Check for updates

Application of artificial intelligence technology in the field of orthopedics: a narrative review

Pengran Liu[1] · Jiayao Zhang[1] · Songxiang Liu[1] · Tongtong Huo[1,3] · Jiajun He[3] · Mingdi Xue[1] · Ying Fang[1] · Honglin Wang[1] · Yi Xie[1] · Mao Xie[1] · Dan Zhang[2] · Zhewei Ye[1]

Abstract

Artificial intelligence (AI) was a new interdiscipline of computer technology, mathematic, cybernetics and determinism. These years, AI had obtained a significant development by the improvement of core technology Machine Learning and Deep Learning. With the assistance of AI, profound changes had been brought into the traditional orthopedics. In this paper, we narratively reviewed the latest applications of AI in orthopedic diseases, including the severity evaluation, triage, diagnosis, treatment and rehabilitation. The research point, relevant advantages and disadvantages of the orthopedic AI was also discussed combined with our own research experiences. We aimed to summarize the past achievements and appeal for more attentions and effective applications of AI in the field of orthopedics.

Pengran Liu, Jiayao Zhang and Songxiang Liu have contributed equally to this paper.

✉ Mao Xie
  xiemao886@163.com

✉ Dan Zhang
  zhangdan_1@foxmail.com

✉ Zhewei Ye
  yezhewei@hust.edu.cn

1 Department of Orthopedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

2 Department of Pharmacy, Traditional Chinese and Western Medicine Hospital of Wuhan (Wuhan No.1 Hospital), Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China

3 School of Artificial Intelligence and Automation, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

1 Introduction

With the development of science and technology, a series of advanced technologies had been applied in the field of traditional medicine, which marked the arrival of the era of intelligent medicine. Artificial intelligence (AI) was one of the most representative technologies, whose appearance had brought great convenience to current clinical works (Zhewei 2020). As an interdiscipline of computer technology, mathematic, cybernetics and determinism, the aim of AI was to study, even surpass, the human intelligence based on intelligent computer algorithm (Myers et al. 2020). AI was originally put forward by the Alan Turing in 1950, however, limited by the poor computer hardware and calculation power (Dutt et al. 2020; Liu et al. 2021a, b). Endured the cold winter, Machine Learning (ML) and Deep Learning (DL) appeared and brought AI with huge developments and further industrial promotions (Kaul et al. 2020). Among the numerous algorithms for AI realizing, ML was one of the best developed branches. ML was a method of learning and analyzing data through computer programs based on statistics and mathematical models to automatically discover regularities and patterns in data and make predictions and decisions. ML, generally consisted of linear model, decision tree, Bayes classifier, random forest, support vector machine (SVM) model, was relatively simple and suitable for some relatively simple application scenarios. DL was a branch of ML basing on neural networks. DL could automatically learn features and patterns from raw data and use them to classify or predict the data. Different from traditional ML, DL was able to learn features with multiple layers of abstraction, which allowed it to work with more complex and high-dimensional data. The most important element of DL was the neural network, which consisted of multiple neurons. Each of the neurons could receive multiple inputs and outputs a result. The core of neural network was the hierarchical structure (such as input and output layer, as well as multiple hidden layers). Each layer was composed of multiple neurons and the output of each layer served as the input to the next one. By increasing the number of layers of the neural network, DL could learn more complex and abstract features to achieve more accurate classification and prediction. The convolutional neural networks (CNNs), deep neural networks (DNNs) and recurrent neural networks (RNNs), generative adversarial networks (GANs), long short-term memory (LSTM) and reinforcement learning (RL) were the representative models of DL. DL was suitable for more complex application scenarios, however, compared with ML, it also required more computing power and data. Recently, DL had surpassed many traditional ML algorithms and become the most promising algorithm to truly implement AI. With the assistance of ML and DL in the field of computer vision, image classification, intelligent identification, natural language processing (NLP), programmed decision-making and big data analysis, AI had obtained a significant improvement and been gradually applied into orthopedics, which brought new innovation to the diagnostic and therapeutic methods of orthopedic diseases (Muthukrishnan et al. 2020). In this paper we comprehensively introduced and reviewed the recent applications of AI in orthopedics, including the severity evaluation, triage, diagnosis, treatment and rehabilitation (as shown in Fig. 1). And as the special feature of this paper, we also firstly and particularly reviewed the most important applications of AI in orthopedics at current stage (AI-aided diagnosis of fracture) in detail, which almost included all the human skeletons (upper limb, lower limb, and axial skeleton). And the AI-diagnosis of other orthopedic diseases, such as osteoporosis, arthritis, ligaments and cartilage injuries, spinal diseases, bone tumor and bone age, was also introduced. Moreover, combined with our own previous studies, we also summarized the research points, relevant advantages and disadvantages of
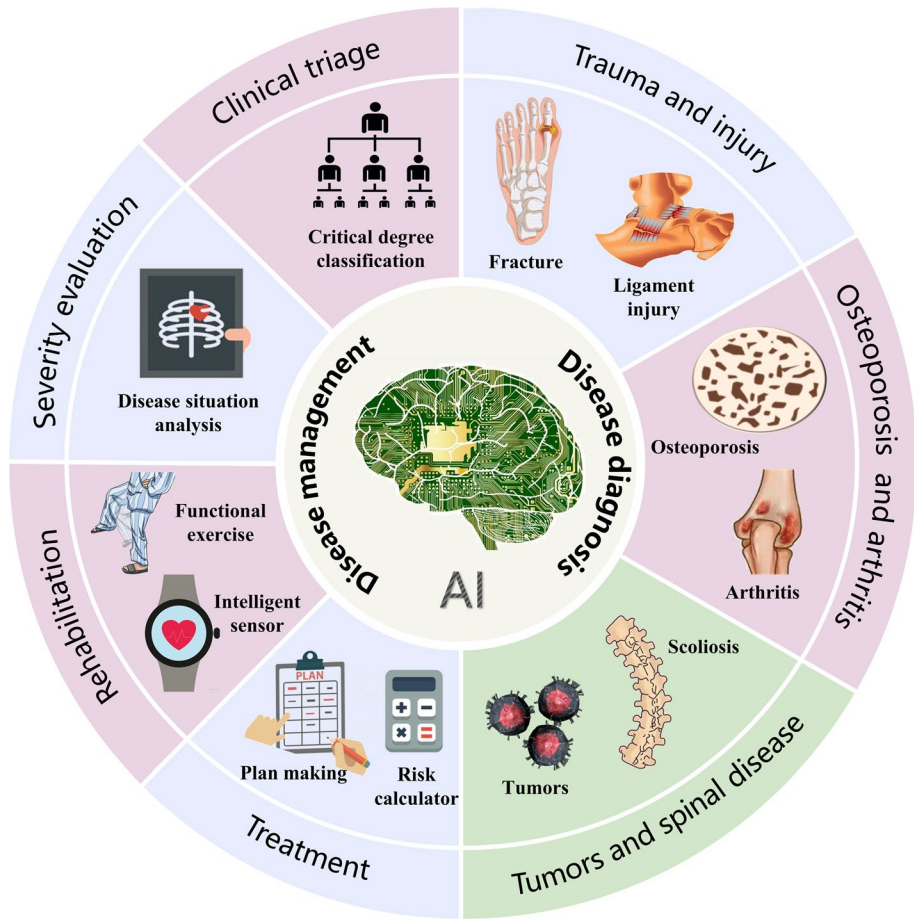
Fig. 1 The applications of AI in orthopedic severity evaluation, triage, diagnosis, treatment and rehabilitation

orthopedic AI, and discussed and shared the relevant research experience about the study of orthopedic AI, such as the balance of database and algorithm, the division of the database, the method of data labeling, the performance indexes of algorithm and other matters requiring attention. This paper could provide readers with a quick overview of the development of orthopedic AI and a deeper understanding of current clinical applications. We hoped to appeal for more attentions and effective applications and promote deeper integration of AI and orthopedics in the future.

2 AI in orthopedic diseases severity evaluation and triage

Most of the orthopedic patients coming into the emergency department for medical care were critical illness patients with open traumatic fracture, joint dislocation or multiple system merged injuries. However, the general crowding of emergency department, combined with insufficient medical resources and overloading works, usually resulted in

delayed patient treatment and a universal health care problem (Kim et al. 2018; Candel et al. 2022). Hence, implementing the rapid diseases severity evaluation and clinical triage of emergency patients, in such a severe environment, was crucial for subsequent medical treatment. The current clinical triage system such as emergency severity index (ESI) had effectively improved the process of the severity evaluation and triage, followed which the lifesaving could be made with the severity priorities (Ganjali et al. 2020). However, the ESI mostly relied on the human judgment of medical staffs at present. And with the patient's individual difference, misjudgments were quite normal and unavoidable in such conditions (Hussain et al. 2019). Hence, a more advanced and safe method was demanded to help clinicians accurately evaluate the conditions of patients.

With the application of NLP of AI technology, this issue had been greatly improved. Based on DL algorithm, the intelligent model could accurately process the clinical data and evaluate the conditions of patients, whose performance would be superior than that in traditional triage scale (Kang et al. 2020). Yao et al. proposed a DL-based model for patients' triage using the 5 years medical records of 864,043 patients in emergency department. In this study, the structured medical data was transferred into text form and imported to the CNN, combined with RNN and attention mechanisms to accomplish the supervised model training. The effects and performance were evaluated by the accuracy and area under the receiver operating characteristic curve (AUROC), which showed 0.83 and 0.87 in the internal testing dataset, and 0.83 and 0.88 in the external testing dataset. The model was also applied to predict the mortality and admission, whose results also expressed 0.3–0.5% higher in accuracy than other conventional methods (Yao et al. 2021). Raita et al. established four ML or DL models (lasso regression, random forest, gradient boosted decision tree, and CNNs) with the medical data of 135,470 patients in emergency department, 70% of the database was set as training dataset and 30% was set as testing dataset. The routinely available triage data was set as predictors (including demographics, triage vital signs, chief complaints, comorbidities) during the training process. After the supervised training, the performance of algorithms was evaluated with the testing dataset to predict the possible clinical outcomes of the injured patients (hospitalization (conventional hospital admission), critical care (admission to intensive care unit) and in-hospital death). The results showed that in the outcomes prediction, all of the four algorithms performed better than traditional ESI, which would enhance clinical triage making, achieving better clinical care and optimal resource utilization for the injured patients (Raita et al. 2019). Similarly, in a Korean 11,656,559-sample study, the in-hospital mortality, critical care and hospitalization of the patients in emergency department were also predicted using the clinical information as predictor variables, including age, sex, chief complaint, time from symptom onset to ED visit, arrival mode, trauma, initial vital signs and mental status. The results showed that the AUROC and area under the precision and recall curve (P-R curve) were 0.93 and 0.26, which significantly outperformed Korean triage and acuity score (AUROC: 0.78, AUPRC: 0.19), modified early warning score (AUROC: 0.81, AUPRC: 0.11), logistic regression (AUROC: 0.90, AUPRC: 0.2), and random forest (AUROC: 0.91, AUPRC: 0.17) (Kwon et al. 2018). The ML-based (XGBoost) triage and acuity score could make predictions more accurate than the existing scales, which was a further life-guarantee of the injured patients in emergency department (Klang et al. 2021). Clinical decision support system (CDSS), an intelligent model based on logistic regression analysis, was developed after the exploration and summary of big-volume clinical historical database, which finally realized the disease triage and offered an objective suggestion for clinicians to improve healthcare (Fernandes et al. 2020a, b). Emergency Department Early Warning Score (TREWS) was also established based on the univariable and multivariable regression analysis, which

made it better in promoting the diseases evaluation and triage of patients (Lee et al. 2020). Moreover, a research tested the predictive performance of several ML algorithms with a same database from the patients in emergency department and indicated that the decision trees model, LASSO logistic regression model, random forests model and gradient boosting machines model performed outstanding in the severity evaluation and triage for the injured patients (Patel et al. 2018). And there were also various analogous attempts of ML and DL algorithms (logistic regression, XGBoost, DNN) in emergency triage, which successfully realized the remote-triage in prehospital situations by the wearable device and saved more time for life-salvation of the injured patients (Hong et al. 2018; Fernandes et al. 2020a, b).

In summary, the application of AI in emergency department could effectively prompt the severity evaluation and emergency triage for patients with a scientific method, which also provided clinicians with reliable reference and reduced the occurrence of clinical adverse events. The AI-aided method was completely crucial for life rescue of patients. The summary of representative AI-severity evaluation and triage was shown in Table 1.

3 AI in orthopedic diagnosis

In the several aspects of AI in orthopedics, the AI-aided diagnosis was the most common application with confirmed effects. With the advantages of AI-computer vision and image identification technologies, the orthopedic diagnosis process was greatly improved. Image identification was the integration of a group of algorithms, which was used to understand the image content. It belonged to the subset of computer vision, which was the representative technology of AI. The core technology of image identification was the recognition of gray difference, with which the image content could be processed and understood and the different targets and objects could also be marked and identified. By inputting images with explicit classifications to train the model, the pre-defined labels could be outputted for the new images without labels. This process realized the intelligent diagnosis of medical image. The applications of AI-aided medical diagnosis had appeared in the identification of lung lesions (including pulmonary nodules, cancer, pneumothorax, mediastinal widening, consolidation, pleural effusion, atelectasis, fibrosis, calcification and even acute respiratory distress syndrome) on X-ray and CT images, which already achieved satisfying effects and entered the stage of clinical application (Nam et al. 2019, 2021; Sim et al. 2020; Sjoding et al. 2021; Seah et al. 2021).

In the field of orthopedics, X-ray, CT and MRI detection were also the most common way for clinical diagnosis of musculoskeletal diseases. Generally, the works of image reading were achievable for orthopedic clinicians in normal situations. However, owing to the overloaded clinical works, inadequate medical resources and lacking senior orthopedists, misdiagnosis and missed diagnosis usually occurred in emergency situations especially in the diagnosis of micro, occult or non-displaced fractures and other orthopedic diseases with nonspecific presentation (such as osteoporosis, arthritis, ligaments and cartilage injury, bone deformity, tumor and bone age assessment). This could generate severe consequences and be critical hindering for patients' treatment (Pinto et al. 2018). Guly et al. indicated that there were 953 diagnostic errors in an emergency department during past four years, and the most common reasons for the errors were misreading radiographs (about 77.8%) (Guly 2001). Duron et al. also illustrated that the physicians suffered from an ever-increasing workload of radiographs interpretation, and the missed fractures represent up

Table 1 The summary of representative AI-orthopedic diseases severity evaluation and triage

Application	Application type	Database		Type	Pre-process	Algorithm	Algorithm type	Framework	Supervised/ unsupervised learning	References
		Source	Size							
Remote triage in the prehospital environment	NLP	American National Trauma Data Bank	460,865	Text	Complete case analysis	Logistic regression, random forest, DNN	ML, DL	Tensor flow	Supervised	Kim et al. (2018)
Need prediction for critical care in prehospital emergency medical services	NLP	Korean National Emergency Department Information System	8,981,181	Text	N/A	DNN	DL	Tensor flow	Supervised	Kang et al. (2020)
Triage system in emergency department	NLP	Self-built database	864,043	Text	Data transformation	CNN, RNN	DL	Keras	Supervised	Yao et al. (2021)
Clinical outcomes prediction of ED	NLP	Self-built database	135,470	Text	N/A	Lasso regression, random forest, gradient boosted decision tree, DNN	ML, DL	Keras	Supervised	Raita et al. (2019)
Triage and acuity score of ED	NLP	Korean National Emergency Department Information System	11,656,559	Text	Data anonymization	DTAS	DL	N/A	Supervised	Kwon et al. (2018)

Table 1 (continued)

Application	Application type	Database Source	Size	Type	Pre-process	Algorithm	Algorithm type	Framework	Supervised/ unsupervised learning	References
Prediction of intensive care unit admission of ED	NLP	Self-built database	412,858	Text, tubular data	Lower-casing and removing punctuation and rare words	XGBoost	ML	N/A	Supervised	Klang et al. (2021)
Prediction of hospital admission of ED	NLP	Self-built database	560,486	Text	Data normalization	Logistic regression, XGBoost, DNN	ML, DL	Keras	Supervised	Hong et al. (2018)
Risk prediction of mortality and cardiopulmonary arrest in ED	NLP	Self-built database	599,276	Text	Data normalization	Logistic regression, random forests, XGBoost	ML	N/A	Supervised	Fernandes et al. (2020a, b)

It was not stated in the literature

ED emergency department, *N/A* not available

to 80% of diagnostic errors in the emergency department (Duron et al. 2021). Hence, it was still necessary to develop an automated and intelligent system for assisting orthopedists to complete the clinical diagnosis. The application of image identification of AI in the diagnosis of orthopedic diseases showed immense potential for the problem (as shown in Fig. 2).

