid in preoperative planning, ensuring accurate implant selection and alignment, whereas robotic systems provide real-time intraoperative guidance, improving precision and minimizing complications.

However, despite the significant role of AI and robotics in joint replacement, current research still has limitations. Limited sample size and few complex cases reduce model generalizability and recognition accuracy [45, 121]. Multicenter and multioperator involvement introduce variability, affecting measurement standardization [121, 125]. Lack of blinding, absence of correlation between short- and long-term outcomes, and inconsistent anesthesia and rehabilitation protocols restrict generalizability [125]. Future research should expand samples, standardize procedures, incorporate multimodal imaging, and strengthen long-term follow-up to increase the clinical value of these technologies.

Shoulder arthroplasty

The integration of navigation systems and robotic platforms in reverse shoulder arthroplasty has improved the accuracy of glenoid component placement, reducing manual errors such as improper inclination and off-center positioning. Giorgini et al. [124] demonstrated that computer-assisted navigation significantly improves the accuracy of glenoid component placement, reducing the number of errors associated with manual techniques. However, as Twomey-Kozak et al. [129] noted, despite these promising findings, the application of navigation systems in shoulder arthroplasty remains in its early

stages, especially compared with their established use in hip and knee replacements. Darwood et al. [122] emphasized the benefits of robotic platforms for the real-time manufacturing of patient-specific instruments, reporting that the system achieved version and inclination angle accuracies of 1.9° [standard deviation (SD) 1.3] and 1.2° (SD 0.7), respectively, along with a positional accuracy of 1.1 mm (SD 0.7) relative to the preoperative plan.

Although these innovations hold promise, shoulder arthroplasty faces unique challenges that impact the effectiveness of AI-based navigation. Giorgini et al. [124] highlighted that humeral component navigation is still underdeveloped, limiting the ability to achieve optimal joint stability. Surgeons must manually align the humeral implant, which introduces variability in surgical precision. Additionally, the initial adoption of AI-based navigation systems increases the operative time, as surgeons must adapt to new workflows. Future research should aim to expand navigation capabilities to both the humeral and glenoid components, optimizing joint biomechanics and implant longevity. The integration of augmented reality and virtual reality for intraoperative guidance could enhance real-time surgical precision and efficiency [129].

Fracture reduction

Achieving excellent reduction is crucial for fracture healing. Ankle trauma is highly prevalent, with an estimated 2 million acute sprains occurring annually in the United States [130]. Due to the complexity of the anatomy, even minor injuries can lead to serious consequences [131], and up to 70% of patients may experience residual physical disability. Improper reduction of the joint during surgery can result in an insufficient contact area or excessive contact force, thereby increasing the risk of postoperative complications [132, 133]. AI-based fracture reduction has been developed to minimize intraoperative fluoroscopy exposure while improving reduction accuracy. Gebremeskel et al. [123] proposed an image-guided robotic system for ankle fracture reduction, utilizing the contralateral ankle as a reference to determine optimal manipulative forces and displacement parameters. This study quantified the manipulative forces and their displacement necessary to reduce the ankle symphysis accurately, marking a crucial first step in defining the design requirements for robotic assistance in terms of force and displacement. Robotic-assisted fracture reduction has notable limitations. The effectiveness of AI-based force prediction models is dependent on the variability of individual patient anatomy, and real-time adaptation remains a challenge. Additionally, current robotic reduction techniques lack intraoperative feedback mechanisms, which may lead to unintended overcorrection or soft tissue

complications. Future efforts should focus on real-time AI adjustments using intraoperative fluoroscopic or CT-based feedback systems to ensure optimal reduction without excessive manual interventions.

Ligament reconstruction

AI has improved ligament reconstruction procedures by enhancing surgical planning and intraoperative precision. Sakakibara et al. [127] introduced a robotic biomechanical testing system for simulating joint kinematics and tendon graft placement, suggesting reconstruction at 30° of plantar flexion for optimal outcomes. Additionally, Yang et al. [128] utilized AI-based 3D CT and MRI analysis for personalized ACL reconstruction planning, integrating robotic assistance for real-time surgical adjustments. Their study demonstrated that robotic guidance improved bone tunnel drilling accuracy, maintaining deviations within 1.5 mm.

Accelerating drug development

New drug development is a long and complex process characterized by high R&D costs and significant uncertainty [134]. In recent years, traditional methods of new drug R&D have become increasingly difficult, with increasing investments and extended development times. The field is currently at a bottleneck stage, necessitating new technologies to achieve cost reduction and increased efficiency. The rapid advancement of AI technology has introduced new opportunities for high-quality developments in the biopharmaceutical industry [135]. As early as the 1980s, Merck began to design drugs using computer-aided design [136]. With the progression of computer technology, AI has gradually predominated, becoming increasingly involved in the drug development process [137]. This approach ensures the quality of analysis while significantly reducing drug R&D costs, shortening development time, and improving overall efficiency, thereby putting new drug development on a fast and efficient path [138, 139]. AI has found applications in predicting protein structure and functional properties [46–48, 140–142], forecasting drug–protein interactions [143, 144], and facilitating high-throughput drug screening [145].

Predicting protein 3D structures, properties, and functions

Understanding the 3D structures, properties, and functions of proteins is a critical step in drug discovery and development. An important example of such efforts is the study by Jumper et al. [46], who developed AlphaFold, a neural network model that provides a computational method for predicting protein structures with atomic accuracy, even in the absence of a known analogous structure. The results showed that, in most cases, its

accuracy is comparable to that of the experimental structures and greatly outperforms that of the other methods.

In 2023, Yuan et al. [142] proposed a DL framework that combines bidirectional temporal convolutional networks, bidirectional long short-term memory, and a multiscale bidirectional temporal convolutional network for secondary protein structure prediction, achieving superior performance over existing methods on the Critical Assessment of protein Structure Prediction 10 – 14 (CASP10 – 14) and CullPDB 513 (CB513) benchmark datasets. Wang et al. [48] developed a DL framework, a language model with a geometric vector perceptron, consisting of a protein language model and a graph neural network, which can make predictions about protein properties by utilizing one-dimensional amino acid sequences and 3D structural information of proteins. Gligorijević et al. [140] introduced deep functional residue identification (DeepFRI), a graphical convolutional network that outperforms current major methods and sequence-based CNNs. DeepFRI uses sequence features extracted from protein language models and protein structures to predict protein functions. AI technology has great potential for accurately predicting protein structures and functions, which will lead to unprecedented technological innovations for new drug development.

Predicting drug-protein interactions

Identifying the target proteins of a drug and predicting the drug-target protein interactions play extremely important roles in the drug development process. By predicting drug-receptor or -protein interactions, researchers can better understand the efficacy of a drug and design it most effectively. For example, Offensperger et al. [143] discovered hundreds of interactions between fragments and proteins through a large-scale chemical proteomics survey. They reported that the data generated are suitable for ML-based models, which, when the chemical structure is used as input, can predict how the chemical interacts with the native proteome in intact cells. Investigating protein–ligand interactions, Wang et al. [144] trained an SVM model on 15,000 ligand–protein interactions involving 626 proteins and 10,000 active compounds. This model successfully identified 9 new compounds and their interactions with 4 key targets (G protein-coupled receptor 4, sirtuin 1, p38 mitogen-activated protein kinase, and glycogen synthase kinase 3 beta). Predicting drug-protein interactions will accelerate the drug development process by reducing the time required for target screening.

High-throughput drug screening

Although AI has not yet been widely applied to screen drug targets for orthopedic diseases, its notable success

in other disease areas has driven a growing number of companies and large pharmaceutical firms to invest in AI-based drug discovery efforts [135]. The PandaOmics platform exemplifies how multiple AI engines can accelerate drug development by identifying promising targets [145]. In the case of INS018_055 [145], PandaOmics integrated multiomics data, literature trends, and biological network analysis to identify TNIK as a novel therapeutic target within just 18 months.

Although AI technology provides unprecedented opportunities for the discovery and development of novel orthopedic drugs by significantly improving the efficiency of drug screening and optimization, several challenges remain. First, in terms of data, there is a notable scarcity of high-quality, consistent, and accessible datasets [137, 146]. Unlike in fields such as image recognition, drug discovery suffers from limited annotated experimental data due to the inherent complexity of biological systems and variability in experimental conditions, leading to inconsistent and unreliable results. Second, challenges with molecular representation persist. Model performance is highly sensitive to the type and quality of molecular input used. Most current representations fail to account for essential aspects such as stereochemistry, conformational flexibility, the molecular surface, and compactness, all of which are crucial to molecular function [146]. Moreover, selecting an appropriate representation often involves a trade-off between simplicity and expressiveness, complicating the goal of model interpretability [137]. To fully realize the potential of AI, a collaborative effort is necessary to improve data quality, increase model transparency, and foster interdisciplinary collaboration.

Contributing to intelligent rehabilitation

Surgery is not the endpoint of orthopedic disease treatment. Postoperative complications such as joint stiffness and ossification can severely affect patient prognosis, making rehabilitation a crucial part of orthopedic postoperative care. With advancements in AI R&D, rehabilitation robots and other AI-based technologies are playing a significant role in assisting in rehabilitative treatment.

Recent advancements in intelligent rehabilitation systems have demonstrated the potential of AI to support complex motor function recovery. Averta et al. [147] designed a novel human-like motion generation algorithm that analyzes human motion through functional principal component analysis features, enabling efficient synthesis of complex movements such as free motion and obstacle avoidance in free space. Building upon the concept of human motion analysis, Zhao et al. [148] applied these principles to rehabilitation by designing a tele-rehabilitation system that integrates natural human responses with big data analytics. Their system includes

an upper limb rehabilitation robot that combines flexible ropes and exoskeleton components and is intended to assist clinicians in optimizing individualized rehabilitation programs. To further enhance rehabilitation management, they proposed a multidimensional training and assessment database to support a multilevel linked rehabilitation system. Extending robotic applications to lower extremity rehabilitation, Miller-Jackson et al. [149] designed a soft pneumatic actuator-driven exoskeleton for hip flexor rehabilitation, providing more options for individuals with lower extremity mobility issues. Compared with not wearing the device, subjects experienced a 43.5% reduction in muscle signals when lifting their legs while wearing the device, suggesting that the device is effective in assisting with hip flexion and that a pneumatic rotary actuator-driven exoskeleton is a viable solution. In conjunction with these hardware-based solutions, digital platforms such as smartphone-based applications are also playing an increasingly important role in rehabilitation. Rossi et al. [150] investigated the use of a smartphone-based care management platform, myMobility, in the context of TKA rehabilitation.

Personalized treatment planning

Precision medicine aims to tailor treatment strategies by classifying patients into subgroups based on variations in prognosis or treatment response, thereby optimizing therapeutic benefits while minimizing unnecessary interventions [151]. AI algorithms are well equipped to handle high-dimensional data, support predictive modeling, and inform personalized management strategies in clinical settings [152]. AI is transforming surgical planning and clinical decision-making by analyzing patient-specific data to predict outcomes, optimize procedures, and guide personalized strategies (Table 4) [32–36, 49, 153–160].

Predictive analytical models have been applied extensively in adult spinal deformity surgeries. For example, Ames et al. [32] used an ML model to predict functional outcome and quality of life after instrumental correction of adult spinal deformity. This model takes multiple factors into account, tailoring predictions to the unique circumstances of each patient, thereby improving preoperative counseling and decision-making. Building on this work, the same team later introduced a refined classification framework based on hierarchical clustering [153], which enhances decision-making by identifying data patterns. This system offers 2-year risk–benefit grids, guiding surgeons in selecting the most appropriate strategies, such as posterior fusion, intervertebral implantation, or pedicle subtraction osteotomy.

Model interpretability plays a crucial role in clinical adoption. Transparent models allow clinicians to

Table 4 Application of AI in personalized treatment planning

Algorithm	Application	References
ML model	Predicting functional outcome and quality of life after instrumental correction of adult spinal deformity	[32]
Cluster hierarchy	Predicting the outcome of postoperative patient self-assessment and providing an accurate risk-benefit analysis for various combinations of surgical techniques	[153]
CART model and cluster analysis	Predicting the risk of postoperative complications and unplanned readmission in surgically treated adult spinal deformity cases	[157]
Factorial analysis	Predicting the establishment of a specific gait model for THA patients	[155]
ML model	Predicting opioid use after THA	[33]
ML model	Predicting the ASES score after ARCR with relatively high accuracy	[36]
ML model	Predicting short-term outcomes after ORIF treatment of ankle fractures	[35]
ML model	Predicting the recurrence of PJI after TKA	[34]
Random forest ML analytical model	Predicting outcomes after irrigation and debridement for PJI	[158]
CNN	Analyzing movement during activities like walking, running, and sidestepping to detect ACL risks	[156]
ML model	Measuring the ACL injury risks	[159]
ML model	Predicting primary ACL injuries	[160]
DL model	Predicting the need for TKA based on knee radiographs	[49]
Classification tree algorithm	Assessing the risk of cervical spine injury in pediatric patients	[154]

ML machine learning, CART classification and regression tree, THA total hip arthroplasty, ASES American shoulder and elbow surgeons, ARCR arthroscopic rotator cuff repair, ORIF open reduction internal fixation, PJI periprosthetic joint infection, TKA total knee arthroplasty, CNN convolutional neural network, ACL anterior cruciate ligament, DL deep learning

understand key predictive factors, such as age, comorbidities, and surgical complexity, ensuring alignment with clinical reasoning. This fosters trust between patients and providers and mitigates potential biases, improving the generalizability of models across diverse populations. Interpretability also enhances personalized treatment planning, as demonstrated by Pellisé et al. [157]. The classification and regression tree algorithm was used to predict complications such as nonunion, vertebral kyphosis, and pedicle screw loosening, providing preoperative insights into the risk of implant failure and guiding individualized interventions. Similarly, Cattaneo et al. [155] applied factorial analysis to study gait changes post-THA, demonstrating how different surgical approaches, i.e., posterior vs. anterior, impact recovery trajectories. These findings emphasize that AI-powered predictive models align surgical strategies with individual patient needs, enhancing postoperative outcomes.

AI is likewise used for postsurgical evaluation and prediction of complications after other procedures, such as THA, arthroscopic rotator cuff repair (ARCR), ankle fracture open reduction internal fixation (ORIF), TKA, and flushing and debridement after periprosthetic joint infection (PJI). To predict opioid use after THA, Karhade et al. [33] designed a preoperative ML algorithm that analyzed a total of 5507 patients. The study revealed that 345 patients had long-term postoperative opioid use and identified predictors of long-term opioid prescription. The best model had an AUC of 0.77, indicating high net

returns. Potty et al. [36] proposed a new ML algorithm to predict the American Shoulder and Elbow Surgeons score after ARCR with relatively high accuracy. Merrill et al. [35] used ML to predict short-term outcomes after ORIF for ankle fractures. ML accurately predicted several comorbidities associated with poor short-term outcomes after ORIF for ankle fractures. Klemm et al. [34] developed and validated 3 ML models to predict the recurrence of PJI after TKA. The factors related to recurrence after TKA were irrigation and debridement, >4 previous open surgeries, metastatic disease, drug abuse, acquired immune deficiency syndrome, enterococcal infection, and obesity. All the ML models achieved good discrimination performance. Shohat et al. [158] designed a retrospective study that collected data on 1174 THA and TKA procedures. They used a randomized forest ML model in assessing 52 variables to predict outcomes after irrigation and debridement for PJI. The model achieved good discrimination (AUC=0.74) and high accuracy. In addition, the model identified 10 significant factors associated with failure, such as positive blood cultures and elevated C-reactive protein levels.