3.1 In fracture

AI-assisted orthopedic diagnosis had got a great success especially in bone fracture, which almost covered most bone of the body prone to fracture. Through extensive literature review, we found that the studies of AI-aided fracture diagnosis were mainly about the bones around joints (including carpal joint, elbow joint, shoulder joint, ankle joint, knee joint and hip joint) as well as irregular and short bones (such as tarsal bone, vertebra, pelvis, clavicle, rib and skull). Their imaging manifestations were not typical to recognized and the overlapping and staggered bones also made it more difficult to locate the fracture lines, which could easily lead to missed diagnosis and misdiagnosis. Correspondingly, the fractures of long bones away from the joints (such as the normal fracture of ulna, radius, humerus, tibia, fibula and femur) were barely studied owing to the easy diagnosis of human level. Moreover, the relevant studies principally based on the database of X-rays and less based on the CT scans. On the basis of our previous work, we thought the reason could be attributed to the heavy workload of pre-classification and image labeling processes in the early stage of database establishment (more than 100 CT scans per patient versus 1–3 X-ray images per patient). And the individual difference, imaging diversity and complexity of CT scans also made the AI-diagnosis more difficult. We would introduce the mainstream studies of AI-aided fracture diagnosis in the order of upper limb, lower limb, and axial skeleton (pelvis, spine and skull) from distal part to proximal part.

3.1.1 The upper limb

For hand fracture, most patients with hand trauma were usually examined in emergency departments of hospitals. AI-aided method could assist physicians with hand X-rays
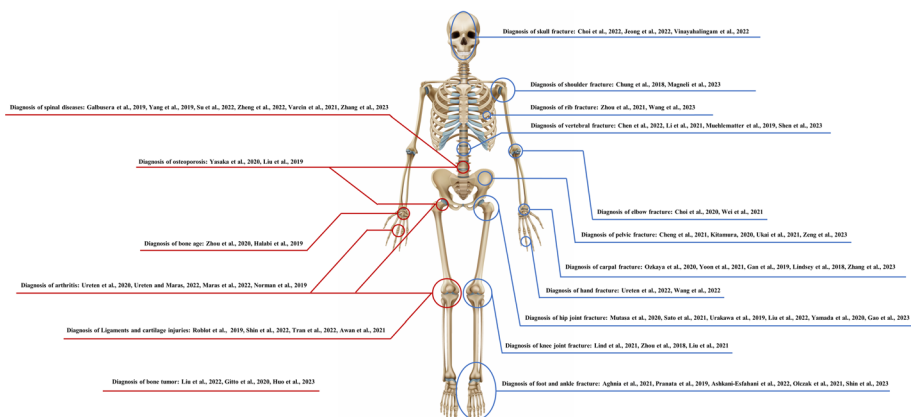

Fig. 2 The applications of AI in the diagnosis of orthopedic diseases

interpreting in the emergency department, especially in the situations lacking senior doctors, such as night shifts and weekends. Ureten et al. proposed a CNN algorithm and applied several DL algorithms (VGG-16, GoogLeNet and ResNet-50) in the supervised learning of image features of 275 fractured wrists, 257 fractured phalanx, and 270 normal hand X-rays. In the study, the data was resized as 224×224 pixels, and random translating and rotating were executed for data augmentation. After training, the accuracy, sensitivity, specificity, and precision in the classification of wrist fracture achieved 0.93, 0.96, 0.90 and 0.89, respectively, with VGG-16, achieved 0.88, 0.94, 0.84 and 0.82, respectively, with Resnet-50, and achieved 0.88, 0.90, 0.85 and 0.85, respectively, with GoogLeNet. And the accuracy, sensitivity, specificity, and precision in the detection of phalanx fracture were 0.84, 0.84, 0.83 and 0.82, respectively, with VGG-16, were 0.81, 0.81, 0.82 and 0.81, respectively, with GoogLeNet, and were 0.79, 0.78, 0.80 and 0.79, respectively, with Resnet-50. And the models performed better than human level, which greatly enhanced the fracture diagnosis of irregular bones on hand (Ureten et al. 2022). Wang et al. also built and trained a DL framework WrisNet based on the self-built database, including 4346 anteroposterior, lateral and oblique hand X-rays. The gray stretch and data augmentation (flipping, brightness and affine transformation, and sharpening) were applied for data preprocessing and augmentation. When the intersection over union (IOU) was set as 0.5, the network achieved 0.55 average precision (AP) in the hairline finger detection, which had an improvement of at least 0.05 over the other frameworks (Wang et al. 2022).

For the carpal fracture, the missed and untreated plan could lead to a progressive pattern of debilitating wrist arthritis, which might ultimately require negative salvage procedures, including wrist fusion. Scaphoid fractures were the most common carpal fracture, but as many as 20% of which were not visible in the initial injury radiograph. Hence, the occult scaphoid fracture was easily neglected in clinical diagnosis and usually resulted in osteonecrosis. The establishment DL model ResNet-50 effectively changed the unfavorable situation. After the supervised training with the X-ray from 390 patients with occult scaphoid fracture, the performance of the model reached 0.76 sensitivity and 0.92 specificity in the automatic rigorization of occult scaphoid fracture, and the AUROC was 0.84, F1 score value was 0.82. While the final performance of the algorithm was similar to a less experienced orthopedic specialist, it was better than the physician in the emergency department, which could effectively reduce the misdiagnosis and missed diagnosis of scaphoid fracture (Ozkaya et al. 2020). A study also built the X-ray-dataset compiled for 11,838 patients with possible scaphoid fractures, who presented to Chang Gung Memorial Hospital and Michigan Medicine between January 2001 and December 2019. In this study, the DL model EfficientNetB3 were trained to classify the occult scaphoid fractures, which achieved an overall sensitivity and specificity of 0.87and 0.92, respectively, with an AUROC of 0.95 in distinguishing the scaphoid fractures from normal scaphoids (Yoon et al. 2021). Distal radius fractures (DRF) were also common carpal fracture. Gan et al. trained the DL algorithm Inception-v4 by 2340 anterior–posterior wrist X-ray from patients with DRF. After the supervised training, when IOU was 0.5, the Inception-v4 performed well in the detection of DRF: the accuracy was 0.93, sensitivity was 0.90, specificity was 0.96 and the Youden Index was 0.86, which were better than the performances of orthopedists and radiologists. Moreover, the author also proposed a Faster R-CNN model (one of the fast object detection algorithms) to serve as an auxiliary algorithm for the Inception-v4 model in locating the regions of interest (ROI) on images, which had a 100% success rate in automatically annotating the ROIs on images. The participation of Faster R-CNN further simplified the workflow and reduced the workload of manual labeling (Gan et al. 2019). Lindsey et al. also developed the DL algorithm U-NET to detect and localize the DRF in X-rays.

The algorithm was trained to accurately emulate the expertise of 18 senior subspecialized orthopedists with their annotated 135,409 DRF X-rays. And a controlled experiment was also run with emergency medicine clinicians to evaluate their ability to detect DRF in wrist X-rays with and without the assistance of AI. The results showed that the average clinician's sensitivity in DRF detection was 0.80 unaided and 0.91 aided and specificity was 0.87 unaided and 0.93 aided. With the assistance of AI, the average clinician experienced a reduction in misinterpretation rate of 0.47 (Lindsey et al. 2018). In our own study, we also established an ensemble model consisted of three DL algorithms (RetinaNet, Faster RCNN and Cascade RCNN) to diagnosis the DRF. After training with 3276 wrist joint anteroposterior X-ray films and 3260 wrist joint lateral X-rays, the ensemble model got excellent accuracy (0.97), sensitivity (0.95) and specificity (0.98) in the DRF detection. The data was resized as 800×800 pixels, and the flipping of data augmentation was also performed. When IOU was set as 0.5, it performed better than clinical orthopedists and radiologists (Zhang et al. 2023a, b). The fracture of the styloid process of the ulna was also detected by the DL algorithm VGG16, which got a diagnostic accuracy of 0.91 ± 0.02 and AUROC of 0.95 (Oka et al. 2021).

For elbow fracture, Choi et al. developed a dual-input CNN-based DL algorithm that utilized both anteroposterior and lateral elbow X-rays to realize the automated detection of supracondylar fracture in conventional radiography. In the study of Choi et al., 1266 pairs of anteroposterior and lateral elbow X-rays were included, and the flipping, rotating, shifting, shearing, and zooming were performed for data augmentation. Finally, the database was split into the training set (1012 pairs, 79.9%) and a testing set (254 pairs, 20.1%). The AUROC, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of algorithm and human level were calculated and compared. The results showed that algorithm got a comparable AUROC (0.97), sensitivity (0.93) and NPV (0.97) to the human readers, and a better specificity (0.92) and PPV (0.80) than human level, which indicated that AI could provide an accurate diagnosis of supracondylar fracture comparable to radiologists (Choi et al. 2020). Radiography was an essential basis for the diagnosis of elbow fractures. To achieve better AI-assisted elbow diagnosis, the bone instance segmentation was necessary upstream task for automatic radiograph interpretation. Bone instance segmentation was a process with which each bone could be extracted separately from radiography. However, the arbitrary directions and the overlapping of bones posed issues for it. To solve this problem, Wei et al. designed a detection-segmentation pipeline by using rotational bounding boxes to detect bones and proposing a robust segmentation method. The proposed pipeline including (1) ResNet architecture for detecting and locating bones, (2) Oriented Bounding Box (OBB) for improving the location accuracy and (3) Global–Local Fusion Segmentation Network for combining the global and local contexts of the overlapped bones. The performance of the network was verified by a dataset that contained 1274 well-annotated elbow X-rays, and the qualitative and quantitative results indicated that the network significantly improved the performance of bone extraction (Wei et al. 2021). The methodology had good potential for applying DL in the X-ray's bone instance segmentation, which could be further enhancement for the AI-aided diagnosis of elbow fracture.

For the shoulder fracture, proximal humeral fractures accounted for a significant proportion, whose classification of the type and severity were important for clinical decision making. Chuang et al. trained the DL model ResNet-152 based on 1,891 proximal humeral X-ray images (515 normal images, 346 humeral greater tuberosity fractures, 514 humeral surgical neck fractures, 269 3-part fractures, 247 4-part fracture, and the data was cropped and resized as 200×200 pixels). After training the model showed a high performance

of 0.96 accuracy, 1.00 AUROC, 0.99 sensitivity, 0.97 specificity and 0.97 Youden index for distinguishing normal shoulders from proximal humerus fractures. In addition, when classifying proximal humeral fracture according to Neer-classification, the algorithm also got promising results with 0.65–0.86 accuracy, 0.90–0.98 AUROC, 0.88–0.97 sensitivity, 0.83–0.94 specificity and 0.71–0.90 Youden index. Compared with the human level, the CNN showed better performance to that of general physicians and orthopedists, and similar performance to orthopedists specialized in the shoulder. The superior performance of the CNN was more markable in the classification of complex 3- and 4-part fractures (Chung et al. 2018). There was also a study achieving the automatic diagnosis of proximal humeral fracture merely through the database of radiology text. The text reports of X-ray or CT from 1324 proximal humerus fracture patients were imported into the BERT model to make the training and characteristics extraction. The model finally achieved good accuracy of 0.61, precision of 0.5, recall of 0.39 and F1 score of 0.39, which were considered reasonable scores for sparse text data in the context of ML (Dipnall et al. 2022). For the whole diagnosis of shoulder fractures, Martin et al. also trained the DL algorithm ResNet based 7189 shoulder plain X-rays. The data was pre-processed by 256×256 pixels resizing and cropping, rotating, and inverting for data augmentation. After the supervised training, the got excellent overall AUROC for the detection of proximal humeral fractures (0.90), diaphyseal humeral fracture (0.97), clavicle fractures (0.96) and scapula fractures (0.87) (Magneli et al. 2023). And it was also the rare study involving scapular fracture, we thought it might be due to the lower incidence of scapular fracture than others and its atypical expression on X-rays. But we believed the relevant studies would appear more and more in the future, and the applications of the algorithms must be potential to speed up the diagnosis and classification task, which could be well assistance for radiologists and orthopedists. Except the clavicle fractures, the DL algorithm was also applied in the dating of clavicle fracture and realized encouraging results (Tsai et al. 2022). The summary of AI-diagnosis in upper limb fracture was shown in Table 2.

3.1.2 The lower limb

For the foot and ankle fracture, early and accurate detection were crucial for optimizing treatment and reducing future complications. Radiographs were also the most abundant imaging techniques for the fractures assessing. Hence, the AI-aided method would also be faster and more accurate in analyzing radiographic images than human intervention. Aghnia et al. applied the principal component analysis network (PCANet) as the architecture to detect the calcaneal fractures on CT scans. And the data augmentation (rotating, distorting and flipping) was also applied during training process, which improved network accuracy by almost 0.35 in classifying calcaneal fractures according to Sanders-classification. Finally, the proposed model achieved 0.72 accuracy in classifying CT calcaneal images into the four Sanders categories, which meant the AI-aided method was a feasible and efficient approach in assisting physicians in evaluating calcaneal fracture types (Aghnia et al. 2021). Pranata et al. also compared two types of DL architectures with different network depths (ResNet and VGG) in the recognition of calcaneus fractures on CT scans (including coronal, sagittal, and transverse views). The speeded-up robust features (SURF) method, canny edge detection and contour tracing were also applied in the bone fracture detection algorithm. The results showed that ResNet was comparable in accuracy (0.98) to the VGG network for calcaneus fractures detection but achieved better performance for involving a DNN architecture (Pranata et al. 2019). In a retrospective case–control study,

Table 2 The summary of AI-diagnosis in upper limb fracture

Application	Application type	Database				Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	Ref
		Source	Size	Type	Pre-process						
Hand fracture	Classification	Self-built database	802	X-ray	Cropping, 224 × 224 pixels resizing Data augmentation: translating and rotating	VGG-16, ResNet-50, GoogleNet	DL	N/A	Intel Core i7-9750H/ Y540/16G/512 GeForce RTX2060 computer and MATLAB	Supervised	Ureten et al. (2022)
	Detection	Self-built database	4346	X-ray	Gray stretch Data augmentation: flipping, brightness and affine transformation, and sharpening	WrisNet	DL	N/A	4 GPU NVIDIA GeForce RTX 3090	Supervised	Wang et al. (2022)
Carpal fracture	Classification	Self-built database	390	X-ray	Cropping, 196 × 196 pixels resizing	ResNet-50	DL	N/A	Python, OpenCV, Tkinter, NVIDIA GTX 1050 Ti graphics card	Supervised	Ozkaya et al. (2020)
	Segmentation	Self-built database	11, 838	X-ray	Cropping Data augmentation: flipping, scaling, brightness, rotating	Efficient-NetB3	DL	Pytorch	Python, Pydicom, OpenCV	Supervised	Yoon et al. (2021)

Table 2 (continued)

Application type	Database Source	Size	Type	Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	Ref
Classification	Self-built database	2340	X-ray	200×200 pixels resizing Data augmentation: inversing, offsetting, rotating, scaling, shearing	Inception-v4	DL	Tensor flow	eWorld Viewer, NVIDIA Titan X GPU	Supervised	Gan et al. (2019)
Segmentation	Self-built database	135,409	X-ray	1,024×512 pixels resizing Data augmentation: rotating, cropping, mirroring, brightness and contrast	U-NET	DL	N/A	N/A	Supervised	Lindsey et al. (2018)
Detection	Self-built database	6536	X-ray	800×800 pixels resizing Data augmentation: flipping	Retina Net, Faster RCNN, Cascade RCNN	DL	Pytorch	Python, Labelme, NVIDIA V100 GPU	Supervised	Zhang et al. (2023)