AI has gradually transformed sports medicine, especially in the area of ACL injury risk prediction. AI methods have facilitated the identification of biomechanical factors associated with ACL injury risk, supporting subsequent developments in imaging-based and sensor-driven predictive models [156, 159–161]. Pedoia et al. [161] used 3D MRI-based statistical shape modeling to

uncover anatomical features associated with ACL injury, such as specific intercondylar width and posterior tibial slope values. Johnson et al. [156] developed a DL-based system that uses a pretrained CNN to monitor knee joint movements in real time. Movement during activities such as walking, running, and sidestepping is analyzed to detect ACL injury risk, offering valuable insights for athletic training and injury prevention. Taborri et al. [159] quantified ACL injury risk by assessing stability and load absorption through inertial sensors and optoelectronic devices. Research has demonstrated that the landing error score system strongly correlates with the risk of ACL injury. Tamimi et al. [160] built a supervised ML-based predictive model using knee morphology data from MRI scans to predict primary ACL injury, achieving 92% testing accuracy and highlighting the role of AI in structural analysis and early intervention.

AI has also demonstrated value in predicting the progression of chronic diseases. In a recent study, Leung et al. [49] developed a DL prediction model for OA progression risk that could directly predict the need for TKA based on knee radiographs. Compared with the binary outcome model using a standard grading system, the proposed DL model better predicted the risk of TKA for OA. Bertsimas et al. [154] demonstrated the value of ML in improving clinical decision-making in pediatric trauma patients. They retrospectively assessed the risk of cervical spine injury in pediatric patients in conjunction with an optimal classification tree algorithm, achieving a sensitivity of 93.3% and a specificity of 82.3%. These models enable health care professionals to make informed, data-driven decisions, ensuring timely interventions and optimal patient outcomes.

By facilitating the early detection of potential risk factors, interpretable computational models contribute to the development of individualized preventive strategies and may assist in improving clinical decision-making. Within orthopedic practice, such models can aid in tailoring treatment plans, refining preoperative assessments, and supporting postoperative management. In areas such as complication risk assessment, rehabilitation planning, and injury prevention, data-driven approaches offer valuable support for predictive analyses. The capacity to process complex clinical data, recognize subtle associations, and provide interpretable outputs underscores the potential utility of these tools in the context of orthopedics and sports medicine.

Limitations and future perspectives

Progress in AI applications for musculoskeletal imaging and orthopedic care has been substantial; however, numerous limitations persist that affect the reliability and integration of AI-based methods into clinical practice.

Limitations in data and model design

The training of AI models heavily depends on large-scale, high-quality, and diverse datasets. However, in orthopedics, data scarcity remains a major issue [162]. Most current studies have relied on imaging data from single centers, specific equipment, and defined populations, which limits the representativeness and generalizability of the results [40, 56, 163, 164]. Additionally, the annotation process involves subjectivity and lacks uniform standards [41, 56]; for example, the use of radiology reports or manual segmentation as reference standards introduces potential bias into the model.

The inherent “black box” nature of AI systems further increases uncertainty in clinical applications [165]. Although deep models can achieve high predictive accuracy, they usually lack interpretability and fail to provide clinicians with understandable decision paths, especially in complex clinical scenarios. Although post hoc visualization methods can assist with partial interpretation, they do not fully replace traceable clinical logic [162]. Furthermore, AI models are vulnerable to adversarial perturbations [166], where small modifications to input images may result in incorrect predictions, posing risks to model safety and reliability.

Technical and operational challenges in clinical application

The clinical application of AI is currently limited by various technical and operational challenges. First, most AI systems are designed for specific tasks such as image classification or structure recognition, and are unable to address broader clinical workflows, including disease correlation and individualized treatment planning [66, 85]. Moreover, AI performs poorly in subjective assessments and cannot handle ambiguous or uncertain information [92], which further limits its utility in clinical scenarios. Second, AI systems are highly dependent on data and computational resources. In practice, many models require significant computing power and specific software-hardware configurations [93, 162], making deployment in resource-limited settings difficult. Model performance is also affected by image quality and the presence of anatomical landmarks, and variations in imaging protocols between hospitals add further complexity to deployment [93].

From a translational perspective, clinical validation remains insufficient. Many studies are based on data from single institutions, and model generalizability across different regions, equipment, and populations has not been adequately tested. For example, only 11% of fracture detection studies reported external geographical validation [167], significantly limiting the clinical applicability of the results. Adaptability to specific populations and complex cases also remains a critical concern.

Current AI systems have yet to receive FDA approval for pediatric fracture detection [85], as pediatric fractures often exhibit unique radiographic features that necessitate specialized detection algorithms [168]. Additionally, the performance of AI models in detecting stress fractures, pathological fractures, and abnormalities in bone structure still requires improvement [121]. Beyond these challenges, the lack of long-term follow-up data and limited validation of real-world clinical outcomes hinder the translation of AI-based methods from research settings to routine orthopedic practice. Short-term diagnostic accuracy does not necessarily equate to long-term therapeutic benefits, and few studies have established definitive links between AI-assisted interventions and patient prognoses over time [41, 125].

Therefore, to achieve effective translation and clinical adoption of AI-based methods in orthopedics, it is necessary to optimize deployment processes; enhance model functionality, strengthen multicenter, multimodal, large-scale clinical validation; and promote real-world research design to improve adaptability and reliability in complex clinical environments.

Acceptance, cost, and ethical considerations

AI still faces considerable barriers to clinical acceptance. One major issue is the lack of clinician trust in “black box” decision-making systems, especially in critical diagnostic scenarios where the reasoning path is not transparent [169, 170]. Additionally, health care providers, particularly less experienced interns or junior doctors, may become overly reliant on AI systems, which could impair the development of their independent clinical judgment [66].

Another significant challenge involves ethical and legal regulation [171]. At present, there is no unified standard for obtaining informed consent for the use of data in model training, and issues such as algorithmic fairness, justice, and data privacy lack systematic evaluation. As developers are not medical professionals, their legal responsibility remains undefined, and clearer regulatory guidance and stakeholder negotiation are urgently needed. Furthermore, AI models often inherit biases from training data, which may amplify systemic inequities related to race or gender. To date, most studies have not considered population diversity or ethnic differences [172].

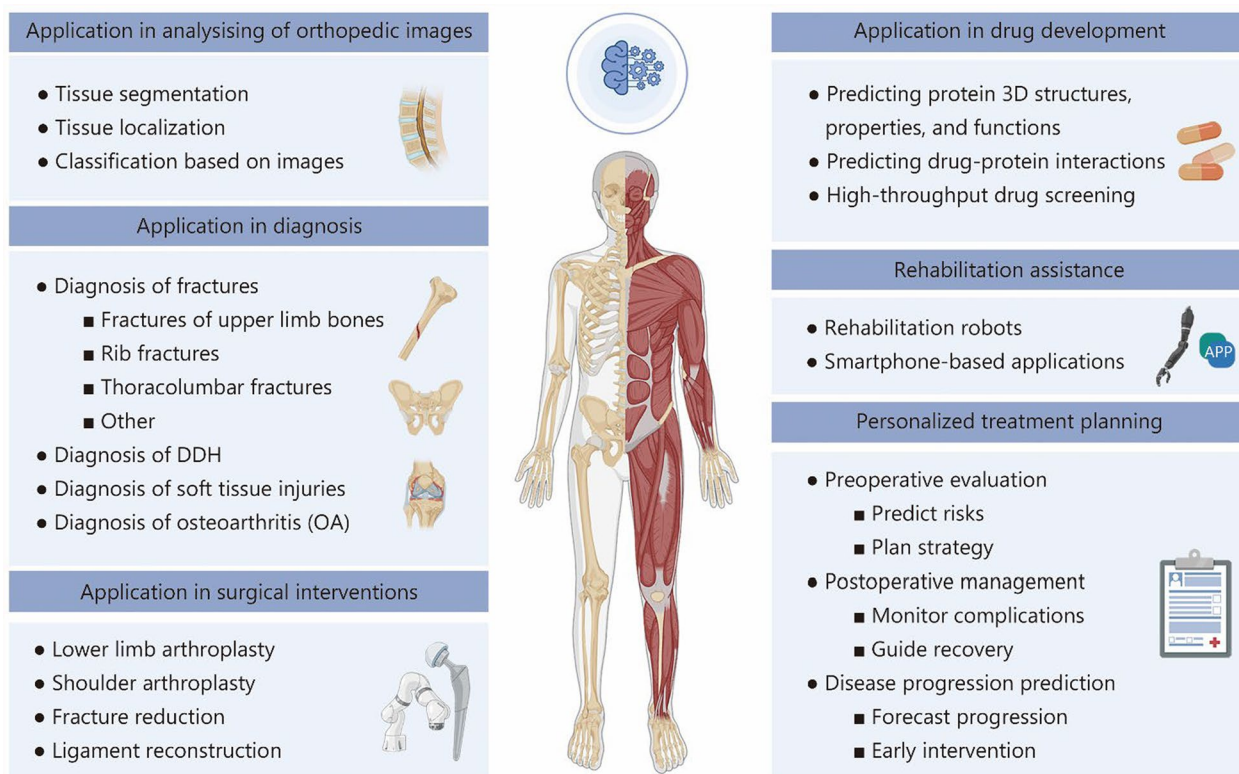


Fig. 3 Comprehensive applications of AI in orthopedics. DDH developmental dysplasia of the hip, 3D three-dimensional

Future perspectives

In summary, although AI methods are developing rapidly and extensively in orthopedics, their clinical application is still limited by several factors, including data quality and scale, model stability, the complexity of clinical environments, medical trust, and regulatory mechanisms. To advance the field, future research should focus on the following aspects: establishing high-quality, multicenter, ethnically diverse data platforms to promote standardized data sharing; developing more interpretable model architectures to increase clinical trust; expanding AI capabilities to cover the full diagnostic and treatment chain, including disease progression prediction, surgical planning, and long-term outcome evaluation; strengthening ethical review and regulatory frameworks to clarify the responsibilities of developers and users; and promoting long-term, multiregional, real-world clinical validation and iterative optimization. Only by addressing these challenges can AI in orthopedics truly evolve from a “laboratory technology” to a “clinical productivity tool”.

Conclusions

This review provides a comprehensive overview of the current applications of computational technologies in orthopedics (Fig. 3), with a particular focus on their roles in imaging analysis, clinical diagnosis, and various aspects of treatment, including surgical planning, drug development, rehabilitation, and personalized care. These tools have been employed to enhance the precision and efficiency of image interpretation, support clinical decision-making, and contribute to individualized therapeutic strategies. At the same time, this review also identifies a range of limitations that continue to hinder their broader implementation. These include constraints related to data availability and model generalizability, challenges in integrating such tools into complex clinical workflows, and concerns surrounding ethical oversight, regulatory standards, and long-term clinical validation. Future research should prioritize the development of interpretable and robust systems, the construction of diverse and high-quality datasets, and the establishment of multidisciplinary frameworks to ensure the responsible and effective incorporation of these technologies into orthopedic practice.

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
ACL	Anterior cruciate ligament
AI	Artificial intelligence
ARCR	Arthroscopic rotator cuff repair
AUC	Area under the curve
CASP	Critical assessment of protein structure prediction
CB513	CuIPDB 513
CBCT	Cone-beam computed tomography
CNN	Convolutional neural networks

CT	Computed tomography
DCNN	Deep convolutional neural network
DDH	Developmental dysplasia of the hip
DeepFRI	Deep functional residue identification
DL	Deep learning
DNN	Deep neural networks
HOG	Histogram of oriented gradient
IHDI grade	International Hip Dysplasia Institute grade
ML	Machine learning
MR	Magnetic resonance
MRI	Magnetic resonance imaging
MURA	Musculoskeletal radiograph
OA	Osteoarthritis
ORIF	Open reduction internal fixation
PHOG	Pyramidal histogram of oriented gradient
PJI	Periprosthetic joint infection
R&D	Research and development
SD	Standard deviation
SVM	Support vector machine
THA	Total hip arthroplasty
TKA	Total knee arthroplasty

Acknowledgements

The authors would like to acknowledge the professional language editing provided by Springer Nature Author Services. Additionally, the authors utilized ChatGPT (version GPT-4o, developed by OpenAI) to assist in improving the clarity and readability of the manuscript.

Authors' contributions

JS drew the figures and wrote the manuscript. GCW, SCW, and CRH helped modify the figures. XC revised the manuscript. JCS and YZZ conceived and supervised the study. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2024YFC2510602), the Key Program of the National Natural Science Foundation of China (82230071), the General Program of the National Natural Science Foundation of China (82172098, 82371603), and the Shanghai Hospital Development Center (SHDC2023CRT013).

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 13 September 2024 Accepted: 19 July 2025

Published online: 04 August 2025

References

- Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31–8.
- Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ*. 2016;353:i2139.
- Xiang Y, Zhao L, Liu Z, Wu X, Chen J, Long E, et al. Implementation of artificial intelligence in medicine: status analysis and development suggestions. *Artif Intell Med*. 2020;102:101780.