Table 2 (continued)

Application	Application type	Database Source	Size	Type	Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	Ref
Elbow fracture	Detection	Self-built database	1619	X-ray	200×200 pixels resizing Data augmentation: flipping, rotating, shifting, shearing, and zooming	ResNet-50	DL	Keras	Intel Core i7-7500U 2.7-GHz processor, NVIDIA GeForce 940MX GPU	Supervised	Choi et al. (2020)
	Segmentation	Self-built database	1274	X-ray	Sobel pre-process	ResNet	DL	Pytorch	4 NVIDIA Titan XP GPU	Supervised	Wei et al. (2021)
Shoulder fracture	Classification	Self-built database	1891	X-ray	Cropping, 200×200 pixels resizing	ResNet-152	DL	Caffe	NVIDIA GTX 1070	Supervised	Chung et al. (2018)
	Classification	Self-built database	7189	X-ray	256×256 pixels resizing Data augmentation: cropping, rotating, and inverting	ResNet	DL	N/A	N/A	Supervised	Magneli et al. (2023)

It was not stated in the literature

N/A not available

Soheil et al. assessed the performance of two different DL model (Inception V3 and Renet-50) in detecting ankle fractures using CT scans from 1050 patients with ankle fracture and the 1,050 individuals with healthy ankles. In the data pre-processing, random flipping and rotating were performed for data augmentation. Finally, the results showed a better performance of Inception V3 than ResNet-50, which got the sensitivity of 0.98 and specificity of 0.98 in the detection of ankle fractures. During the testing process, only one fracture was missed, which meant that AI could be used for developing the currently used image interpretation programs or used as a separate assistant solution for the clinicians to detect ankle fractures precisely (Ashkani et al. 2022). According to the 2018 AO Foundation/Orthopedic, Trauma Association (AO/OTA) were often too complex for human observers to learn and use, there was also research trained a DL network based on ResNet architecture with 4941 ankle X-rays to classify them according to the 2018 AO/OTA-classification. The average AUROC was 0.90 for correctly classifying malleolar type B fractures. However, it performed poorly in the classification of malleolar A fractures, which might be caused by the atypical expression of fibular tip avulsions (Olczak et al. 2021). Talus fracture with the osteochondral lesions was also one kind of ankle injury that's easily missed in the radiological diagnosis. To improve the clinical situation, Shin et al. developed a CNN framework and trained with 379 anteroposterior ankle X-rays. And the results showed that the performance of the framework for the AUROC, accuracy, PPV and NPV in the talus fracture detection were 0.77, 0.81, 0.81 and 0.82, respectively, which was very meaningful in diagnosing lesion (Shin et al. 2023).

Until now, there was less AI-diagnosis research about the fracture of the three cuneiform bones, metatarsal bones, navicular bone and cuboid bone, which remained a research blank to be attended.

For the knee joint fracture, the inherent complexity in terms of anatomy made it difficult to diagnose on a plain radiograph. Recently, a study had shown promising results for interpreting the radiographs of knee joint fracture. 6003 X-rays of knee joint fractures were included and a ResNet algorithm were constructed to categorize the fractures according to the 2018 AO/OTA-classification system. The results showed that mean AUROC was 0.87 for proximal tibia fractures, 0.89 for patella fractures and 0.89 for distal femur fractures. Almost 3/4 of AUROC estimating were above 0.8 and more than half reached 0.9 or above, which expressed that the DL could be used not only for fracture identification but also for more detailed classification of fractures around the knee joint (Lind et al. 2021). The accurate detection could not be separated from the automatic segmentation of knee joint anatomy. To improve the efficiency and accuracy of knee joint tissue segmentation and achieve higher recognized rate, there were studies discussed the effects of new method such as deep CNN, 3D fully connected conditional random field (CRF) and 3D simplex deformable modeling, through which the femur, tibia, patella, muscle, cartilage, meniscus, quadriceps, patellar tendon, infrapatellar fat pad, joint effusion and Baker's cyst were well segmented (Zhou et al. 2018; Cheng et al. 2022). As the pivotal location of force conduction among the lower extremity, the proximal tibia could be damaged by a compression fracture, split fracture, bone defect or other structural injuries during an excessive violent load. The tibial plateau fracture was one kind of proximal tibia fracture in the knee joint. It was a severe articular injury with a broad damage-spectrum to the locomotor system, which usually accompanied with poor clinical effect and limited articular function. The early and accurate diagnosis of tibial plateau fracture were crucial for the treatment. In our previous study, the DL algorithm named RetinaNet was trained based on 542 anterior X-rays (458 for training dataset, 84 for testing dataset) of knee joint to detect the tibial plateau fractures.

The operating environment of the algorithm was NVIDIA GeForce RTX 3080 GPU. Finally, RetinaNet showed a detecting accuracy of 0.91 for the identification of tibial plateau fractures, which was comparable to the performance of an orthopedic physician panel. And the average time spent of per detection of the algorithm was 0.56 s, which was 16 times faster than human level (Liu et al. 2021a, b).The result further illustrated that DL was a valid and efficient method for the clinical diagnosis of tibial plateau fractures, which could be a useful assistant for orthopedists, and largely promote clinical workflow.

For the hip join fracture, the femoral neck fractures and intertrochanteric fractures were the most common result of violent hip injuries. Mutasa et al. applied the DL algorithm with advanced data augmentation (flipping, rotating, contrast and addition of Gaussian noise matrix) to accurately diagnose and classify femoral neck fractures. The self-built database included 1063 hip X-rays from 550 patients with the labels of Garden fracture classification, which consisted of 127 Garden I and II fracture X-rays, 610 Garden III and IV fracture X-rays and 326 normal hip X-rays. And the results showed that the two-classify prediction between fractures and normal hip achieved AUROC of 0.92, accuracy of 0.92, sensitivity of 0.91, specificity of 0.93, PPV of 0.96 and NPV of 0.86. And the three-classify prediction of Garden I/II, Garden III/IV or no fracture got the performance of 0.96 AUROC, 0.86 accuracy, 0.79 sensitivity, 0.90 specificity, 0.80 PPV and 0.90 NPV (Mutasa et al. 2020). Sato et al. also trained a DL model Net-B4 by 5242 hip X-rays with femoral neck fracture from 4851 cases and 5242 images without fracture site, whose accuracy, sensitivity, specificity, F-value and AUROC were 0.96, 0.95, 0.96, 0.96 and 0.99, respectively. The controlled experiment was also performed, which illustrated that the diagnostic accuracy of the young residents in orthopedics department was significantly improved with the assistant of the model (Sato et al. 2021). The automatic detection of femoral intertrochanteric fractures was also accomplished by the DL algorithm VGG-16 with the database of 3346 hip images, whose accuracy, sensitivity, and specificity were 0.95, 0.93 and 0.97 respectively, exceeding that of orthopedic surgeons (Urakawa et al. 2019). In our previous study, we also realized the detection of femoral intertrochanteric fractures by the DL algorithm, Faster-RCNN. 700 X-rays of femoral intertrochanteric fractures patient were collected and resized as 600×800 pixels. Then the image was labeled by the labeling software LabelImg and the database were divided into the training dataset and test dataset in a ratio of 9:1. Finally, compared with orthopedic physicians, the Faster-RCNN algorithm performed better in accuracy (0.88), specificity (0.87), misdiagnosis rate (0.13) and time consumption (5 min). And as for the sensitivity and missed diagnosis rate, there was no significant difference between the Faster-RCNN and human level. The operating environment was NVIDIA GeForce RTX 3080 GPU (Liu et al. 2022a, b). Our study further proved that DL was an effective assistant for the diagnosis of femoral intertrochanteric fractures. A study multi-center study from Stanford University School of Medicine and University of Adelaide trained the DL algorithm DenseNet with 172 layers on the database from Royal Adelaide Hospital, which consisted of 45,786 proximal femoral X-rays with a fracture prevalence of 11%. In the internal testing dataset (200 fracture cases and 200 non-fractures), the DenseNet got strong AUROC of 0.99, which was better than five human radiologists (0.96). Furthermly, in the external validation study, an external testing dataset from Stanford University Medical Center, consisted of 40 fracture X-rays (22 involved fractures in the trochanteric region, 18 involved fractures of the femoral neck) and 41 negative cases, it also arrived 0.98 AUROC (Oakden-Rayner et al. 2022). Moreover, there was also study realized the classification of different hip fractures (three-classify of femoral neck fractures, intertrochanteric fractures and normal hip), whose average accuracy, recall, precision, and F1

score of the DL model Xception achieved 0.98, 0.98, 0.98 and 0.98, respectively. And the performance of the model was significantly better than that of the orthopedists (Yamada et al. 2020).

Including our own study, to eliminate the visual interference and get better training effect and the recognition ability, most of the prior studies with large training database had set the wide exclusion criterion and exclude radiographs with implants, other non-hip fractures, poor positioning or suboptimal image quality, which could potentially introduce the selection bias and restrict the applicability of the trained framework in the real-world population. To avoid the limitation, Gao et al. developed and examined the performance of the DL model DenseNet based on the database of 40,000 X-rays, which particularly included all kind of frontal pelvic X-rays regardless of perceived image quality, presence of other non-hip fractures or metallic implants (more than 34.3%) to simulate the real clinical situations. The performance of the model was surprising, which also achieved high sensitivity (0.94) and specificity (0.96). It meant that the wide exclusion criterion of low-quality data during the database establishment might not be necessary, and the comprehensive performance of the algorithm should be consider with the algorithm property and data annotation level (Gao et al. 2023). The summary of AI-diagnosis in lower limb fracture was shown in Table 3.

3.1.3 The axial skeleton

For the pelvic fractures, pelvic fracture was a severe trauma with high rate of morbidity and mortality. The pelvic X-ray was essential for detecting the fracture lines in trauma patients, which was also the key component for trauma survey. Cheng et al. developed a multiscale DL algorithm named PelviXNet and trained it with 5204 pelvic X-rays with supervised point annotation. In this study, the data was cropped and resized as 1024×1024 pixels, and the random translating, rescaling, flipping and rotating were also performed for data augmentation. After training, the PelviXNet yielded an AUROC of 0.97 in the clinical population testing set of 1,888 pelvic X-rays. And the accuracy, sensitivity, and specificity were 0.92, 0.90 and 0.93, respectively, which demonstrated a comparable performance with radiologists and orthopedics in detecting pelvic and hip fractures (Cheng et al. 2021). Kitamura created and tested the DL model Densenet-121 to detect pelvic X-rays position, hardware presence, and pelvic and acetabular fractures. The database including 14,374 pelvic X-rays and random flipping, cropping and adjustment of brightness and contrast were applied for data augmentation. The results showed that the position and hardware models performed well with AUROCC of 0.99–1.00, the pelvic and acetabular fracture detection model got the performance as low as 0.70 for the pelvis fracture and as high as 0.85 for the acetabular fracture (Kitamura 2020). Accurate and automatic diagnosis and surgical planning of pelvic fracture required effective identification and localization of the fracture area. In addition to X-ray detection, the CT scans diagnostic system was proposed based on the YOLOv3 models (multiple, real-time object detection system), in which each YOLOv3 model was trained using differently orientated CT scans. The system was validated in 93 patients with pelvic fractures, which got AUROC of 0.82, recall of 0.80 and precision of 0.90 (Ukai et al. 2021). Similarly, the group of Zeng et al. had also developed a novel DL framework UNET for the automatic identification and localization of complex pelvic fractures in the CT scans. The framework was implemented with supervised learning and consisted of two weight-shared branches with a structural attention mechanism, to minimize the confusion of local complex structures of the pelvic bones with the fracture

Table 3 The summary of AI-diagnosis in lower limb fracture

Application	Application type	Database				Algorithm	Algorithm type	Framework	Tools	Supervised/unsupervised learning	References
		Source	Size	Type	Pre-process						
Foot and ankle fracture	Classification	Self-built database	760	CT	360×360 pixels resizing, noise removing Data augmentation: rotating, distorting, and flipping	PCANet	DL	N/A	Python, OpenCV	Supervised	Aghnia et al. (2021)
	Classification and detection	Self-built database	1931	CT	224×224 pixels resizing	ResNet, VGG	DL	N/A	N/A	Supervised	Pranata et al. (2019)
	Detection	Self-built database	2100	X-ray	Data augmentation: flipping and rotating	Inception V3, ResNet-50	DL	Tensor flow	Tesla K80 GPU	Supervised	Ashkani et al. (2022)
	Classification	Self-built database	4941	X-ray	Cropping, 256×256 pixels resizing	ResNet	DL	N/A	N/A	Supervised	Olczak et al. (2021)
	Classification	Self-built database	379	X-ray	Cropping, 224×224 pixels resizing	CNN	DL	Tensor flow	N/A	Supervised	Shin et al. (2023)
Knee joint fracture	Detection	Self-built database	6003	X-ray	Cropping, 256×256 pixels resizing	ResNet	DL	Pytorch	N/A	Supervised	Lind et al. (2021)
	Segmentation	Self-built database	3420	MRI	Cropping, 320×224 pixels resizing. Data augmentation: translating, shearing and rotating	CNN	DL	Keras	MATLAB, NVIDIA GeForce GTX 1080 Ti graphic card with total 3584 cores, and 11 GB GDDR-5RAM	Supervised	Zhou et al. (2018)

Table 3 (continued)

Application	Application type	Database		Size	Type	Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	References
		Source										
	Detection	Self-built database		542	X-ray	Cropping, 1080×1080 pixels resizing	RetinaNet	DL	Pytorch	Python, Labelme, NVIDIA GeForce RTX 3080 GPU	Supervised	Liu et al. (2021a, b)
Hip join fracture	Classification	Self-built database		1063	X-ray	256×256 pixels resizing Data augmentation: flipping, rotating, contrast and addition of Gaussian noise matrix	CNN	DL	Tensor flow	Python, DICOM Web Viewer, NVIDIA Titan X Pascal GPU, Intel Core i7 CPU and 32 GB RAM	Supervised	Mutasa et al. (2020)
	Detection	Self-built database		10,484	X-ray	256×256 pixels resizing	EfficientNet-B4	DL	Pytorch	N/A	Supervised	Sato et al. (2021)
	Detection	Self-built database		3346	X-ray	Cropping, 300×300 pixels resizing Data augmentation: rotating, shifting, shearing, flipping,	VGG-16	DL	Tensor flow	N/A	Supervised	Urakawa et al. (2019)
	Detection	Self-built database		700	X-ray	600×800 pixels resizing	Faster RCNN	DL	Tensor flow	Python, FreeRename, LabelImg, NVIDIA GeForce RTX 3080 GPU	Supervised	Liu et al. (2022a, b)

Table 3 (continued)

Application	Application type	Database			Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	References
		Source	Size	Type	Pre-process						
	Classification	Self-built database	1703	X-ray	Cropping, 299×299 pixels resizing Data augmentation: rotating, shifting, brightness, shearing, flipping	Xception	DL	Tensor flow	Python, Image-DataGenerator, NVIDIA GeForce RTX 2060 GPU, Intel Core i7-9750, 16 GB RAM	Supervised	Yamada et al. (2020)
	Detection	Self-built database	40,203	X-ray	512×512 pixels resizing Data augmentation: zooming, translating, flipping, rotating, brightness	DenseNet-121	DL	Tensor flow	NVIDIA Quadro TRX 6000 GPU with 32 GB DDR4	Supervised	Gao et al. (2023)

It was not stated in the literature

N/A not available

zones. It also allowed to combine the symmetry properties of the pelvic anatomy and capture the symmetric feature differences on both the left and right sides, which overcame the limitations of existing methods usually considering only image or geometric features. The comprehensive experiments on 103 clinical CT scans from the publicly available database showed that the framework achieved accuracy and sensitivity of 0.92 and 0.93 (Zeng et al. 2023).