4. Malik AT, Khan SN. Predictive modeling in spine surgery. *Ann Transl Med.* 2019;7(Suppl 5):S173.
5. Wu E, Wu K, Daneshjouri R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med.* 2021;27(4):582–4.
6. Hui AT, Alvandi LM, Eleswarapu AS, Fornari ED. Artificial intelligence in modern orthopaedics: current and future applications. *JBJS Rev.* 2022. <https://doi.org/10.2106/JBJS.RVW.22.00086>.
7. Charles YP, Lamas V, Ntilikina Y. Artificial intelligence and treatment algorithms in spine surgery. *Orthop Traumatol Surg Res.* 2023;109(15):103456.
8. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *AJR Am J Roentgenol.* 2019;213(3):506–13.
9. Ho D, Quake SR, McCabe ERB, Chng WJ, Chow EK, Ding X, et al. Enabling technologies for personalized and precision medicine. *Trends Biotechnol.* 2020;38(5):497–518.
10. Patel AA, Schwab JH, Amanatullah DF, Divi SN. AOA critical issues symposium: shaping the impact of artificial intelligence within orthopaedic surgery. *J Bone Joint Surg Am.* 2023;105(18):1475–9.
11. Cheng K, Guo Q, He Y, Lu Y, Xie R, Li C, et al. Artificial intelligence in sports medicine: could GPT-4 make human doctors obsolete? *Ann Biomed Eng.* 2023;51(8):1658–62.
12. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotechnol.* 2018;6:75.
13. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med.* 2023;388(13):1201–8.
14. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. *Minim Invasive Ther Allied Technol.* 2019;28(2):73–81.
15. Li Z, Song P, Li G, Han Y, Ren X, Bai L, et al. AI energized hydrogel design, optimization and application in biomedicine. *Mater Today Bio.* 2024;25:101014.
16. Chang M, Canseco JA, Nicholson KJ, Patel N, Vaccaro AR. The role of machine learning in spine surgery: the future is now. *Front Surg.* 2020;7:54.
17. Chang TC, Seufert C, Eminaga O, Shkoliar E, Hu JC, Liao JC. Current trends in artificial intelligence application for endourology and robotic surgery. *Urol Clin North Am.* 2021;48(1):151–60.
18. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–44.
19. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. *Gastrointest Endosc.* 2020;92(4):807–12.
20. Zhang YP, Zhang XY, Cheng YT, Li B, Teng XZ, Zhang J, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. *Mil Med Res.* 2023;10(1):22.
21. Zeng S, Wang XL, Yang H. Radiomics and radiogenomics: extracting more information from medical images for the diagnosis and prognostic prediction of ovarian cancer. *Mil Med Res.* 2024;11(1):77.
22. Ramkumar PN, Luu BC, Haeberle HS, Karnuta JM, Nwachukwu BU, Williams RJ. Sports medicine and artificial intelligence: a primer. *Am J Sports Med.* 2022;50(4):1166–74.
23. Avanzo M, Stancanello J, Pirrone G, Drigo A, Retico A. The evolution of artificial intelligence in medical imaging: from computer science to machine and deep learning. *Cancers (Basel).* 2024;16(21):3702.
24. Bai L, Wu Y, Li G, Zhang W, Zhang H, Su J. AI-enabled organoids: construction, analysis, and application. *Bioact Mater.* 2024;31:525–48.
25. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial intelligence and orthopaedics: an introduction for clinicians. *J Bone Joint Surg Am.* 2020;102(9):830–40.
26. Deo RC. Machine learning in medicine. *Circulation.* 2015;132(20):1920–30.
27. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. *JOR Spine.* 2019;2(1):e1044.
28. Chafai N, Luigi B, Sara B, Badaoui B. Emerging applications of machine learning in genomic medicine and healthcare. *Crit Rev Clin Lab Sci.* 2024;61(2):140–63.
29. Hassan M, Awan FM, Naz A, deAndrés-Galiana EJ, Alvarez O, Cernea A, et al. Innovations in genomics and big data analytics for personalized medicine and health care: a review. *Int J Mol Sci.* 2022;23(9):4645.
30. Kotti M, Duffell LD, Faisal AA, McGregor AH. Detecting knee osteoarthritis and its discriminating parameters using random forests. *Med Eng Phys.* 2017;43:19–29.
31. Luu BC, Wright AL, Haeberle HS, Karnuta JM, Schickendantz MS, Makhni EC, et al. Machine learning outperforms logistic regression analysis to predict next-season NHL player injury: an analysis of 2322 players from 2007 to 2017. *Orthop J Sports Med.* 2020;8(9):2325967120953404.
32. Ames CP, Smith JS, Pellisé F, Kelly M, Gum JL, Alanay A, et al. Development of predictive models for all individual questions of SRS-22R after adult spinal deformity surgery: a step toward individualized medicine. *Eur Spine J.* 2019;28(9):1998–2011.
33. Karhade AV, Schwab JH, Bedair HS. Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty. *J Arthroplast.* 2019;34(10):2272–7.e1.
34. Klemt C, Laurencin S, Uzosike AC, Burns JC, Costales TG, Yeo I, et al. Machine learning models accurately predict recurrent infection following revision total knee arthroplasty for periprosthetic joint infection. *Knee Surg Sports Traumatol Arthrosc.* 2022;30(8):2582–90.
35. Merrill RK, Ferrandino RM, Hoffman R, Shaffer GW, Ndu A. Machine learning accurately predicts short-term outcomes following open reduction and internal fixation of ankle fractures. *J Foot Ankle Surg.* 2019;58(3):410–6.
36. Potty AG, Potty ASR, Maffulli N, Blumenschein LA, Ganta D, Mistovich RJ, et al. Approaching artificial intelligence in orthopaedics: predictive analytics and machine learning to prognosticate arthroscopic rotator cuff surgical outcomes. *J Clin Med.* 2023;12(6):2369.
37. Martin RK, Wastvedt S, Pareek A, Persson A, Visnes H, Fenstad AM, et al. Unsupervised machine learning of the combined Danish and Norwegian knee ligament registers: identification of 5 distinct patient groups with differing ACL revision rates. *Am J Sports Med.* 2024;52(4):881–91.
38. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* 2021;13(1):152.
39. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* 2019;11(1):70.
40. Hemke R, Buckless CG, Tsao A, Wang B, Torriani M. Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment. *Skeletal Radiol.* 2020;49(3):387–95.
41. Norman B, Padoia 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(1):177–85.
42. Chung SW, Han SS, Lee JW, Oh K-S, Kim NR, Yoon JP, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop.* 2018;89(4):468–73.
43. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol.* 2018;73(5):439–45.
44. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci U S A.* 2018;115(45):11591–6.
45. Borjali A, Chen AF, Muratoglu OK, Morid MA, Varadarajan KM. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. *J Orthop Res.* 2020;38(7):1465–71.
46. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–9.
47. Wang S, Peng J, Ma J, Xu J. Protein secondary structure prediction using deep convolutional neural fields. *Sci Rep.* 2016;6(1):18962.
48. Wang Z, Combs SA, Brand R, Calvo MR, Xu P, Price G, et al. LM-GVP: an extensible sequence and structure informed deep learning framework for protein property prediction. *Sci Rep.* 2022;12(1):6832.
49. 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(3):584–93.
50. Burström G, Buerger C, Hoppenbrouwers J, Nachabe R, Lorenz C, Babic D, et al. Machine learning for automated 3-dimensional segmentation of the spine and suggested placement of pedicle screws based on