For the vertebral fractures, they were the most common fractures in high fall injured patients or osteoporotic fractures in older individuals. Chen et al. developed a DL model ResNet-50 for classifying fresh vertebral compression fractures from X-rays, with MRI as the reference standard. 1877 X-rays of vertebral compression fractures in 1099 patients were included, and the model reaching an AUROC of 0.80, accuracy of 0.74, sensitivity of 0.80 and specificity of 0.68. Chen also indicated that in the detection process the lateral (AUROC, 0.83) views exhibited better performance than anteroposterior views (AUROC, 0.77) (Chen et al. 2022). Li et al. demonstrated the DL model YOLOv3 (consisted of object detection, data pre-processing and classification to detect vertebral fractures) with excellent accuracy (0.93), sensitivity (0.91), and specificity (0.93) for detecting vertebral fractures of the lumbar spine, and the AUROCs for the classifying of Grades I, II and III vertebral fractures were 0.91, 0.98 and 0.99, respectively. The interobserver reliability (Kappa value) of the DL performance and human observers was also calculated to estimate the effects of the model, which got 0.72 and 0.77 for thoracic and lumbar vertebrae (generally, Kappa value ≤ 0.4 meant poor consistency, $0.40 <$ Kappa value ≤ 0.60 meant moderate consistency, $0.60 <$ Kappa value ≤ 0.80 meant high consistency, Kappa value > 0.80 meant excellent consistency) (Li et al. 2021a, b). Derkatch et al. also set up a CNN not only realized the identification of vertebral fractures on X-ray with high performance (0.94 AUROC, 0.87 sensitivity and 0.88 specificity), but also achieved the prediction of vertebral fractures by the bone mineral density measurements on the picture (Derkatch et al. 2019). With the development of DL predictive model, through the analysis of bone texture on the standard lumbar CT scans the prediction could also be generated for the patients at risk of vertebrae fractures (Muehlematter et al. 2019). Osteoporotic vertebral fracture was a risk factor for morbidity and mortality in elderly population, which meant the accurate diagnosis was crucially important for improving clinical outcomes. In recent study of Shen et al. the detection and segmentation of osteoporotic vertebral fracture were also realized by the DL algorithm named AI-OVF-SH. After training with the 11,397 lumbar lateral X-rays from six clinical centers, the algorithm got the accuracy, sensitivity, and specificity of 0.97, 0.84 and 0.97 for all fractures in the internal testing dataset and 0.96, 0.83 and 0.94 in the 1,276 X-rays of external testing dataset (Shen et al. 2023).

For the rib fractures, they usually occurred in 40–80% of thoracic blunt trauma events and might lead to severe complications, such as pneumonia, lung contusion, haemothorax, even death. However, the interpretation of all ribs on more than hundreds of CT scans was a time-consuming and labor-intensive work, which had already been reported that the missed diagnosis rate of rib fractures was as high as 20.9%, significantly higher than the fractures of other position (Urbaneja et al. 2019). A retrospective study collected the CT scans from 2658 rib fracture patients and applied a Faster R-CNN model to detect the fracture site, which yielded good classification performance for the classifying of fresh, healing, and old rib fractures. Compared with experienced radiologists, the DL model achieved a higher sensitivity (0.95 vs. 0.77), comparable precision (0.91 vs. 0.87), and a shorter diagnosis time (a reduction of 126.15 s) (Zhou et al. 2021). And with the assistance of DL method the diagnostic performance of rib fractures of orthopedist was greatly improved with precision from 0.80 to 0.91, sensitivity from 0.62 to 0.86, and a reduction of 73.9 s

time consuming (Zhou et al. 2020a, b). In the study of Wang et al. the DNN algorithm RB Net was developed and trained on the database of 13,821 thoracic CT scans from 15 different hospitals to realize the rib segmentation and fracture detection. The model performance varied greatly with different fracture patterns. Both in the internal testing dataset and external testing dataset, the model achieved the highest sensitivity for displaced fractures (0.98, 0.98), followed by old fractures (0.93, 0.92), non-displaced fractures (0.89, 0.85), and buckle fractures (0.82, 0.70), which was in accordance with the different conspicuousness of these types of rib fractures. The study also indicated that the buckle fracture was the most visually inconspicuous and hence the most common type of missed fractures both for human and algorithm (Wang et al. 2023).

In summary, with human-AI collaboration, orthopedists would achieve higher performance in the detection of rib fractures than human-only, which provided a clinically applicable method to assist the works in clinical practice (Jin et al. 2020).

For the skull fracture, the head trauma was a significant cause of morbidity and mortality worldwide. The increasing number of emergency department visits for head trauma had become a public health concern. Based a database of 508 skull X-rays, Choi et al. trained the object detection DL frameworks (YOLOv3) to detect the skull fractures. After the testing by the internal or external testing dataset, the model expressed an AUROC of 0.92 and 0.87, a sensitivity of 0.81 and 0.78, a specificity of 0.91 and 0.88, respectively. With the assistance of YOLOv3, a significant AUROC improvement was observed in radiologists and emergency physicians with the difference from reading without AI assistance of 0.094 and 0.069, respectively (Choi et al. 2022). Similarly, the RetinaNet architecture in the DL model trained with 2026 skull X-rays (991 fracture, 1035 normal) also got a precision of 0.72, 0.66 and 0.36, respectively, when IOU was set as 0.1, 0.3 and 0.5 (Jeong et al. 2022). The DL algorithm Faster RCNN was also proposed on 6404 mandibular X-rays with manually annotated and labelled as a reference to detect mandibular fractures. In the testing dataset consisting of 149 X-rays with fracture and 171 X-rays without fracture, the trained algorithm got F1 score of 0.94 and an AUROC of 0.97 for the automatic fracture detection, which assisted orthopedists to reduce the misdiagnosis (Vinayahalingam et al. 2022). Besides, there was less AI-aided diagnosis of sternum fracture in the field of axial skeleton. The summary of AI-diagnosis in axial skeleton fracture was shown in Table 4.

3.2 In other orthopedic diseases

Except the common applications for fracture diagnosis, AI technology had also been widely applied in the diagnosis of other orthopedic diseases, such as osteoporosis, arthritis, ligaments and cartilage injury, spinal disorder and deformities, bone tumor and bone age assessment, whose imaging expression might also be uneasy to estimate.

3.2.1 Osteoporosis

Osteoporosis was defined as a systemic skeletal disease, which characterized by low bone mass and microarchitectural deterioration of bone tissue, and a consequently increased bone fragility and susceptibility of fracture. Osteoporosis was also one of the causes of fragility fracture among old population, which relied on the dual-emission X-ray absorptiometry (DXA) as gold standard for determination of bone mineral density (BMD) to make a definite diagnosis (Kanis et al. 2019). However, the disappointed situation of difficult result reading of DXA and the examination noises brought lots of inconvenience to the

Table 4 The summary of AI-diagnosis in axial skeleton fracture

Application	Application type	Database				Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	References
		Source	Size	Type	Pre-process	Algorithm					
Pelvic fracture	Detection	Self-built database	5204	X-ray	Cropping, 1024×1024 pixels resizing Data augmentation: translating, rescaling, flipping, rotating	PelviXNet	DL	Pytorch	Python, Intel Xeon E5-2650 v4 CPU, 2.2 GHz 128 GB RAM, 4 NVIDIA Titan V GPUs	Supervised	Cheng et al. (2021)
	Classification	Self-built database	14,374	X-ray	224×224 pixels resizing Data augmentation: brightness, contrast, flipping, cropping	Densenet-121	DL	Tensor flow	Python, Pydicom, GeForce 1080 GTX GPU	Supervised	Kitamura (2020)
	Segmentation	Self-built database	389	CT	282×282×282 voxel resizing, multi-orientated image synthesis and fracture region extraction	YOLOv3	DL	N/A	Core i9-10900 k CPU, Titan RTX GPU	Supervised	Ukai et al. (2021)
	Segmentation	Open source database	103	CT	64×64×64 voxel resizing Data augmentation: translating, flipping	UNet, DBSCAN, ResNet-50	DL	Pytorch	NVIDIA GeForce RTX 3090 GPU	Supervised	Zeng et al. (2023)

Table 4 (continued)

Application	Application type	Database Source	Size	Type	Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	References
Vertebral fracture	Classification	Self-built database	1877	X-ray	256×256 pixels resizing	ResNet-50	DL	Pytorch	InferScholar Center, 4 NVIDIA GTX 1080Ti GPUs	Supervised	Chen et al. (2022)
	Detection	Self-built database	941	X-ray	224×224 pixels resizing Data augmentation: color inversing, Gaussian blur, contrast, brightness	YOLOv3	DL	Tensor flow	Python, NVIDIA GeForce GTX 1080 Ti	Supervised	Li et al. (2021a, b)
	Detection	Self-built database	120	CT	Scaling, centering, redundant features removing	MLP, RF, SVM, Bayes classifier, ANN	ML, DL	N/A	MaZda	Supervised	Muehlematter et al. (2019)
	Segmentation	Self-built database	12,673	X-ray	640×640 pixels resizing	AI-OVF-SH	DL	N/A	Annotation platform (https:// spine.jying hui.cn)	Supervised	Shen et al. (2023)
Rib fracture	Detection	Self-built database	26,580	CT	Data augmentation: rotating, flipping,	Faster RCNN	DL	N/A	LabelImg	Supervised	Zhou et al. (2021)

Table 4 (continued)

Application	Application type	Database Source	Size	Type	Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/unsupervised learning	References
Skull fracture	Segmentation	Self-built database	13,821	CT	N/A	RB Net	DL	N/A	uAI Research Portal	Supervised	Wang et al. (2023)
	Detection	Self-built database	508	X-ray	Cropping, 608×608 pixels resizing Data augmentation: rotating, flipping	YOLOv3	DL	N/A	VGG Image Annotator	Supervised	Choi et al. (2022)
	Detection	Self-built database	2026		8-bit image converting	RetinaNet	DL	Keras	ImageJ, Python, OpenCV, 4 GeForce RTX 2080Ti GPU, Xeon Gold 6238 CPU and 320 GB RAM	Supervised	Jeong et al. (2022)
	Detection	Self-built database	6404		Data augmentation: flipping	Faster RCNN	DL	Pytorch	NVIDIA Titan V GPU 12G	Supervised	Vinayahalingam et al. (2022)

It was not stated in the literature

N/A not available

orthopedists. Hence, Yasaka et al. employed an DL model BMD-CNNs by the database of 1665 lumbar CT scans from 183 patients to extract BMD of lumbar vertebra, the 60-fold data augmentation was also applied to achieve 99,900 images by noise adding, random parallel shifting and rotating. The result showed that the predicted BMD values from the CNN model were significantly correlated with the BMD values from DXA (Pearson's correlation coefficient was 0.852, and osteoporosis was diagnosed with AUROC of 0.96, which realized the automatic diagnosis of osteoporosis by normal CT scans) (Yasaka et al. 2020). Moreover, there was also study further achieved the gradation of osteoporosis by the improved U-Net model, which achieved the accuracy of 0.81 (Liu et al. 2019).

3.2.2 Arthritis

Arthritis was a disease arising from the degeneration of joint, which presented discomfort symptoms such as swelling, pain, snapping and effusion. The middle-aged and elderly people were high-risk people, who were often accompanied by joint swelling and pain, effusion, limited activities, and other complications. However, the imaging manifestations were usually time-consuming and not easy to interpret without experienced orthopedists. Ureten et al. developed a series of algorithms to solve the problems in the diagnosis of arthritis. For the hip osteoarthritis, Ureten applied the VGG-16 network and transfer learning with a database consisted of 221 normal hip X-rays and 213 hip X-rays with osteoarthritis, which were evaluated with performance of 0.90 accuracy, 0.97 sensitivity, 0.83 specificity and 0.84 precision(Ureten et al. 2020). For the hand joint rheumatoid arthritis, the YOLO-v4 algorithm was used for objective detection in 1426 original hand X-rays without data loss, and classification was made by the application of transfer learning with a pre-trained VGG-16 network. The results showed that the classification of rheumatoid arthritis and normal hand X-rays got 0.90, 0.92, 0.88, 0.89 and 0.97 accuracy, sensitivity, specificity, precision and AUROC, respectively. And in the classification of rheumatoid arthritis, osteoarthritis and normal hand X-rays, an 0.80 accuracy result was obtained (Ureten and Maras 2022). The diagnosis of sacroiliitis and cervical arthritis were also realized with the VGG-16 network, which got accuracy of 0.89 and 0.93, sensitivity of 0.90 and 0.95, specificity of 0.88 and 0.92, and precision of 0.88 and 0.92, respectively(Ureten et al. 2021; Maras et al. 2022). Except for the automatic diagnosis of arthritis on X-ray, Zhou et al. also investigated the application of DL model in MRI to diagnose knee osteoarthritis. The MRI scans of 104 patients with knee osteoarthritis were selected as the research subjects and an image super-resolution algorithm based on multiscale wide residual network model was proposed and compared with the single-shot multibox detector (SSD) algorithm, superresolution convolutional neural network (SRCNN) algorithm and enhanced deep superresolution (EDSR) algorithm. Moreover, the diagnostic performances upon different MRI sequences were also analyzed to determine the best optimal sequence in the automatic recognition, which applied the arthroscopic results as the gold standard. The results showed that the model performed better than others and the 3D-DS-WE and T2* sequences were found to be the best sequence for diagnosing knee osteoarthritis, which got high diagnostic accuracy of over 0.95 in grade IV lesions. And the consistency test also indicated that the 3D-DS-WE and T2* sequences had a strong consistency with the results of arthroscopy (Kappa values=0.74 and 0.68, respectively) (Hu et al. 2022). Moreover, Norman et al. also designed a knee osteoarthritis detection neural network based on Kallgren Lawrence (KL) classification system. After training with the database of 4490 images, for non, mild, moderate, and severe knee osteoarthritis, the algorithm achieved the sensitivity rate of 0.83, 0.70,

0.68 and 0.86, the specificity of 0.86, 0.83, 0.97 and 0.99, which provided orthopedists with more accurate arthritis judgment (Norman et al. 2019). The detection of patellofemoral osteoarthritis on the knee lateral view X-rays was also realized, which got AUROC of 0.95(Bayramoglu et al. 2021).

3.2.3 Ligaments and cartilage injuries

Ligaments and cartilage injuries were the most frequent injuries in the motor system, such as meniscus tear and cruciate ligament rupture. And MRI was a useful method for detecting the ligaments and cartilage injuries with high sensitivity and specificity for that task. However, the MRI imaging reading might be difficult for inexperienced orthopedic junior doctors, which remained potential medical risks. To help orthopedists with MRI diagnosis of meniscus tears, Roblot et al. proposed a Faster-R CNNs algorithm based on 1123 knee MRI images, which yielded an AUROC of 0.94 in meniscus tear detection. What's more, the orientation of the tear could also be recognized with AUROC of 0.83 (Roblot et al. 2019). Qiu et al. also fused two CNNs models based on 2460 MRI scans collected from 205 patients in the hospital to diagnose meniscus injury, which got accuracy 0.93, sensitivity of 0.91, specificity of 0.94 and AUROC of 0.96 (Qiu et al. 2021). In the study of Shin et al. all the types of meniscal tears (medial, lateral or medial and lateral) could be accurately differentiated, and the classifying of horizontal, complex, radial and longitudinal tears were also recognized with AUROC of 0.76, 0.85, 0.60 and 0.85, respectively (Shin et al. 2022). For the detection of anterior cruciate ligament tears, CNN model also played a crucial part with a database of 19,765 knee MRI scans from 17,738 patients, which finally achieved a satisfying performance with AUROC of 0.93, sensitivity of 0.87, specificity of 0.9 in two external open-source datasets (KneeMRI and MRNet) (Tran et al. 2022). Moreover, Awan et al. also proposed a customized 14 layers ResNet-14 architecture of CNN with six different directions by using class balancing and data augmentation. The algorithm performed well not only in the detection of anterior cruciate ligament tears, but also in the classifying of healthy tear, partial tear or fully ruptured tear with AUROC of 0.98, 0.97 and 0.99, respectively (Awan et al. 2021). The AI-based MRI technology of ligaments and cartilage injuries had high practical value in clinical practice, which could effectively improve the accuracy of diagnosis, reduce the rate of misdiagnosis and time consumption.

3.2.4 Spinal diseases

Spinal diseases were commonly diagnosed by the radiological examinations and the accurate angle and dimension measurements were also required, which could be hard and time-consumed to operated manually. The application of AI was eagerly anticipated to support the diagnosis of spinal diseases which required highly specialized expertise. There were already AI models achieved outstanding performance in the automatic diagnosis of spinal diseases, such as scoliosis, disc herniation and lumbar spondylolisthesis. For instance, the fully standard convolutional network (FCN) model was trained with the database of 493 spine-images of patients suffering from various disorders, including adolescent idiopathic scoliosis, adult deformities, and spinal stenosis. The end plate centers, hip joint centers, and margins of the S1 end plate were set as landmarks for the calculation of anatomical parameters (including T4-T12 kyphosis, L1–L5 lordosis, Cobb angle of scoliosis, pelvic incidence, sacral slope and pelvic tilt). As a results, the FCN performed well in the recognition of spinal sagittal/coronal deformities and degenerative phenomena, and the standard

errors of the estimated parameters merely ranged from 2.7° (for the pelvic tilt) to 11.5° (for the L1–L5 lordosis) (Galbusera et al. 2019). Even more striking, there was also DL model could directly utilize the unclothed back images to detect scoliosis, whose accuracy was superior to those of human specialists in detecting scoliosis, detecting cases with a curve ≥ 20° and severity grading for both binary classifications and the four-class classification. This method could be potentially applied in routine scoliosis screening and periodic follow-ups of pretreatment cases without radiation exposure (Yang et al. 2019). Watanabe et al. also created a scoliosis screening system to estimate the spinal alignment, the Cobb angle, and vertebral rotation from moiré images. In the system, the positions of 12 thoracic and 5 lumbar vertebrae, 17 spinous processes and the vertebral rotation angle of each vertebra could also be accurately located and calculated by the algorithm. Finally, the mean absolute error (MAE) of the estimated vertebral positions was 5.4 mm per person, was 3.42° of the Cobb angles and was $2.9° \pm 1.4°$ of the angle of vertebral rotation. And the MAE was 4.38° in normal spines, was 3.13° in spines with a slight deformity, and was 2.74° in spines with a mild to severe deformity, which greatly enhanced the diagnosis accuracy of scoliosis (Watanabe et al. 2019). The recognition and grading of disc herniation, central canal stenosis and nerve roots compression were also realized by the ResNet-50 algorithm with the database of 1273 axial T2-MRI scans, which got the accuracies of 0.84 for disc herniation, 0.86 for central canal stenosis and 0.81 for nerve roots compression. And the internal and external testing also showed almost substantially perfect agreement (Kappa value was 0.67–0.85) for the multi-task classification model, which further approved the performance of ResNet-50 in the diagnosis of the three spinal diseases (Su et al. 2022). The semantic segmentation network (BianqueNet) composed of three innovative modules also achieved high-precision in the evaluating of lumbar intervertebral disc degeneration (IVDD), which diagnosed and quantified the IVDD accurately and efficiently on the T2-MRI scans (Zheng et al. 2022). The symptoms in lumbar spondylolisthesis (LS) were not obvious in the early stages of LS, which usually led to severe disease progress without identifying. Hence, advanced treatment mechanisms were required to implement for diagnosis of LS, which was crucial in terms of early diagnosis, rehabilitation, and treatment planning. A transfer learning based MobileNet CNN model was developed with 2707 lumbar X-rays, which could extract the ROIs via Yolov3 and classify the images as spondylolisthesis or normal. And the model reached the testing accuracy of 0.99, sensitivity of 0.98 and the specificity of 0.99, whose performance encouragingly stated that the model could be used in outpatient clinics where any experts were not present (Varcin et al. 2021). And in our previous study, we also trained the DL algorithms Faster RCNN and RetinaNet with 1596 lumbar lateral X-rays of LS patients from three hospitals. Finally, the Faster RCNN got the better performance in LS detection (0.93 of precision, recall and F1-score), which was better than medical group (Zhang et al. 2023).

3.2.5 Bone tumor and bone age assessment

Bone tumor and bone age assessment also could not be separated from the imaging diagnosis, which might require doctors with more experiments in radiological interpreting. Chianca et al. extracted the features of bone tumor and created a ML classifier by tenfold iterations and cross-validation. The classifier could label the bone tumors as benign or malignant (2-label classification), and benign, primary malignant or metastases (3-label classification), which obtained 0.94 accuracy in the detection of bone tumor and provided significant help for clinical diagnosis (Chianca et al. 2021). Liu et al. also proposed

a multi-model weighted fusion framework (WFF) for benign and malignant diagnosis of spinal tumors based on MRI scans and age information. With the import of reference age information, the accuracy of WFF in the recognition of benign and malignant tumors on MRI scans was higher than that of three orthopedists (0.82 versus 0.68, 0.73 and 0.63) (Liu et al. 2022a, b). The lesions of low-grade or high-grade cartilaginous bone tumors on MRI scans were also correctly classified by the AdaboostM1 algorithm (with accuracy and AUROC of 0.85), whose performance was no significant difference compared with the radiologist (Gitto et al. 2020). Even in the confirmation of cancer bone metastasis, AI also had a place in the prediction and diagnosis. Zhao et al. developed a DL model DNN with 12,222 cases of 99mTc-MDP bone scintigraphy. The model demonstrated a considerable diagnostic performance of bone metastasis detection, 0.98 AUROC for breast cancer, 0.95 for prostate cancer, 0.95 for lung cancer and 0.97 for other cancers, which represented comparable performance to that of individually classifying by human physicians. The further AI-consulted interpretation also improved human diagnostic sensitivity and accuracy (Zhao et al. 2020). AI could support imaging-driven diagnosis of musculoskeletal malignancies, however, the data quality and quantity needed further increasing to achieve better performance. A systematic, structured data collection and the establishment of national or international networks to obtain substantial datasets were important points for the critical advancement (Hinterwimmer et al. 2022). In our own study, we also realized the automatic detection and segmentation of lung cancer bone metastases based on the training of DL algorithm 3D UNet with the spinal CT from 126 patients. The model finally achieved a detection sensitivity of 0.89 and a segmentation dice coefficient of 0.85 (Huo et al. 2023).

Bone age reflected the true growth and development status of children, which played a critical role in evaluating growth and endocrine disorders. Greulich and Pyle (GP) and Tanner-Whitehouse 3 (TW3) were the most prevalently used techniques for bone age assessment (BAA). In the procedures of BAA, the 20 bones (13 radius, ulna and short bones and 7 carpal bones) were identified with a categorized stage, then, the stage were replaced by a score to calculate the total score and transform into the bone age. However, errors in terms of months were still unavoidable and the doctor's subjectivity usually caused significant variation. And at least a time-consuming of 20 min was also required to complete the BAA manually (Roche et al. 1970). Although the conventional computer-aided detection system was adopted, the assessment still partly relied on manual interpretation, which imposed unavoidable inter- and intra-reviewer variability. To solve this issue, Zhou et al. established and validated an optimized TW3-AI BAA system based on a CNN with a database of 9059 clinical X-rays of the left hand. After training, the performance of TW3-AI model was highly consistent with human level. And the final accuracy of TW3-AI was better than the estimate of reviewers. Further study also revealed that manual interpreting of the male capitate, hamate, the first distal and fifth middle phalanx and female capitate, the trapezoid, and the third and fifth middle phalanx were most inconsistent, which were quite satisfying in AI model. Moreover, the average image processing time was 1.5 ± 0.2 s of the algorithm, which was significantly shorter than manual efficiency (Zhou et al. 2020a, b). The Radiological Society of North America (RSNA) Pediatric Bone Age Machine Learning Challenge in 2019 also solicited researchers to create an algorithm or model using ML techniques that would accurately determine bone age in a curated data set of pediatric hand X-rays. The mean absolute distance (MAD) in months was set as a primary evaluation measure, which was calculated by the mean of the absolute values of the difference between model estimates and the bone age reference standard. Processing with the database consisting of 14,236 hand X-rays (12,611 training set, 1425 validation set, 200 test set) available for participants, the best three algorithms achieved the MADs of 4.2, 4.4 and

4.5 months, respectively (Halabi et al. 2019). The summary of AI-diagnosis in other orthopedic diseases was shown in Table 5.

Except for these classical orthopedic diseases, AI also expressed a crucial role in the diagnosis of other atypical orthopedic problems, including the detection of shoulder pain (dislocation or periarticular calcification) (Grauhan et al. 2022), developmental dysplasia of hip(Park et al. 2021), patellar dysplasia (AI-aided assessing of insall-salvati index (ISI), caton-deschamps index (CDI) and Keerati index (KI)) (Ye et al. 2020). And there was also study built the database of 1,023 dorsoplantar X-rays and trained a CNN framework to realize the automatically labeling and calculating of the first–second intermetatarsal angle, hallux valgus angle, hallux interphalangeal angle and distal metatarsal articular angle, which got the standard deviation ranged from 2.25 to 4.47° compared with the reference standard. The results promoted the clinical detection and severity evaluation of the hallux valgus (Li et al. 2022). In addition, for the common people who has not yet been diagnosed from orthopedic diseases (such as osteoporotic fracture), the AI-predictor could export the risk population from the analysis of health examination data, providing early warning to the people concerned (Gorelik and Gyftopoulos 2020; Villamor et al. 2020; Ferizi et al. 2019). In summary, the application of AI in orthopedic diseases diagnosis significantly improved the accuracy and efficiency, helping clinicians with reduction of misdiagnosis and missed diagnosis as well as workload. Although, some scholars also expressed concern about the algorithmic error in clinical diagnosis (Langerhuizen et al. 2020), but with the development of larger database and superior algorithm updating, this worry could be solved perfectly.

4 AI in orthopedic treatment

A surgery was the primary and effective treatment for most orthopedic diseases, such as bone fracture, locomotor system injury and bone tumor. The intelligent surgical robots cut a conspicuous figure in the field of orthopedic surgery, which was also the representative application in the field of intelligent medicine (Zhewei 2020). Since 1980s, the first generation of intelligent surgical robots named PUMA was invented, which could help surgeons with highly difficulty surgeries (Drake et al. 1991). This was the first attempt to apply the robot-assisted surgical procedure in surgery. With the improving of precision and stability of the mechanical arms, the surgical robots developed rapidly with increasing attention in these years. Da Vinci robot had been proposed and applied in multidisciplinary surgeries with remarkable outcomes (Tamhankar et al. 2020; Lippross et al. 2020). The orthopedic proprietary robots such as Mako (Stryker Corporation) and Ti-Robot (Beijing Jishuitan Hospital) intelligently realized the surgical tactile feedback, path planning, intraoperative warning and navigation, which enormously improved orthopedic surgery with higher accuracy, efficiency and security (Zhang et al. 2022; Han et al. 2019; Fan et al. 2020a, b). However, a wrong cognition had confused the general public and even lots of professional orthopedists for a long time, who believed that the surgical robots were also the embodiment of AI in medicine. Hence, we thought it was necessary to clarify in this review that current surgical robots could not be called AI-robot, whose functions totally depended on manual operation, rather than the independent judgment and decision making based on algorithms. As lacking the intelligent and automatic elements, it would be better to regard them as a more flexible scalpel or more advanced surgical mechanical arms, which could achieve difficult operations in traditional surgeries flexibility and precisely with the flexible

Table 5 The summary of AI-diagnosis in other orthopedic diseases

Application	Application type	Database		Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	References
		Source	Size	Type							
Osteoporosis	Detection	Self-built database	1665	CT	Cropping, 96×96 pixels resizing Data augmentation: noise adding, shifting, rotating	BMD-CNN	DL	Chainer	Python, Pydicom, Core i9-7900X CPU, NVIDIA Quadro P6000 GPU, 128 GB RAM	Supervised	Yasaka et al. (2020)
Arthritis											
	Detection	Self-built database	89	X-ray	N/A	U-NET	DL	N/A	N/A	Supervised	Liu et al. (2019)
Hip	Detection	Self-built database	434	X-ray	Cropping, 224×224 pixels resizing Data augmentation: translating, flipping, scaling	VGG-16	DL	Keras	MATLAB, Intel Core ™ i7-9750H, 16G, 512 GeForce RTX2060 GPU	Supervised	Ureten et al. (2020)
Finger	Classification	Self-built database	1426	X-ray	Cropping Data augmentation: rotating, translating, flipping	VGG-16	DL	Keras	MATLAB, 512 GeForce RTX2060 GPU	Supervised	Ureten and Maras (2022)

Table 5 (continued)

Application	Application type	Database			Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/unsupervised learning	References
		Source	Size	Type							
Cervical vertebra	Classification	Self-built database	331	X-ray	Cropping, 224×224 pixels resizing	VGG-16/19, Resnet-101, DenseNet-201	DL	Keras	MATLAB, Intel Core ™ i7-9750H, 16G, 512 GeForce RTX2060 GPU	Supervised	Maras et al. (2022)
Knee	Classification	Open-source database	4504	X-ray	Cropping, 500×500 pixels resizing Data augmentation: rotating, translating	DenseNet	DL	Tensor flow	MATLAB	Supervised	Norman et al. (2019)
Ligaments and cartilage injuries	Detection	Self-built database	1123	MRI	256×256 pixels resizing Data augmentation: flipping	Faster RCNN	DL	N/A	N/A	Supervised	Roblot et al. (2019)
Meniscus	Detection/classification	Self-built database	1048	MRI	Normalizing between 0 and 1 (pixel value/255)	AlexNet	DL	PyTorch	NVIDIA GeForce RTX 2080TI	Supervised	Shin et al. (2022)

Table 5 (continued)

Application	Application type	Database			Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	References
		Source	Size	Type							
ACL	Detection	Self-built and open-source databases	19,765	MRI	Cropping, 64×64 pixels resizing	CNN	DL	Keras	MATLAB, NVIDIA V100 GPU	Supervised	Tran et al. (2022)
	Detection	Self-built database	917	MRI	Cropping, 75×75 pixels resizing Data augmentation: real-time in model running	ResNet-14	DL	Keras	Python, Tesla K80 processor with 2496 CUDA cores and 12 GB RAM	Supervised	Awan et al. (2021)
Spinal diseases											
Scoliosis	Detection	Self-built database	986	X-ray	256×512 pixels resizing. Data augmentation: flipping, rotating, brightness, contrast	FCN	DL	Keras	Python, NVIDIA Titan Xp GPU	Supervised	Galbusera et al (2019)
	Detection	Self-built database	3240	Back image	N/A	DLA	DL	Caffe	MATLAB, LabelImg, 4Titan Xp GPU	Supervised	Yang et al. (2019)

Table 5 (continued)

Application	Application type	Database			Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	References
		Source	Size	Type							
Intervertebral disc disorder	Classification	Self-built database	15,254	MRI	Data augmentation: rotating, cropping	ResNet-50	DL	Pytorch	MRIcro, RTX 2080Ti GPU	Supervised	Su et al. (2022)
	Segmentation	Self-built database	1051	MRI	320×320 or 512×512 pixels resizing	BianqueNet	DL	N/A	LabelMe	Supervised	Zheng et al. (2022)
LS	Detection	Self-built database	600	X-ray	400×600 pixels resizing Data augmentation: flipping, zooming, brightness	MobileNet CNN	DL	N/A	N/A	Supervised	Varcin et al. (2021)
	Detection	Self-built database	1596	X-ray	800×800 pixels resizing Data augmentation: fliping	Faster R-CNN, RetinaNet	DL	Pytorch	LabelMe, Python, NVIDIA V100 GPU, 32 GB RAM	Supervised	Zhang et al. (2023)

Table 5 (continued)

Application	Application type	Database				Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/unsupervised learning	References
		Source	Size	Type								
Bone tumor	Classification	Self-built database	2150	MRI	112×112 pixels resizing	WFF	DL	N/A	LabelMe, Intel E5-2640 CPU and an NVIDIA GTX1080Ti GPU	Supervised	Liu et al. (2022a, b)
	Detection	Self-built database	13,477	BS	256×768 pixels resizing	ResNet-50	DL	N/A	N/A	Supervised	Gitto et al. (2020)
	Segmentation	Self-built database	126	CT	Cropping Data augmentation: rotating, scaling, Gaussian noise, Gaussian blur, brightness, contrast, mirroring	3D U-Net	DL	PaddlePaddle	LabelMe, NVIDIA GeForce GTX 1080Ti	Supervised	Huo et al. (2023)

Table 5 (continued)

Application	Application type	Database				Algorithm	Algorithm type	Framework	Tools	Supervised/unsupervised learning	References
		Source	Size	Type	Pre-process						
Bone age	Detection	Self-built database	9059	X-ray	256×256 pixels resizing Data augmentation: rotating, shifting, scaling	TW3-AI	DL	Tensor flow	Python	Supervised	Zhou et al. (2020a, b)
	Detection	Open-source database	14,236	X-ray	500×500 pixels resizing	Inception V3	DL	N/A	N/A	Supervised	Halabi et al. (2019)

It was not stated in the literature

ACL anterior cruciate ligament, *LS* lumbar spondylolisthesis, *BS* bone scintigraphy, *N/A* not available

and fine cutting-tool and the convenient control-panel. The confusion between surgical robots and AI might be caused by the excessive function publicity and highly subjective expectations in the medical market. Moreover, as a cutting-edge technology, the related conception or surgical robots was still immature, which also led to confusion. But the surgical robots based on computer system still had the potential to realize the total conception of AI and its final developed form must include the complete combination with AI. Only at this stage, the automatic and intelligent AI-surgical robots could be truly realized.