- intraoperative cone-beam computer tomography. *J Neurosurg Spine*. 2019;31(1):147–54.
51. Chae J, Kang Y-J, Noh Y. A deep-learning approach for foot-type classification using heterogeneous pressure data. *Sensors*. 2020;20(16):4481.
 52. Fan G, Liu H, Wu Z, Li Y, Feng C, Wang D, et al. Deep learning-based automatic segmentation of lumbosacral nerves on CT for spinal intervention: a translational study. *AJNR Am J Neuroradiol*. 2019;40(6):1074–81.
 53. Ghidotti A, Vitali A, Regazzoni D, Cohen MW, Rizzi C. Comparative analysis of convolutional neural network architectures for automated knee segmentation in medical imaging: a performance evaluation. *J Comput Inf Sci Eng*. 2024;24(5):051005.
 54. Glocker B, Feulner J, Criminisi A, Haynor DR, Konukoglu E, editors. Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. *Med Image Comput Comput Assist Interv*. 2012;15(Pt 3):590–8.
 55. Jamaludin A, Loots M, Kadir T, Zisserman A, Urban J, Battié MC, 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(5):1374–83.
 56. Lind A, Akbarian E, Olsson S, Näsell H, Sköldenberg O, Razavian AS, et al. Artificial intelligence for the classification of fractures around the knee in adults according to the 2018 AO/OTA classification system. *PLoS ONE*. 2021;16(4):e0248809.
 57. Oktay AB, Akgul YS. Simultaneous localization of lumbar vertebrae and intervertebral discs with SVM-based MRF. *IEEE Trans Biomed Eng*. 2013;60(9):2375–83.
 58. Shah RF, Martinez AM, Padoia V, Majumdar S, Vail TP, Bini SA. Variation in the thickness of knee cartilage. The use of a novel machine learning algorithm for cartilage segmentation of magnetic resonance images. *J Arthroplast*. 2019;34(10):2210–5.
 59. Jakubicek R, Chmelik J, Jan J, Ourednicek P, Lambert L, Gavelli G. Learning-based vertebra localization and labeling in 3D CT data of possibly incomplete and pathological spines. *Comput Methods Programs Biomed*. 2020;183:105081.
 60. Kwan JL, Calder LA, Bowman CL, MacIntyre A, Mimeault R, Honey L, et al. Characteristics and contributing factors of diagnostic error in surgery: analysis of closed medico-legal cases and complaints in Canada. *Can J Surg*. 2024;67(1):E58.
 61. Federico CA, Trotsyuk AA. Biomedical data science, artificial intelligence, and ethics: navigating challenges in the face of explosive growth. *Annu Rev Biomed Data Sci*. 2024;7(1):1–14.
 62. Rabie AH, Saleh AI. Diseases diagnosis based on artificial intelligence and ensemble classification. *Artif Intell Med*. 2024;148:102753.
 63. Guermazi A, Tannoury C, Kompel AJ, Murakami AM, Ducarouge A, Gillibert A, et al. Improving radiographic fracture recognition performance and efficiency using artificial intelligence. *Radiology*. 2022;302(3):627–36.
 64. Link TM, Padoia V. Using AI to improve radiographic fracture detection. *Radiology*. 2022;302(3):637–8.
 65. Wu AM, Bisignano C, James SL, Abady GG, Abedi A, Abu-Gharbieh E, et al. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet Healthy Longev*. 2021;2(9):e580–92.
 66. Chen K, Stotter C, Klestil T, Nehrer S. Artificial intelligence in orthopedic radiography analysis: a narrative review. *Diagnostics*. 2022;12(9):2235.
 67. Ajmera P, Kharat A, Botchu R, Gupta H, Kulkarni V. Real-world analysis of artificial intelligence in musculoskeletal trauma. *J Clin Orthop Trauma*. 2021;22:101573.
 68. Shen L, Gao C, Hu S, Kang D, Zhang Z, Xia D, et al. Using artificial intelligence to diagnose osteoporotic vertebral fractures on plain radiographs. *J Bone Miner Res*. 2023;38(9):1278–87.
 69. Beyaz S, Açıkcı K, Sümer E. Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches. *Jt Dis Relat Surg*. 2020;31(2):175–83.
 70. Burns JE, Yao J, Summers RM. Vertebral body compression fractures and bone density: automated detection and classification on CT images. *Radiology*. 2017;284(3):788–97.
 71. Guan B, Zhang G, Yao J, Wang X, Wang M. Arm fracture detection in X-rays based on improved deep convolutional neural network. *Comput Electr Eng*. 2020;81:106530.
 72. Jones RM, Sharma A, Hotchkiss R, Sperling JW, Hamburger J, Ledig C, et al. Assessment of a deep-learning system for fracture detection in musculoskeletal radiographs. *NPJ Digit Med*. 2020;3:144.
 73. Langerhuizen DWG, Bulstra AEJ, Janssen SJ, Ring D, Kerckhoffs GMMJ, Jaarsma RL, et al. Is deep learning on par with human observers for detection of radiographically visible and occult fractures of the scaphoid? *Clin Orthop Relat Res*. 2020;478(11):2653–9.
 74. Li YC, Chen HH, Horng-Shing LuH, Hondar Wu HT, Chang MC, Chou PH. Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists? *Clin Orthop Relat Res*. 2021;479(7):1598–612.
 75. Niya A, Murakami K, Kobayashi R, Sekimoto A, Saeki M, Toyofuku K, et al. Development of an artificial intelligence-assisted computed tomography diagnosis technology for rib fracture and evaluation of its clinical usefulness. *Sci Rep*. 2022;12(1):8363.
 76. Pranata YD, Wang KC, Wang JC, Idram I, Lai JY, Liu JW, et al. Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. *Comput Methods Programs Biomed*. 2019;171:27–37.
 77. Sato Y, Takegami Y, Asamoto T, Ono Y, Hidetoshi T, Goto R, et al. Artificial intelligence improves the accuracy of residents in the diagnosis of hip fractures: a multicenter study. *BMC Musculoskelet Disord*. 2021;22(1):407.
 78. Yao L, Guan X, Song X, Tan Y, Wang C, Jin C, et al. Rib fracture detection system based on deep learning. *Sci Rep*. 2021;11(1):23513.
 79. Wu J, Liu N, Li X, Fan Q, Li Z, Shang J, et al. Convolutional neural network for detecting rib fractures on chest radiographs: a feasibility study. *BMC Med Imaging*. 2023;23(1):18.
 80. Briody H, Hanneman K, Patlas MN. Applications of artificial intelligence in acute thoracic imaging. *Can Assoc Radiol J*. 2025. <https://doi.org/10.1177/08465371251322705>.
 81. Rosenberg GS, Cina A, Schiró GR, Giorgi PD, Gueorguiev B, Alini M, et al. Artificial intelligence accurately detects traumatic thoracolumbar fractures on sagittal radiographs. *Medicina (Kaunas)*. 2022;58(8):998.
 82. Murata K, Endo K, Aihara T, Suzuki H, Sawaji Y, Matsuoka Y, et al. Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. *Sci Rep*. 2020;10(1):20031.
 83. Weng YS, Wang LJ, Huang JQ, Cai LJ. Factors associated with new fractures in adjacent vertebrae after percutaneous vertebroplasty for osteoporotic vertebral compression fractures. *Am J Transl Res*. 2024;16(11):6972–9.
 84. Lorentzon M, Litsne H, Axelsson KF. The significance of recent fracture location for imminent risk of hip and vertebral fractures—a nationwide cohort study on older adults in Sweden. *Osteoporos Int*. 2024;35(6):1077–87.
 85. Zech JR, Santomartino SM, Yi PH. Artificial intelligence (AI) for fracture diagnosis: an overview of current products and considerations for clinical adoption, from the AJR special series on AI applications. *AJR Am J Roentgenol*. 2022;219(6):869–78.
 86. Oakden-Rayner L, Gale W, Bonham TA, Lungren MP, Carneiro G, Bradley AP, et al. Validation and algorithmic audit of a deep learning system for the detection of proximal femoral fractures in patients in the emergency department: a diagnostic accuracy study. *Lancet Digit Health*. 2022;4(5):e351–8.
 87. Tadavarthi Y, Vey B, Krupinski E, Prater A, Gichoya J, Safdar N, et al. The state of radiology AI: considerations for purchase decisions and current market offerings. *Radiol Artif Intell*. 2020;2(6):e200004.
 88. Chen M, Cai R, Zhang A, Chi X, Qian J. The diagnostic value of artificial intelligence-assisted imaging for developmental dysplasia of the hip: a systematic review and meta-analysis. *J Orthop Surg Res*. 2024;19(1):522.
 89. Bradley CS, Verma Y, Maddock CL, Wedge JH, Gargan MF, Kelley SP. A comprehensive nonoperative treatment protocol for developmental dysplasia of the hip in infants. *Bone Joint J*. 2023;105-B(8):935–42.
 90. Liu C, Xie H, Zhang S, Mao Z, Sun J, Zhang Y. Misshapen pelvis landmark detection with local-global feature learning for diagnosing developmental dysplasia of the hip. *IEEE Trans Med Imaging*. 2020;39(12):3944–54.