As for the real participant of AI in the treatment of orthopedic diseases, the most common application was the AI-aided medical decision making, which had been extensively applied in the treatment protocol designation. Traditionally, the surgery for patients depended on the condition of illness and was also inevitably infected by the orthopedists' subjective experience, which possibly led to different surgical planning for one patient (Kraemer et al. 2016). Moreover, on account of the individual differences among patients, the most appropriate plan and relevant surgical risks were difficult to precisely confirm. The participant of AI could be a reliable way to cover the shortage and provide a scientific reference with comprehensive consideration for medical decision making (Shortliffe and Sepulveda 2018). And the appearance of AI-surgical risk prediction calculator had achieved satisfying results. For instance, apart from patients with severe neurological deficits, it was still not clear whether surgical or conservative treatment for lumbar disc herniations was more effective for the patients. Wirries et al. collected the clinical data (including treatment planning and clinical outcomes) of 60 orthopedic patients with lumbar disc herniations to develop a DL algorithm. After the model fitting and a tenfold cross-validation, it could predict the possible 6 month-later outcomes for patients with treatment of lumbar disc herniations, which precisely got a 0.34 difference compared with real situation (Wirries et al. 2020). Surgeries of pelvic bone tumors were very challenging due to the complexity of anatomical structures and the irregular bone shape. To solve the challenges, Du et al. applied ML-assisted CT/MRI image fusion technique and built a personalized 3D model for preoperative plan making, such as the operation selecting and tumor margin assessment (Du et al. 2020). Moreover, DL model also provided personalized prediction for pelvic fracture patients in the extraperitoneal hematoma volumes quantitative visualization and measurement, which was helpful for decision making and potential outcome forecasting (Dreizin et al. 2020). Furtherly, based on the open-source database ACS-NSQIP, Bertsimas et al. presented an original Optimal Classification Trees (OCT) model upon ML algorithm named POTTER to calculate surgical complications in terms of mortality, morbidity, sepsis as well as infection in the period of 30 days postoperatively, whose accuracy and stability were higher than that in traditional American Society of Anesthesiologists (ASA), Emergency Surgery Score (ESS), and ACS-NSQIP calculators (Bertsimas et al. 2018). According to the clinical practicability and popularity of POTTER, 1 year later the authors subsequently created the "My Surgical Risk" calculator based on the database of more than 50,000 patients, which could further predict the complications in 24 months after operation, including wound condition, sepsis, venous thrombosis, intensive care unit admission, mechanical ventilation requirements, neurologic and cardiovascular complications, and death. The AUROC of the model arrived 0.94, which could be advice and reference for doctors to minimize the surgical risks (Bihorac et al. 2019). In addition, the infection risk of tibial shaft fractures after surgery (Machine Learning Consortium 2021), the risk of bone cement leakage in percutaneous vertebroplasty (Li et al. 2021a, b), the relapsed risk of kyphoplasty in osteoporotic vertebral compression fractures and the reherniation rate following lumbar microdiscectomy (Dong et al. 2022; Harada et al. 2021), the risk of femoral head osteonecrosis after internal fixation of femoral neck fracture (Zhu

et al. 2020), the length-prediction of hospital stays following femoral neck fracture (Zhong et al. 2021), the individual difficulty-prediction of percutaneous endoscopic transforaminal discectomy at L5/S1 level (Fan et al. 2020a, b), and the possible adverse clinical outcomes of sarcopenia (Pickhardt et al. 2022) were well predicted with the assistance of AI. The summary of AI in orthopedic treatment was shown in Table 6.

In summary, with the property of predicting risk and complications, the application of AI algorithm could acquire whether patients would benefit from a surgical procedure or a conservative treatment at initial medical phase. It could help to avoid the negative results and avert the unnecessarily invasive and harmful injury for patients. Besides, for the patients required a surgery, AI could also provide powerful assistance for the individually optimal surgical decision making, when there were controversially different treatment selections.

5 AI in orthopedic rehabilitation

For the orthopedic surgery such as the internal fixation of fractures, the most important three items were intraoperative reduction, fixation and postoperative rehabilitation. A feasible and effective rehabilitation was crucial to patients. However, due to the impossibility of one-to-one full-guidance functional training during the hospitalization and the lack of professional guidance after discharge, the effect of rehabilitation exercise was very limited. For this problem, there were also many studies applied AI technology in the postoperative rehabilitation to promote patient recovery. The relevant studies were mostly in the field of rehabilitation exercise movement recognition and evaluation, as well as medical information collection and analysis. For instance, the routine rehabilitation treatments for postoperative motor dysfunction were usually unsatisfying. The traditional assessment was quite subjective, which mostly depended on the experience and expertise of clinicians, lacking the standardization and precision. Hence, it might be inconveniently to track the valid functional changes during the rehabilitation process. The emerging intelligent rehabilitation platform provided objective and accurate functional assessment for patients, which also promoted the informationalized and standardized improvement of clinical guidance (Huo et al. 2021). With the enhancement of DL algorithm, automatic high-level feature extraction had been applied in optimizing the performance of human motion recognition (HAR). Moreover, in the healthcare and eldercare, DL were also applied in the intelligent sensors based on HAR to analyze the health data of users (Nafea et al. 2021). Combined with DL algorithm, the depth camera and inertial sensors could capture and classify the video actions in HAR, which realized the recovering training monitoring for patients during orthopedic rehabilitation (Xing et al. 2020). Similarly, the feature representation and data augmentation based on wearable IMU sensor data and a deep LSTM neural network also achieved the human activity classification, which could monitor the standard degree of rehabilitation exercise movements (Steven and Han 2018). The depth video sensor based life-logging HAR system for elderly care in smart indoor environments was also proposed to recognize the activity and generate the life logs, which could directly monitor healthcare problems for elderly people, or examine the indoor activities of people at home, office and hospital (Jalal et al. 2014). And there were also orthopedic rehabilitation robots assisting patients with strength training and functional rehabilitation, which combined with the AI-sensors to collect and analyze the rehabilitative data. They could automatically provide the passive, active and assisted exercising (Padilla-Castaneda et al. 2018). For example, the

Table 6 The summary of AI in orthopedic treatment

Application	Application type	Database			Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	Referenes
		Source	Size	Type							
Lumbar disc herniations	Decision-making	Sefl-built database	60	Text	Encoding of categorical variables	DNN	DL	Keras	Python, LabelEn-coder	Supervised	Wirries et al. (2020)
Extraperi-toneal hematoma	Decision-making segmenta-tion	Sefl-built database	253	Text/CT	Assignment of variables	FCN	DL	N/A	N/A	Supervised	Dreizin et al. (2020)
Surgical outcome prediction	Decision-making	Open-source database	382,960	Text	Imputing of missing values	OCT	ML	N/A	N/A	Supervised	Bertsimas et al. (2018)
Surgical outcome prediction	Decision-making	Open-source database	51,457	Text	Errors and outliers removing	TRIPOD	ML	N/A	N/A	Supervised	Bihorac et al. (2019)
Bone cement leakage	Decision-making	Sefl-built database	385	Text	Variables logistic regression analysis	LR, GBM, XGB, RF, DT, MLP	ML	N/A	N/A	Supervised	Li et al. (2021a, b)
Recollapse of vertebral compres-sion fractures	Decision-making	Sefl-built database	346	Text	Variables logistic regression analysis	SVM	ML	N/A	Python	Supervised	Dong et al. (2022),
Re-herniation after lumbar microdis-cectomy	Decision-making	Sefl-built database	2630	Text	k-nearest neighbors imputation	XGBoost	ML	Streamlit	Python	Supervised	Harada et al. (2021)

Table 6 (continued)

Application	Application type	Database				Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	Referenes
		Source	Size	Type	Pre-process						
Prediction of hospital stay of PNF	Decision-making	Sefl-built database	182	Text	Encoding of variables, case deletion and mean filling of missing values	BP, SVM	ML	N/A	N/A	Supervised	Zhong et al. (2021)

It was not stated in the literature

LR logistic regression, *GBM* gradient boosting machine, *XGB* extreme gradient boosting, *RF* random forest, *DT* decision tree, *MLP* multilayer perceptron, *SVM* support vector machine, *BP* back propagation, *N/A* not available

training assisted by robot after proximal humeral fracture (Kroger et al. 2021). In summary, the application of AI in orthopedic rehabilitation improved the rehabilitation training and clinical outcomes, brought the traditional rehabilitation medicine a creative approach. The summary of AI in orthopedic rehabilitation was shown in Table 7.

6 Conclusion and outlook

AI had demonstrated a promising future in the application among orthopedic diseases in terms of severity evaluation, triage, diagnosis, treatment and rehabilitation. It would be comprehensive and scientific intelligent-assistant for clinicians to avoid clinical risk and design an individual medical plan for the sake of optimal remedies. The researches of AI in medicine had drawn an increasing attention to researchers, but there was also a lack of uniform industry standards, with which the relevant studies could be constructed more valuable. With our own studies on intelligent medicine and orthopedic AI, there were several research points we thought needing to be summarized: (1) Database and algorithm. The feature extraction, generalization and summarization of the database were the essence of orthopedic AI. And large database was recommended for the algorithm learning, training and better performance. However, the structural innovation of algorithms was equally important. To achieve optimum working conditions, it required the engineers to further modify algorithm parameters (even design a new algorithm) according to the structure and characteristics of medical data. Plenty of current studies ignored the algorithm innovation and excessively pursued the large database. The direct application of existing algorithms in the medical analysis without any modification might be adverse effect for final study results. Hence, during the orthopedic AI research, both of database size and the algorithm-weight needed to be addressed equally. (2) The division of database. In the study of orthopedic AI, the database would be commonly divided into 3 datasets, training dataset (for the data feature extraction and learning), validation dataset (for the algorithm parameters adjusting to improve performance) and testing dataset (for the evaluation of the algorithm performance). The proportion of division could be flexibly set around the approximate standard of 6:2:2 or 7:2:1 to achieve the optimal results. Of course, the setting of validation dataset could also be omitted according to the size of total database, and the recommended proportion of training dataset and testing dataset approximately was 6:4, 7:3 or 8:2. No matter how to set the proportion, the training dataset should be a majority, which ensured the algorithm could learn as many data features as possible to avoid the diagnostic errors. (3) The data sources and labeling. The data sources could be self-collection establishment or existing database publicly available on the web. A multi-center database (cross time and space, national or international) was also recommended, with which the internal and external testing could be realized to further verify the universality and generalization of the algorithm in different data environments. The labeling process was regarded as the most time-consuming work in the orthopedic AI study, which was also the most crucial procedure. It directly determined the quality of training dataset and training effect. Hence, labeling process should be operated with extra care by the senior and experienced orthopedists. For example, in the AI-diagnosis on medical images, a precise and professional outlining of lesion was better than the simple box notation. And the labeling tools such as labelImg (https://github.com/tzutalin/LabelImg) and labelme (https://github.com/wkentaro/labelme) were recommended. (4) Overfitting and underfitting. When the

Table 7 The summary of AI in orthopedic rehabilitation

Application	Application type	Database			Pre-process	Algorithm	Algorithm type	Framework	Tools	Supervised/ unsupervised learning	References
		Source	Size	Type							
Human activity recognition	Computer vision	Open-source database	1,098,209	Wearable sensor data	Series data segmentation	CNN-BiL-STM	DL	Keras	Python, E5-2686 v4 Xeon CPU, 32 GB RAM, NVIDIA V100	Unsupervised	Nafea et al. (2021)
Human activity recognition	Computer vision	Self-built and open-source database	5405	Wearable sensor data	Data augmentation: local average (G=4)	LSTM	DL	N/A	N/A	Unsupervised	Steven and Han (2018)
Human activity recognition	Computer vision	Self-built database	4800	Video clips	Depth silhouette preprocessing converting	Hidden Markov model	ML	N/A	Open NI/ NITE	Supervised	Jalal et al. (2014)

It was not stated in the literature

N/A not available

Fig. 3 Confusion matrix. TP: The real situation of the target was positive, and the predicted result was also positive; FN: The real situation of the target was positive, but the predicted result was negative; FP: The real situation of the target was negative, but the predicted result was positive; TN: The real situation of the target was negative, and the predicted result was also negative

		Prediction	
		Positive	Negative
Real situation	Positive	True positive（TP）	False negative（FN）
	Negative	False positive（FP）	True negative（TN）

database size was limited but the model structure was overcomplex, the algorithm was easy to appear an overfitting (the loss was small in the training dataset, but abnormally high in the verification or testing dataset), which meant the model was hypersensitive. On the contrary, if the algorithm got a large loss in both training and testing dataset, it was called underfitting, which could be attributed to the weak algorithm structure. Both overfitting and underfitting would cause a poor performance. For the algorithm overfitting, the data cleaning and modification to reduce the noise and errors, simplifying the model to limit its computational power, or further expansion of training dataset could solve it well. For the algorithm underfitting, improving and modifying the model to further fit the training database could be beneficial. Both two items should be avoided in the orthopedic AI study. (5) The performance indexes. The relevant performance indexes would be calculated based on the result of prediction (in the form of confusion matrix), as shown in Fig. 3.

The indexes such as (1) accuracy, (2) sensitivity, (3) missed diagnosis rate, (4) specificity, (5) misdiagnosis rate, (6) PPV, (7) NPV, (8) ROC, (9) AUROC, (10) P-R curve, (11) F1 score, (12) AP and mean AP (mAP) were applied to describe the results in most target detection and classification studies of orthopedic AI. For instance:

(1) Accuracy: the proportion of the targets that were predicted correctly to the total targets.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

(2) Sensitivity: the proportion of positive targets that were correctly diagnosed as positive (also known as recall).

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

(3) Missed diagnosis rate: the proportion of positive targets who were wrongly diagnosed as negative.

$$Misseddiagnosisrate = 1 - \frac{TP}{TP + FN} \tag{3}$$

(4) Specificity: the proportion of negative targets that were correctly diagnosed as negative.

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

(5) Misdiagnosis rate: the proportion of negative targets who were wrongly diagnosed as positive.

$$Misdiagnosisrate = 1 - \frac{TN}{TN + FP} \tag{5}$$

(6) PPV: the proportion of targets diagnosed as positive were indeed positive (also known as precision).

$$PPV = \frac{TP}{TP + FP} \tag{6}$$

(7) NPV: the proportion of targets diagnosed as negative were indeed negative.

$$NPV = \frac{TN}{TN + FN} \tag{7}$$

(8) ROC: a curve reflected the relationship between TP and FP, with FP as the horizontal coordinate and TP as the vertical coordinate.
(9) AUROC: the area under the receiver operating characteristic curve (ROC). The larger value indicated a better algorithm performance.
(10) P-R curve: a curve reflected the relationship between precision and recall, with the recall as horizontal coordinate and precision as the vertical coordinate.
(11) F1 score: the balance point of precision and recall. F1 score was an important index used to evaluate the performance of the algorithm, which took into account both the precision and recall of the algorithm, and could be regarded as the collaborative average of them. The larger value indicated a better algorithm performance.

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{8}$$

(12) AP: an index for one classification of the targets, which actually was the area under the P-R curve. mAP: the average of all the APs of each classification of the targets. Both, the larger value indicated a better algorithm performance.

Commonly, the accuracy, sensitivity, missed diagnosis rate, specificity, misdiagnosis rate, PPV and NPV directly reflected the recognition ability of the algorithm after training, which were the main indexes to evaluate its clinical performance. In the specific application scenarios such as performance assessment of orthopedic AI-diagnosis, they deserved more attention. The ROC, AUROC, P-R curve, F1 score, AP and mAP were used to comprehensively evaluate the model's property and compare the different algorithms. They represented the learning ability and superiority of the algorithm. In the algorithm study such as model improvement, they deserved more inclining. (6) Patients'

privacy. The privacy concerns also needed attention. Before the study, the patient information on the data required a thorough cleaning and desensitization.

Moreover, despite that AI has brought surprising improvements to the management of orthopedic diseases, it was just an assistance instead of complete human-replacement in current stage. And the merits and demerits also came together. For the merits: (1) With the better performance of AI than human level, the clinical failures such as underestimated illness states, wrong triages, misdiagnosis, missed diagnosis, risky treatment plans and inappropriate rehabilitation situations were largely avoided, which further benefited the patient security. (2) The credibility of clinical decision making was further enhanced. (3) The clinical workflow and efficiency were accelerated, which promoted the medical resources rearrangement. (4) The clinical burden was reduced, which improved the working environment for doctors. (5) The continuous learning of junior doctors was also realized with the accurate AI-guidance. (6) The less developed areas and primary hospitals lacking medical experts could be benefited from professional help with the assistance of AI. (7) The diseases could also be automatically graded according to the severity and treatment difficulty, and patients would be treated in order of priority, which gradually realized the medical reform of hierarchical diagnosis and treatment system. These could be seemed as the advantages of AI in medicine. While profiting by the conveniences of AI, the relevant demerits should not be ignored, which need more attention to avoid the risks. For the demerits: (1) According to the immature algorithm structure and insufficient data availability, the possibly underlying errors of algorithm still existed, which required human supervision and amendment. (2) The standardized database was lacking. Owing to the diversity of data from different hospitals and countries, and the inconsistent labeling manners of different studies, the universality and generalization of the algorithm needed to be further confirmed in different data environments. (3) Current AI algorithm in medicine were mostly established by professional engineers based on existing models, and few medical experts were involved during the process. Hence, the model might lack favorable consistence with the characteristics of medical data, which would cause unknown drawbacks and risks. (4) The medical AI also lacked transparency and interpretability, which mostly relied on the generalization and summarization of data. There was no way to know how the medical predictions were generated. (5) Most of the medical AI was still in the stage of retrospective research and had not been widely applied in clinical practice. More clinical evidence and prospective review were required, such as systematic commissioning, auditing, stability test, extensive simulation and validation. (6) Although AI owned the excellent ability of computing power, storage capacity, deep searching and fast learning, there still some inevitable drawbacks such as the issue of robustness. In the face of systemic disturbances, AI might not perform as robust as human logic. (7) The responsibility assignment of AI-medical negligence was not clear yet, which was prone to potential medical disputes. (8) The AI-medical insurance charging measures, medical policies and ethics were still undefined. (9) The over-reliance on external AI-assistance would also be adverse for the cultivation of doctors' clinical ability. (10) Potential risk of patient privacy disclosure. While facing the enhancement of AI in medicine, these disadvantages needed more noticing. A rational attitude was also required to obtain the profits and avoid the harms. We believed with the rapid development and updating of AI technology these worries would not take long to be solved. The future of AI in medicine and orthopedics remained bright and promising.

Declarations

Competing interests The authors declare no competing interests.

Conflict of interest The authors declare that there are no conflicts of interest relevant to this article.

Ethical approval The study was reviewed and approved by the ethics committee of Wuhan Union Hospital, Tongji Medical College, Huazhong University of Science and Technology.

References

Aghnia FN, Lai JY, Wang JC et al (2021) Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques. Injury 52(3):616–624

Ashkani ES, Mojahed YR, Bhimani R, Kerkhoffs GM et al (2022) Detection of ankle fractures using deep learning algorithms. Foot Ankle Surg S1268–7731(22):00102–00103

Awan MJ, Rahim M, Salim N, Mohammed M et al (2021) Efficient detection of knee anterior cruciate ligament from magnetic resonance imaging using deep learning approach. Diagnostics 11(1):105

Bayramoglu N, Nieminen MT, Saarakkala S (2021) Automated detection of patellofemoral osteoarthritis from knee lateral view radiographs using deep learning: data from the multicenter osteoarthritis study (MOST). Osteoarthritis Cartilage 29(10):1432–1447

Bertsimas D, Dunn J, Velmahos GC, Kaafarani H (2018) Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based predictive optimal trees in emergency surgery risk (POTTER) calculator. Ann Surg 268(4):574–583

Bihorac A, Ozrazgat T, Ebadi A, Motaei A, Madkour M, Pardalos M et al (2019) My surgery risk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. Ann Surg 269(4):652–662

Candel B, Raven W, Lameijer H, Thijssen W, Termorshuizen F, Boerma C et al (2022) The effect of treatment and clinical course during emergency department stay on severity scoring and predicted mortality risk in intensive care patients. Crit Care 26(1):112

Chen W, Liu X, Li K, Luo Y, Bai S, Wu J et al (2022) A deep-learning model for identifying fresh vertebral compression fractures on digital radiography. Eur Radiol 32(3):1496–1505

Cheng CT, Wang Y, Chen HW, Hsiao PM, Yeh CN, Hsieh CH et al (2021) A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs. Nat Commun 12(1):1066

Cheng R, Crouzier M, Hug F, Tucker K, Juneau P, Mccreedy E et al (2022) Automatic quadriceps and patellae segmentation of MRI with cascaded U(2) -Net and SASSNet deep learning model. Med Phys 49(1):443–460

Chianca V, Cuocolo R, Gitto S, Albano D, Merli I, Badalyan J et al (2021) Radiomic machine learning classifiers in spine bone tumors: a multi-software multi-scanner study. Eur J Radiol 137:109586

Choi JW, Cho YJ, Lee S, Lee J, Lee S, Choi YH et al (2020) Using a dual-input convolutional neural network for automated detection of pediatric supracondylar fracture on conventional radiography. Invest Radiol 55(2):101–110

Choi JW, Cho YJ, Ha JY, Lee YY, Koh SY, Seo JY et al (2022) Deep learning-assisted diagnosis of pediatric skull fractures on plain radiographs. Korean J Radiol 23(3):343–354

Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP et al (2018) Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 89(4):468–473

Derkatch S, Kirby C, Kimelman D, Jozani MJ, Davidson JM, Leslie WD (2019) Identification of vertebral fractures by convolutional neural networks to predict nonvertebral and hip fractures: a registry-based cohort study of dual X-ray absorptiometry. Radiology 293(2):405–411

Dipnall JF, Lu J, Gabbe BJ, Cosic F, Edwards E, Page R et al (2022) Comparison of state-of-the-art machine and deep learning algorithms to classify proximal humeral fractures using radiology text. Eur J Radiol 153:110366

Dong ST, Zhu J, Yang H, Huang G, Zhao C, Yuan B (2022) Development and internal validation of supervised machine learning algorithm for predicting the risk of recollapse following minimally invasive kyphoplasty in osteoporotic vertebral compression fractures. Front Public Health 10:874672

Drake JM, Joy M, Goldenberg A, Kreindler D (1991) Computer- and robot-assisted resection of thalamic astrocytomas in children. Neurosurgery 29(1):27–33

Dreizin D, Zhou Y, Chen T, Li G, Yuille AL, Mclenithan A et al (2020) Deep learning-based quantitative visualization and measurement of extraperitoneal hematoma volumes in patients with pelvic fractures: potential role in personalized forecasting and decision support. J Trauma Acute Care Surg 88(3):425–433

Du X, Wei H, Li P, Yao WT (2020) Artificial intelligence (AI) assisted CT/MRI image fusion technique in preoperative evaluation of a pelvic bone osteosarcoma. Front Oncol 10:1209

Duron L, Ducarouge A, Gillibert A, Laine J, Allouche C, Cherel N et al (2021) Assessment of an AI aid in detection of adult appendicular skeletal fractures by emergency physicians and radiologists: a multicenter cross-sectional diagnostic study. Radiology 300(1):120–129

Dutt S, Sivaraman A, Savoy F, Rajalakshmi R (2020) Insights into the growing popularity of artificial intelligence in ophthalmology. Indian J Ophthalmol 68(7):1339–1346

Fan G, Liu H, Wang D, Feng C, Li Y, Yin B et al (2020a) Deep learning-based lumbosacral reconstruction for difficulty prediction of percutaneous endoscopic transforaminal discectomy at L5/S1 level: a retrospective cohort study. Int J Surg 82:162–169

Fan M, Liu Y, He D, Han X, Zhao J, Duan F et al (2020b) Improved accuracy of cervical spinal surgery with robot-assisted screw insertion: a prospective, randomized, controlled study. Spine 45(5):285–291

Ferizi U, Honig S, Chang G (2019) Artificial intelligence, osteoporosis and fragility fractures. Curr Opin Rheumatol 31(4):368–375

Fernandes M, Mendes R, Vieira SM, Leite F, Palos C, Johnson A et al (2020a) Risk of mortality and cardiopulmonary arrest in critical patients presenting to the emergency department using machine learning and natural language processing. PLoS ONE 15(4):e0230876

Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa J (2020b) Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 102:101762

Galbusera F, Niemeyer F, Wilke HJ, Bassani T, Casaroli G, Anania C et al (2019) Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. Eur Spine J 28(5):951–960

Gan K, Xu D, Lin Y, Shen Y, Zhang T, Hu K et al (2019) Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. Acta Orthop 90(4):394–400

Ganjali R, Golmakani R, Ebrahimi M, Eslami S, Bolvardi E (2020) Accuracy of the emergency department triage system using the emergency severity index for predicting patient outcome; a single center experience. Bull Emerg Trauma 8(2):115–120

Gao Y, Soh N, Liu N, Lim G, Ting D, Cheng LT et al (2023) Application of a deep learning algorithm in the detection of hip fractures. iScience 26(8):107350

Gitto S, Cuocolo R, Albano D, Chianca V, Messina C, Gambino A et al (2020) MRI radiomics-based machine-learning classification of bone chondrosarcoma. Eur J Radiol 128:109043

Gorelik N, Gyftopoulos S (2020) Applications of artificial intelligence in musculoskeletal imaging: from the request to the report. Can Assoc Radiol J 72(1):45–59

Grauhan NF, Niehues SM, Gaudin RA, Keller S, Vahldiek JL, Adams LC et al (2022) Deep learning for accurately recognizing common causes of shoulder pain on radiographs. Skeletal Radiol 51(2):355–362

Guly HR (2001) Diagnostic errors in an accident and emergency department. Emerg Med J 18(4):263–269

Halabi SS, Prevedello LM, Kalpathy Cramer J, Mamonov AB, Bilbily A, Cicero M et al (2019) The RSNA pediatric bone age machine learning challenge. Radiology 290(2):498–503

Han X, Tian W, Liu Y, Liu B, He D, Sun Y et al (2019) Safety and accuracy of robot-assisted versus fluoroscopy-assisted pedicle screw insertion in thoracolumbar spinal surgery: a prospective randomized controlled trial. J Neurosurg Spine 30:1–8

Harada GK, Siyaji ZK, Mallow GM, Hornung AL, Hassan F, Basques BA et al (2021) Artificial intelligence predicts disk re-herniation following lumbar microdiscectomy: development of the "RAD" risk profile. Eur Spine J 30(8):2167–2175

Hinterwimmer F, Consalvo S, Neumann J, Rueckert D, Von Eisenhart-Rothe R, Burgkart R (2022) Applications of machine learning for imaging-driven diagnosis of musculoskeletal malignancies-a scoping review. Eur Radiol 32(10):7173–7184

Hong WS, Haimovich AD, Taylor RA (2018) Predicting hospital admission at emergency department triage using machine learning. PLoS ONE 13(7):e0201016

Hu Y, Tang J, Zhao S, Li Y (2022) Deep learning-based multimodal 3 T MRI for the diagnosis of knee osteoarthritis. Comput Math Methods Med 2022:7643487

Huo CC, Zheng Y, Lu WW, Zhang TY, Wang DF, Xu DS et al (2021) Prospects for intelligent rehabilitation techniques to treat motor dysfunction. Neural Regen Res 16(2):264–269

Huo T, Xie Y, Fang Y, Wang Z, Liu P, Duan Y et al (2023) Deep learning-based algorithm improves radiologists' performance in lung cancer bone metastases detection on computed tomography. Front Oncol 131:125637

Hussain F, Cooper A, Carson-Stevens A, Donaldson L, Hibbert P, Hughes T et al (2019) Diagnostic error in the emergency department: learning from national patient safety incident report analysis. BMC Emerg Med 19(1):77

Jalal A, Kamal S, Kim D (2014) A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. Sensors 14(7):11735–11759

Jeong TS, Yee GT, Kim KG, Kim YJ, Lee SG, Kim WK (2022) Automatically diagnosing skull fractures using an object detection method and deep learning algorithm in plain radiography images. J Korean Neurosurg Soc 66:53

Jin L, Yang J, Kuang K, Ni B, Gao Y, Sun Y et al (2020) Deep-learning-assisted detection and segmentation of rib fractures from CT scans: development and validation of FracNet. EBioMedicine 62:103106

Kang DY, Cho KJ, Kwon O, Kwon JM, Jeon KH, Park H et al (2020) Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. Scand J Trauma Resusc Emerg Med 28(1):17

Kanis JA, Cooper C, Rizzoli R, Reginster JY (2019) European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int 30(1):3–44

Kaul V, Enslin S, Gross SA (2020) History of artificial intelligence in medicine. Gastrointest Endosc 92(4):807–812

Kim D, You S, So S, Lee J, Yook S, Jang DP et al (2018) A data-driven artificial intelligence model for remote triage in the prehospital environment. PLoS ONE 13(10):e0206006

Kitamura G (2020) Deep learning evaluation of pelvic radiographs for position, hardware presence and fracture detection. Eur J Radiol 130:109139

Klang E, Kummer BR, Dangayach NS, Zhong A, Kia MA, Timsina P et al (2021) Predicting adult neuroscience intensive care unit admission from emergency department triage using a retrospective, tabular-free text machine learning approach. Sci Rep 11(1):1381

Kraemer K, Cohen ME, Liu Y, Barnhart DC, Rangel SJ, Saito JM et al (2016) Development and evaluation of the American college of surgeons NSQIP pediatric surgical risk calculator. J Am Coll Surg 223(5):685–693

Kroger I, Nerz C, Schwickert L, Scholch S, Mussig JA, Studier-Fischer S et al (2021) Robot-assisted training after proximal humeral fracture: a randomised controlled multicentre intervention trial. Clin Rehabil 35(2):242–252

Kwon JM, Lee Y, Lee Y, Lee S, Park H, Park J (2018) Validation of deep-learning-based triage and acuity score using a large national dataset. PLoS ONE 13(10):e0205836

Langerhuizen D, Bulstra A, Janssen SJ, Ring D, Kerkhoffs G, Jaarsma RL et al (2020) Is deep learning on par with human observers for detection of radiographically visible and occult fractures of the scaphoid? Clin Orthop Relat Res 478(11):2653–2659

Lee SB, Kim DH, Kim T, Kang C, Lee SH, Jeong JH et al (2020) Emergency department triage early warning score (TREWS) predicts in-hospital mortality in the emergency department. Am J Emerg Med 38(2):203–210

Li W, Wang J, Liu W, Xu C, Li W, Zhang K et al (2021a) Machine learning applications for the prediction of bone cement leakage in percutaneous vertebroplasty. Front Public Health 9:812023

Li YC, Chen HH, Horng-Shing LH, Hondar WH, Chang MC, Chou PH (2021b) Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists? Clin Orthop Relat Res 479(7):1598–1612

Li T, Wang Y, Qu Y, Dong R, Kang M, Zhao J (2022) Feasibility study of hallux valgus measurement with a deep convolutional neural network based on landmark detection. Skeletal Radiol 51(6):1235–1247