91. Huang B, Xia B, Qian J, Zhou X, Zhou X, Liu S, et al. Artificial intelligence-assisted ultrasound diagnosis on infant developmental dysplasia of the hip under constrained computational resources. *J Ultrasound Med.* 2023;42(6):1235–48.
92. Xu W, Shu L, Gong P, Huang C, Xu J, Zhao J, et al. A deep-learning aided diagnostic system in assessing developmental dysplasia of the hip on pediatric pelvic radiographs. *Front Pediatr.* 2022;9:785480.
93. Jaremo JL, Hareendranathan A, Bolouri SES, Frey RF, Dulai S, Bailey AL. AI aided workflow for hip dysplasia screening using ultrasound in primary care clinics. *Sci Rep.* 2023;13(1):9224.
94. Ghasseminia S, Lim AKS, Concepcion NDP, Kirschner D, Teo YM, Dulai S, et al. Interobserver variability of hip dysplasia indices on sweep ultrasound for novices, experts, and artificial intelligence. *J Pediatr Orthop.* 2022;42(4):e315–23.
95. Libon J, Ng C, Bailey A, Hareendranathan A, Joseph R, Dulai S. Remote diagnostic imaging using artificial intelligence for diagnosing hip dysplasia in infants: results from a mixed-methods feasibility pilot study. *Paediatr Child Health.* 2023;28(5):285–90.
96. Kumm J, Roemer FW, Guermazi A, Turkiewicz A, Englund M. Natural history of intrameniscal signal intensity on knee MR images: six years of data from the osteoarthritis initiative. *Radiology.* 2016;278(1):164–71.
97. Khan M, Evaniew N, Bedi A, Ayeni OR, Bhandari M. Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. *CMAJ.* 2014;186(14):1057–64.
98. Bien N, Rajpurkar P, Ball RL, Irvin J, Park A, Jones E, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. *PLoS Med.* 2018;15(11):e1002699.
99. Fritz B, Marbach G, Civardi F, Fucetese SF, Pfirrmann CWA. Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference. *Skeletal Radiol.* 2020;49(8):1207–17.
100. 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(2):400–10.
101. Sokal PA, Norris R, Maddox TW, Oldershaw RA. The diagnostic accuracy of clinical tests for anterior cruciate ligament tears are comparable but the Lachman test has been previously overestimated: a systematic review and meta-analysis. *Knee Surg Sports Traumatol Arthrosc.* 2022;30(10):1795.
102. Nyland J. The ACL Café Menu: individualised treatment of anterior cruciate ligament injuries—from prevention to conservative treatment, repair and reconstruction. *Knee Surg Sports Traumatol Arthrosc.* 2025.
103. Štajduhar I, Mamula M, Miličić D, Ūnal G. Semi-automated detection of anterior cruciate ligament injury from MRI. *Comput Methods Programs Biomed.* 2017;140:151–64.
104. Richardson ML. MR protocol optimization with deep learning: a proof of concept. *Curr Probl Diagn Radiol.* 2021;50(2):168–74.
105. Tran A, Lassalle L, Zille P, Guillin R, Pluot E, Adam C, et al. Deep learning to detect anterior cruciate ligament tear on knee MRI: multi-continental external validation. *Eur Radiol.* 2022;32(12):8394–403.
106. Liu F, Guan B, Zhou Z, Samsonov A, Rosas H, Lian K, et al. Fully automated diagnosis of anterior cruciate ligament tears on knee MR images by using deep learning. *Radiol Artif Intell.* 2019;1(3):180091.
107. Zeng W, Ismail SA, Pappas E. Detecting the presence of anterior cruciate ligament injury based on gait dynamics disparity and neural networks. *Artif Intell Rev.* 2020;53(5):3153–76.
108. Li X, Huang H, Wang J, Yu Y, Ao Y. The analysis of plantar pressure data based on multimodel method in patients with anterior cruciate ligament deficiency during walking. *Biomed Res Int.* 2016;2016(1):7891407.
109. Wu R, Guo Y, Chen Y, Zhang J. Osteoarthritis burden and inequality from 1990 to 2021: a systematic analysis for the global burden of disease study 2021. *Sci Rep.* 2025;15(1):8305.
110. Conrozier T, Brandt K, Piperno M, Mathieu P, Merle-Vincent F, Vignon E. Reproducibility and sensitivity to change of a new method of computer measurement of joint space width in hip osteoarthritis. Performance of three radiographic views obtained at a 3-year interval. *Osteoarthr Cartil.* 2009;17(7):864–70.
111. Mulford KL, Kaji ES, Grove AF, Saniei S, Girod-Hoffman M, Maradit-Kremers H, et al. A deep learning tool for minimum joint space width calculation on antero-posterior knee radiographs. *J Arthroplast.* 2025. <https://doi.org/10.1016/j.jarth.2025.01.038>.
112. 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(9):1369–74.
113. Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Sci Rep.* 2018;8(1):1727.
114. Norman B, Pedoia 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(3):471–7.
115. Pedoia V, Lee J, Norman B, Link TM, Majumdar S. Diagnosing osteoarthritis from T2 maps using deep learning: an analysis of the entire osteoarthritis initiative baseline cohort. *Osteoarthr Cartil.* 2019;27(7):1002–10.
116. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg.* 2018;268(1):70–6.
117. Hughes TM, Dossett LA, Hawley ST, Telem DA. Recognizing heuristics and bias in clinical decision-making. *Ann Surg.* 2020;271(5):813–4.
118. Jacofsky DJ, Allen M. Robotics in arthroplasty: a comprehensive review. *J Arthroplast.* 2016;31(10):2353–63.
119. McDonnell JM, Ahern DP, Doinn TÓ, Gibbons D, Rodrigues KN, Birch N, et al. Surgeon proficiency in robot-assisted spine surgery. *Bone Joint J.* 2020;102-B(5):568–72.
120. Murphy MP, Brown NM. CORR synthesis: when should the orthopaedic surgeon use artificial intelligence, machine learning, and deep learning?. *Clin Orthop Relat Res.* 2021;479(7):1497–505.
121. Chen X, Li S, Liu X, Wang Y, Ma R, Zhang Y, et al. Acetabular diameter assessment and three-dimensional simulation for acetabular reconstruction in dysplastic hips. *J Arthroplast.* 2023;38(8):1551–8.
122. Darwood A, Hurst SA, Villatte G, Tatti F, El Daou H, Reilly P, et al. Novel robotic technology for the rapid intraoperative manufacture of patient-specific instrumentation allowing for improved glenoid component accuracy in shoulder arthroplasty: a cadaveric study. *J Shoulder Elbow Surg.* 2022;31(3):561–70.
123. Gebremeskel M, Shafiq B, Uneri A, Sheth N, Simmerer C, Zbijewski W, et al. Quantification of manipulation forces needed for robot-assisted reduction of the ankle syndesmosis: an initial cadaveric study. *Int J Comput Assist Radiol Surg.* 2022;17(12):2263–7.
124. Giorgini A, Tarallo L, Novi M, Porcellini G. Computer-assisted surgery in reverse shoulder arthroplasty: early experience. *Indian J Orthop.* 2021;55(4):1003–8.
125. Kayani B, Konan S, Tahmassebi J, Pietrzak JRT, Haddad FS. Robotic-arm assisted total knee arthroplasty is associated with improved early functional recovery and reduced time to hospital discharge compared with conventional jig-based total knee arthroplasty. *Bone Joint J.* 2018;100(B(7)):930–7.
126. Rossi SMP, Sangaletti R, Andriollo L, Matascioli L, Benazzo F. The use of a modern robotic system for the treatment of severe knee deformities. *Technol Health Care.* 2024;32:3737–46.
127. Sakakibara Y, Teramoto A, Takagi T, Yamakawa S, Shoji H, Okada Y, et al. Effects of the ankle flexion angle during anterior talofibular ligament reconstruction on ankle kinematics, laxity, and in situ forces of the reconstructed graft. *Foot Ankle Int.* 2022;43(5):725–32.
128. Yang G, Liu D, Zhou G, Wang Q, Zhang X. Robot-assisted anterior cruciate ligament reconstruction based on three-dimensional images. *J Orthop Surg Res.* 2024;19(1):246.
129. Twomey-Kozak J, Hurley E, Levin J, Anakwenze O, Klifto C. Technological innovations in shoulder replacement: current concepts and the future of robotics in total shoulder arthroplasty. *J Shoulder Elbow Surg.* 2023;32(10):2161–71.
130. Herzog MM, Kerr ZY, Marshall SW, Wikstrom EA. Epidemiology of ankle sprains and chronic ankle instability. *J Athl Train.* 2019;54(6):603–10.
131. Peiffer M, Lewis L, Xie K, Guild TT, Ashkani-Esfahani S, Kwon JY. The influence of talar displacement on articular contact mechanics: a 3D finite element analysis study using weightbearing computed tomography. *Foot Ankle Int.* 2024;45(4):393–405.

132. Gregersen MG, Dalen AF, Skrede AL, Bjelland Ø, Nilsen FA, Molund M. Effects of fibular plate fixation on ankle stability in a Weber B fracture model with partial deltoid ligament sectioning. *Foot Ankle Int.* 2024;45(6):641–7.
133. Spindler FT, Gaube FP, Böcker W, Polzer H, Baumbach SF. Value of intraoperative 3D imaging on the quality of reduction of the distal tibiofibular joint when using a suture-button system. *Foot Ankle Int.* 2022;44(1):54–61.
134. Bajorath J, Kearnes S, Walters WP, Meanwell NA, Georg GI, Wang S. Artificial intelligence in drug discovery: into the great wide open. *J Med Chem.* 2020;63(16):8651–2.
135. Smalley E. AI-powered drug discovery captures pharma interest. *Nat Biotechnol.* 2017;35(7):604–5.
136. Brown FK, Sherer EC, Johnson SA, Holloway MK, Sherborne BS. The evolution of drug design at Merck Research Laboratories. *J Comput Aided Mol Des.* 2017;31(3):255–66.
137. Gangwal A, Lavecchia A. Unleashing the power of generative AI in drug discovery. *Drug Discov Today.* 2024;29(6):103992.
138. Lowe D. AI designs organic syntheses. *Nature.* 2018;555(7698):592–3.
139. Schneider G. Automating drug discovery. *Nat Rev Drug Discov.* 2018;17(2):97–113.
140. Gligorijević V, Renfrew PD, Kosciolok T, Leman JK, Berenberg D, Vatanen T, et al. Structure-based protein function prediction using graph convolutional networks. *Nat Commun.* 2021;12(1): 3168.
141. Spencer M, Eickholt J, Cheng J. A deep learning network approach to ab initio protein secondary structure prediction. *IEEE/ACM Trans Comput Biol Bioinform.* 2015;12(1):103–12.
142. Yuan L, Ma Y, Liu Y. Ensemble deep learning models for protein secondary structure prediction using bidirectional temporal convolution and bidirectional long short-term memory. *Front Bioeng Biotechnol.* 2023;11:1051268.
143. Offensperger F, Tin G, Duran-Frigola M, Hahn E, Dobner S, Ende CWa, et al. Large-scale chemoproteomics expedites ligand discovery and predicts ligand behavior in cells. *Science.* 2024;384(6694):eadk5864.
144. Wang F, Liu D, Wang H, Luo C, Zheng M, Liu H, et al. Computational screening for active compounds targeting protein sequences: methodology and experimental validation. *J Chem Inf Model.* 2011;51(11):2821–8.
145. Ren F, Aliper A, Chen J, Zhao H, Rao S, Kuppe C, et al. A small-molecule TNK inhibitor targets fibrosis in preclinical and clinical models. *Nat Biotechnol.* 2025;43(1):63–75.
146. Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of artificial intelligence for computer-assisted drug discovery. *Chem Rev.* 2019;119(18):10520–94.
147. Averta G, Della Santina C, Valenza G, Bicchi A, Bianchi M. Exploiting upper-limb functional principal components for human-like motion generation of anthropomorphic robots. *J Neuroeng Rehabil.* 2020;17(1):63.
148. Zhao Y, Liang C, Gu Z, Zheng Y, Wu Q. A new design scheme for intelligent upper limb rehabilitation training robot. *Int J Environ Res Public Health.* 2020;17(8):2948.
149. Miller-Jackson TM, Natividad RF, Lim DYL, Hernandez-Barraza L, Ambrose JW, Yeow RC-H. A wearable soft robotic exoskeleton for hip flexion rehabilitation. *Front Robot AI.* 2022;9:835237.
150. Rossi SMP, Panzera RM, Sangaletti R, Andriollo L, Giudice L, Lecci F, et al. Problems and opportunities of a smartphone-based care management platform: application of the Wald principles to a survey-based analysis of patients' perception in a pilot center. *Healthcare.* 2024;12(2):153.
151. Chen ZH, Lin L, Wu CF, Li CF, Xu RH, Sun Y. Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. *Cancer Commun (Lond).* 2021;41(11):1100–15.
152. Sánchez de la Nava AM, Atienza F, Bermejo J, Fernández-Avilés F. Artificial intelligence for a personalized diagnosis and treatment of atrial fibrillation. *Am J Physiol Heart Circ Physiol.* 2021;320(4):H1337–47.
153. Ames CP, Smith JS, Pellisé F, Kelly M, Alanay A, Acaroğlu E, et al. Artificial intelligence based hierarchical clustering of patient types and intervention categories in adult spinal deformity surgery: towards a new classification scheme that predicts quality and value. *Spine.* 2019;44(13):915–26.
154. Bertsimas D, Masiakos PT, Mylonas KS, Wiberg H. Prediction of cervical spine injury in young pediatric patients: an optimal trees artificial intelligence approach. *J Pediatr Surg.* 2019;54(11):2353–7.
155. Cattaneo A, Ghidotti A, Catellani F, Fiorentino G, Vitali A, Regazzoni D, et al. Motion acquisition of gait characteristics one week after total hip arthroplasty: a factor analysis. *Arch Orthop Trauma Surg.* 2024;144(5):2347–56.
156. Johnson WR, Mian A, Lloyd DG, Alderson JA. On-field player workload exposure and knee injury risk monitoring via deep learning. *J Biomech.* 2019;93:185–93.
157. Pellisé F, Serra-Burriel M, Smith JS, Haddad S, Kelly MP, Vila-Casademunt A, et al. Development and validation of risk stratification models for adult spinal deformity surgery. *J Neurosurg Spine.* 2019;31(4):587–99.
158. Shohat N, Goswami K, Tan TL, Yayac M, Soriano A, Sousa R, et al. 2020 frank stinchfield award: identifying who will fail following irrigation and debridement for prosthetic joint infection. *Bone Joint J.* 2020;102-B((7 Supple B)):11–9.
159. Taborri J, Molinaro L, Santospagnuolo A, Vetrano M, Vulpiani MC, Rossi S. A machine-learning approach to measure the anterior cruciate ligament injury risk in female basketball players. *Sensors (Basel).* 2021;21(9):3141.
160. Tamimi I, Ballesteros J, Lara AP, Tat J, Alaquiel M, Schupbach J, et al. A prediction model for primary anterior cruciate ligament injury using artificial intelligence. *Orthop J Sports Med.* 2021;2021(9):23259671211027544.
161. Pedoia V, Lansdown DA, Zaid M, McCulloch CE, Souza R, Ma CB, et al. Three-dimensional MRI-based statistical shape model and application to a cohort of knees with acute ACL injury. *Osteoarthritis Cartil.* 2015;23(10):1695–703.
162. Alzubaidi L, Al-Dulaimi K, Salhi A, Alammam Z, Fadhel MA, Albahri AS, et al. Comprehensive review of deep learning in orthopaedics: applications, challenges, trustworthiness, and fusion. *Artif Intell Med.* 2024;155:102935.
163. Himeur Y, Al-Maadeed S, Kheddar H, Al-Maadeed N, Abualsaud K, Mohamed A, et al. Video surveillance using deep transfer learning and deep domain adaptation: towards better generalization. *Eng Appl Artif Intell.* 2023;119:105698.
164. Jakubovitz D, Giryres R, Rodrigues MRD. Generalization error in deep learning. In: Boche H, Caire G, Calderbank R, Kutyniok G, Mathar R, Petersen P, editors. *Compressed sensing and its applications: third international MATHEON conference 2017.* Cham: Springer; 2019. p. 153–93.
165. Roy S, Pal D, Meena T. Explainable artificial intelligence to increase transparency for revolutionizing healthcare ecosystem and the road ahead. *Netw Model Anal Health Inform Bioinforma.* 2023;13(1):4.
166. Arnab A, Miksik O, Torr PHS. On the robustness of semantic segmentation models to adversarial attacks. *IEEE Trans Pattern Anal Mach Intell.* 2020;42(12):3040–53.
167. Oliveira e Carmo L, van den Merkhof A, Olczak J, Gordon M, Jutte PC, Jaarsma RL, et al. An increasing number of convolutional neural networks for fracture recognition and classification in orthopaedics. *Bone Jt Open.* 2021;2(10):879–85.
168. George MP, Bixby S. Frequently missed fractures in pediatric trauma: a pictorial review of plain film radiography. *Radiol Clin North Am.* 2019;57(4):843–55.
169. Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR data science institute artificial intelligence survey. *J Am Coll Radiol.* 2021;18(8):1153–9.
170. Roy S, Meena T, Lim SJ. Demystifying supervised learning in healthcare 4.0: a new reality of transforming diagnostic medicine. *Diagnostics (Basel).* 2022;12(10):2549.
171. 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.
172. Dallora AL, Anderberg P, Kvist O, Mendes E, Diaz Ruiz S, Sanmartin BJ. Bone age assessment with various machine learning techniques: a systematic literature review and meta-analysis. *PLoS One.* 2019;14(7):e0220242.