Lind A, Akbarian E, Olsson S, Nasell H, Skoldenberg O, Razavian AS et al (2021) Artificial intelligence for the classification of fractures around the knee in adults according to the 2018 AO/OTA classification system. PLoS ONE 16(4):e0248809

Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S et al (2018) Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci USA 115(45):11591–11596

Lippross S, Junemann KP, Osmonov D, Peh S, Alkatout I, Finn J et al (2020) Robot assisted spinal surgery—a technical report on the use of DaVinci in orthopaedics. J Orthop 19:50–53

Liu J, Wang J, Ruan W, Lin C, Chen D (2019) Diagnostic and gradation model of osteoporosis based on improved deep U-net network. J Med Syst 44(1):15

Liu PR, Lu L, Zhang JY, Huo TT, Liu SX, Ye ZW (2021a) Application of artificial intelligence in medicine: an overview. Curr Med Sci 41(6):1105–1115

Liu PR, Zhang JY, Xue MD, Duan YY, Hu JL, Liu SX et al (2021b) Artificial intelligence to diagnose tibial plateau fractures: an intelligent assistant for orthopedic physicians. Curr Med Sci 41(6):1158–1164

Liu H, Jiao M, Yuan Y, Ouyang H, Liu J, Li Y et al (2022a) Benign and malignant diagnosis of spinal tumors based on deep learning and weighted fusion framework on MRI. Insights Imaging 13(1):87

Liu P, Lu L, Chen Y, Huo T, Xue M, Wang H et al (2022b) Artificial intelligence to detect the femoral intertrochanteric fracture: the arrival of the intelligent-medicine era. Front Bioeng Biotechnol 10:927926

Machine Learning Consortium (2021) A machine learning algorithm to identify patients with tibial shaft fractures at risk for infection after operative treatment. J Bone Joint Surg Am 103(6):532–540

Magneli M, Ling P, Gislen J, Fagrell J, Demir Y, Arverud ED et al (2023) Deep learning classification of shoulder fractures on plain radiographs of the humerus, scapula and clavicle. PLoS ONE 18(8):e0289808

Maras Y, Tokdemir G, Ureten K, Atalar E, Duran S, Maras H (2022) Diagnosis of osteoarthritic changes, loss of cervical lordosis, and disc space narrowing on cervical radiographs with deep learning methods. Jt Dis Relat Surg 33(1):93–101

Muehlematter UJ, Mannil M, Becker AS, Vokinger KN, Finkenstaedt T, Osterhoff G et al (2019) Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning. Eur Radiol 29(5):2207–2217

Mutasa S, Varada S, Goel A, Wong TT, Rasiej MJ (2020) Advanced deep learning techniques applied to automated femoral neck fracture detection and classification. J Digit Imaging 33(5):1209–1217

Muthukrishnan N, Maleki F, Ovens K, Reinhold C, Forghani B, Forghani R (2020) Brief history of artificial intelligence. Neuroimaging Clin N Am 30(4):393–399

Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C (2020) Artificial intelligence and orthopaedics: an introduction for clinicians. J Bone Joint Surg Am 102(9):830–840

Nafea O, Abdul W, Muhammad G, Alsulaiman M (2021) Sensor-based human activity recognition with spatio-temporal deep learning. Sensors 21(6):2141

Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY et al (2019) Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. Radiology 290(1):218–228

Nam JG, Kim M, Park J, Hwang EJ, Lee JH, Hong JH et al (2021) Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs. Eur Respir J 57(5):2003061

Norman B, Pedoia V, Noworolski A, Link TM, Majumdar S (2019) Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs. J Digit Imaging 32(3):471–477

Oakden-Rayner L, Gale W, Bonham TA, Lungren MP, Carneiro G, Bradley A et al (2022) 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 4(5):e351–e358

Oka K, Shiode R, Yoshii Y, Tanaka H, Iwahashi T, Murase T (2021) Artificial intelligence to diagnose distal radius fracture using biplane plain X-rays. J Orthop Surg Res 16(1):694

Olczak J, Emilson F, Razavian A, Antonsson T, Stark A, Gordon M (2021) Ankle fracture classification using deep learning: automating detailed AO foundation/orthopedic trauma association (AO/OTA) 2018 malleolar fracture identification reaches a high degree of correct classification. Acta Orthop 92(1):102–108

Ozkaya E, Topal FE, Bulut T, Gursoy M, Ozuysal M, Karakaya Z (2020) Evaluation of an artificial intelligence system for diagnosing scaphoid fracture on direct radiography. Eur J Trauma Emerg Surg 48(1):585–592

Padilla-Castaneda MA, Sotgiu E, Barsotti M, Frisoli A, Orsini P, Martiradonna A et al (2018) An orthopaedic robotic-assisted rehabilitation method of the forearm in virtual reality physiotherapy. J Healthc Eng 2018:7438609

Park HS, Jeon K, Cho YJ, Kim SW, Lee SB, Choi G et al (2021) Diagnostic performance of a new convolutional neural network algorithm for detecting developmental dysplasia of the hip on anteroposterior radiographs. Korean J Radiol 22(4):612–623

Patel SJ, Chamberlain DB, Chamberlain JM (2018) A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. Acad Emerg Med 25(12):1463–1470

Pickhardt PJ, Perez AA, Garrett JW, Graffy PM, Zea R, Summers RM (2022) Fully automated deep learning tool for sarcopenia assessment on CT: L1 versus L3 vertebral level muscle measurements for opportunistic prediction of adverse clinical outcomes. AJR Am J Roentgenol 218(1):124–131

Pinto A, Berritto D, Russo A, Riccitiello F, Caruso M, Belfiore MP et al (2018) Traumatic fractures in adults: missed diagnosis on plain radiographs in the emergency department. Acta Biomed 89(1S):111–123

Pranata YD, Wang KC, Wang JC, Idram I, Lai JY, Liu JW et al (2019) Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. Comput Methods Programs Biomed 171:27–37

Qiu X, Liu Z, Zhuang M, Cheng D, Zhu C, Zhang X (2021) Fusion of CNN1 and CNN2-based magnetic resonance image diagnosis of knee meniscus injury and a comparative analysis with computed tomography. Comput Methods Programs Biomed 211:106297

Raita Y, Goto T, Faridi MK, Brown D, Camargo CJ, Hasegawa K (2019) Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 23(1):64

Roblot V, Giret Y, Bou AM, Morillot C, Chassin X, Cotten A et al (2019) Artificial intelligence to diagnose meniscus tears on MRI. Diagn Interv Imaging 100(4):243–249

Roche AF, Rohmann CG, French NY, Davila GH (1970) Effect of training on replicability of assessments of skeletal maturity (Greulich-Pyle). Am J Roentgenol Radium Ther Nucl Med 108(3):511–515

Sato Y, Takegami Y, Asamoto T, Ono Y, Hidetoshi T, Goto R et al (2021) Artificial intelligence improves the accuracy of residents in the diagnosis of hip fractures: a multicenter study. BMC Musculoskelet Disord 22(1):407

Seah J, Tang C, Buchlak QD, Holt XG, Wardman JB, Aimoldin A et al (2021) Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. Lancet Digit Health 3(8):e496–e506

Shen L, Gao C, Hu S, Kang D, Zhang Z, Xia D et al (2023) Using artificial intelligence to diagnose osteoporotic vertebral fractures on plain radiographs. J Bone Miner Res 38(9):1278–1287

Shin H, Choi GS, Shon OJ, Kim GB, Chang MC (2022) Development of convolutional neural network model for diagnosing meniscus tear using magnetic resonance image. BMC Musculoskelet Disord 23(1):510

Shin H, Park D, Kim JK, Choi GS, Chang MC (2023) Development of convolutional neural network model for diagnosing osteochondral lesions of the talus using anteroposterior ankle radiographs. Medicine 102(19):e33796

Shortliffe EH, Sepulveda MJ (2018) Clinical decision support in the era of artificial intelligence. JAMA 320(21):2199–2200

Sim Y, Chung MJ, Kotter E, Yune S, Kim M, Do S et al (2020) Deep convolutional neural network-based software improves radiologist detection of malignant lung nodules on chest radiographs. Radiology 294(1):199–209

Sjoding MW, Taylor D, Motyka J, Lee E, Co I, Claar D et al (2021) Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation. Lancet Digit Health 3(6):e340–e348

Steven EO, Han DS (2018) Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network. Sensors 18(9):2892

Su ZH, Liu J, Yang MS, Chen ZY, You K, Shen J et al (2022) Automatic grading of disc herniation, central canal stenosis and nerve roots compression in lumbar magnetic resonance image diagnosis. Front Endocrinol 13:890371

Tamhankar AS, Chaturvedi H, Gautam G (2020) Beyond traditional frontiers: robotic total pelvic exenteration. Int Braz J Urol 46(6):1112

Tran A, Lassalle L, Zille P, Guillin R, Pluot E, Adam C et al (2022) Deep learning to detect anterior cruciate ligament tear on knee MRI: multi-continental external validation. Eur Radiol 32:8394

Tsai A, Grant PE, Warfield SK, Ou Y, Kleinman PK (2022) Deep learning of birth-related infant clavicle fractures: a potential virtual consultant for fracture dating. Pediatr Radiol 52(11):2206–2214

Ukai K, Rahman R, Yagi N, Hayashi K, Maruo A, Muratsu H et al (2021) Detecting pelvic fracture on 3D-CT using deep convolutional neural networks with multi-orientated slab images. Sci Rep 11(1):11716

Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N (2019) Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiol 48(2):239–244

Urbaneja A, De Verbizier J, Formery AS, Tobon-Gomez C, Nace L, Blum A et al (2019) Automatic rib cage unfolding with CT cylindrical projection reformat in polytraumatized patients for rib fracture detection and characterization: feasibility and clinical application. Eur J Radiol 110:121–127

Ureten K, Maras HH (2022) Automated classification of rheumatoid arthritis, osteoarthritis, and normal hand radiographs with deep learning methods. J Digit Imaging 35(2):193–199

Ureten K, Arslan T, Gultekin KE, Demir A, Ozer HF, Bilgili Y (2020) Detection of hip osteoarthritis by using plain pelvic radiographs with deep learning methods. Skeletal Radiol 49(9):1369–1374

Ureten K, Maras Y, Duran S, Gok K (2021) Deep learning methods in the diagnosis of sacroiliitis from plain pelvic radiographs. Mod Rheumatol 33:202

Ureten K, Sevinc HF, Igdeli U, Onay A, Maras Y (2022) Use of deep learning methods for hand fracture detection from plain hand radiographs. Ulus Travma Acil Cerrahi Derg 28(2):196–201

Varcin F, Erbay H, Cetin E, Cetin I, Kultur T (2021) End-to-end computerized diagnosis of spondylolisthesis using only lumbar X-rays. J Digit Imaging 34(1):85–95

Villamor E, Monserrat C, Del RL, Romero-Martin JA, Ruperez MJ (2020) Prediction of osteoporotic hip fracture in postmenopausal women through patient-specific FE analyses and machine learning. Comput Methods Programs Biomed 193:105484

Vinayahalingam S, Nistelrooij N, Van Ginneken B, Bressem K, Troltzsch D, Heiland M et al (2022) Detection of mandibular fractures on panoramic radiographs using deep learning. Sci Rep 12(1):19596

Wang W, Huang W, Lu Q, Chen J, Zhang M, Qiao J et al (2022) Attention mechanism-based deep learning method for hairline fracture detection in hand X-rays. Neural Comput Appl 21:18773–18785

Wang S, Wu D, Ye L, Chen Z, Zhan Y, Li Y (2023) Assessment of automatic rib fracture detection on chest CT using a deep learning algorithm. Eur Radiol 33(3):1824–1834

Watanabe K, Aoki Y, Matsumoto M (2019) An application of artificial intelligence to diagnostic imaging of spine disease: estimating spinal alignment from moire images. Neurospine 16(4):697–702

Wei D, Wu Q, Wang X, Tian M, Li B (2021) Accurate instance segmentation in pediatric elbow radiographs. Sensors 21(23):7966

Wirries A, Geiger F, Hammad A, Oberkircher L, Blumcke I, Jabari S (2020) Artificial intelligence facilitates decision-making in the treatment of lumbar disc herniations. Eur Spine J 30(8):2176–2184

Xing M, Wei G, Liu J, Zhang J, Yang F, Cao H (2020) A review on multi-modal human motion representation recognition and its application in orthopedic rehabilitation training. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 37(1):174–178

Yamada S, Maki S, Kishida S, Nagai H, Arima J, Yamakawa N et al (2020) Automated classification of hip fractures using deep convolutional neural networks with orthopedic surgeon-level accuracy: ensemble decision-making with antero-posterior and lateral radiographs. Acta Orthop 91(6):699–704

Yang J, Zhang K, Fan H, Huang Z, Xiang Y, Yang J et al (2019) Development and validation of deep learning algorithms for scoliosis screening using back images. Commun Biol 2:390

Yao LH, Leung KC, Tsai CL, Huang CH, Fu LC (2021) A novel deep learning-based system for triage in the emergency department using electronic medical records: retrospective cohort study. J Med Internet Res 23(12):e27008

Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O (2020) Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network. Eur Radiol 30(6):3549–3557

Ye Q, Shen Q, Yang W, Huang S, Jiang Z, He L et al (2020) Development of automatic measurement for patellar height based on deep learning and knee radiographs. Eur Radiol 30(9):4974–4984

Yoon AP, Lee YL, Kane RL, Kuo CF, Lin C, Chung KC (2021) Development and validation of a deep learning model using convolutional neural networks to identify scaphoid fractures in radiographs. JAMA Netw Open 4(5):e216096

Zeng B, Wang H, Xu J, Tu P, Joskowicz L, Chen X (2023) Two-stage structure-focused contrastive learning for automatic identification and localization of complex pelvic fractures. IEEE Trans Med Imaging 42(9):2751–2762

Zhang J, Ng N, Scott C, Blyth M, Haddad FS, Macpherson GJ et al (2022) Robotic arm-assisted versus manual unicompartmental knee arthroplasty : a systematic review and meta-analysis of the MAKO robotic system. Bone Joint J. 104-B(5):541–548

Zhang J, Li Z, Lin H, Xue M, Wang H, Fang Y et al (2023a) Deep learning assisted diagnosis system: improving the diagnostic accuracy of distal radius fractures. Front Med 10:1224489

Zhang J, Lin H, Wang H, Xue M, Fang Y, Liu S et al (2023b) Deep learning system assisted detection and localization of lumbar spondylolisthesis. Front Bioeng Biotechnol 11:1194009

Zhao Z, Pi Y, Jiang L, Xiang Y, Wei J, Yang P et al (2020) Deep neural network based artificial intelligence assisted diagnosis of bone scintigraphy for cancer bone metastasis. Sci Rep 10(1):17046

Zheng HD, Sun YL, Kong DW, Yin MC, Chen J, Lin YP et al (2022) Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI. Nat Commun 13(1):841

Zhewei Y (2020) Intelligent medicine. Chinese People's Medical Publishing House, Bei Jing, pp 1–10

Zhong H, Wang B, Wang D, Liu Z, Xing C, Wu Y et al (2021) The application of machine learning algorithms in predicting the length of stay following femoral neck fracture. Int J Med Inform 155:104572

Zhou Z, Zhao G, Kijowski R, Liu F (2018) Deep convolutional neural network for segmentation of knee joint anatomy. Magn Reson Med 80(6):2759–2770

Zhou QQ, Wang J, Tang W, Hu ZC, Xia ZY, Li XS et al (2020a) Automatic detection and classification of rib fractures on thoracic CT using convolutional neural network: accuracy and feasibility. Korean J Radiol 21(7):869–879

Zhou XL, Wang EG, Lin Q, Dong GP, Wu W, Huang K et al (2020b) Diagnostic performance of convolutional neural network-based Tanner-Whitehouse 3 bone age assessment system. Quant Imaging Med Surg 10(3):657–667

Zhou QQ, Tang W, Wang J, Hu ZC, Xia ZY, Zhang R et al (2021) Automatic detection and classification of rib fractures based on patients' CT images and clinical information via convolutional neural network. Eur Radiol 31(6):3815–3825

Zhu W, Zhang X, Fang S, Wang B, Zhu C (2020) Deep learning improves osteonecrosis prediction of femoral head after internal fixation using hybrid patient and radiograph variables. Front Med 7:573522