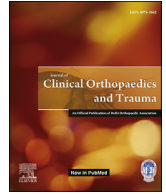




Contents lists available at ScienceDirect

Journal of Clinical Orthopaedics and Trauma

journal homepage: www.elsevier.com/locate/jcot

Real-world analysis of artificial intelligence in musculoskeletal trauma

Pranav Ajmera^{a,*}, Amit Kharat^a, Rajesh Botchu^b, Harun Gupta^c, Viraj Kulkarni^d^a Department of Radiology, Dr D.Y. Patil Medical College, Hospital and Research Center, DPU, Pune, India^b Department of Musculoskeletal Radiology, Royal Orthopedic Hospital, Birmingham, UK^c Department of Musculoskeletal Radiology, Leeds Teaching Hospitals, Leeds, UK^d Data Science Team, DeepTek Inc., USA

ARTICLE INFO

Article history:

Received 6 July 2021

Received in revised form

20 August 2021

Accepted 20 August 2021

Available online 27 August 2021

Keywords:

Artificial intelligence

Musculoskeletal

Machine learning

Imaging

Radiology

Deep learning

ABSTRACT

Musculoskeletal trauma accounts for a large percentage of emergency room visits and is amongst the top causes of unscheduled patient visits to the emergency room. Musculoskeletal trauma results in expenditure of billions of dollars and protracted losses of quality-adjusted life years. New and innovative methods are needed to minimise the impact by ensuring quick and accurate assessment. However, each of the currently utilised radiological procedures, such as radiography, ultrasonography, computed tomography, and magnetic resonance imaging, has resulted in implosion of medical imaging data. Deep learning, a recent advancement in artificial intelligence, has demonstrated the potential to analyse medical images with sensitivity and specificity at par with experts. In this review article, we intend to summarise and showcase the various developments which have occurred in the dynamic field of artificial intelligence and machine learning and how their applicability to different aspects of imaging in trauma can be explored to improvise our existing reporting systems and improvise on patient outcomes.

© 2021 Delhi Orthopedic Association. All rights reserved.

1. Introduction

Musculoskeletal (MSK) imaging has come a long way, particularly in the past two decades with rapid developments in the field of medical imaging technology. With addition of artificial intelligence (AI) to medical imaging technology has opened the doors for innovation. It can be categorized as general purpose or point-based AI solutions focussed on a single pathology. Most of the current developments are happening in the space of point-based solutions, considering the current maturity of the recent technological innovations. The growth in AI has been due to three key reasons: access to high quality annotated data, rapid developments in the space of deep learning and a massive leap in the computational power through cloud-based graphic processing unit (GPU).

Injuries to the MSK system ranks amongst the top causes for self-referred and unscheduled visits to the emergency department.¹ The direct cost of treatment of these injuries was nearly \$130 billion, with a little less than half of all injuries involving tendons and ligaments. Over three-fourths of all missed working days being attributable to their injury. The costs involved include

diagnostic costs, such as radiography, computed tomography (CT), magnetic resonance imaging (MRI) and dedicated ultrasonographic evaluation, followed by their relevant treatment costs.²

MSK pathologies due to various causes, which also includes trauma, are the leading contributor to disability worldwide and also the most significant contributor towards the global need for rehabilitation. They are also the single most significant attributable cause for years lived with disability, contributing to nearly 17% of the total burden. Amongst all the MSK conditions, low back pain is the most common cause of disability, closely trailed by fractures (436 million globally on an annual basis), followed by osteoarthritis.³ Trauma-related disability has seen a significant rise due to enhanced mobility which is the boon and bane of the new era due to high-speed accidents. The risk of fracture has been rising over the previous decades because of an overall increase in the prevalence of bone diseases like osteoporosis. There is a gender preference, with females suffering more frequently than males. Amongst these, the incidence particularly rises for those over 50-years of age. Nearly 200 million are affected by osteoporosis alone.^{4,5} In the US, this translates into 1.5 million fractures annually as per the pre-2000 cross-sectional study.⁶ An American research venture conceived a Markov decision model which predicted the number to be around 2 million, resulting in expenditure costs of over \$17 billion by 2005, most of which pertains to in-patient care.⁷

* Corresponding author. 185, G-Floor, Dayanand Vihar, Delhi, 110092, India.
E-mail address: pranavajmera@gmail.com (P. Ajmera).

Misinterpretation or delayed diagnosis in trauma imaging can lead to an increase in mortality and morbidity. The delayed diagnosis of occult fractures, especially at sites like the scaphoid or pediatric long bone fractures, can have serious morbidity issues as intervention gets delayed. A few surveys have shown that radiological error closely trails improper clinical assessment as the contributor towards missed diagnosis. In radiology, the primary cause is attributable to the fact that patients who present to the emergency, the interpretation is performed by registrars and not consultants who have significant experience in interpreting and correlating these findings clinically.⁸ Multiple studies have been conducted to assess the rate of missed injuries and delayed diagnosis, and while the results have been variable, varying between 1.3 and 40% based on the study population and the datasets, it would be fair to say that the number approaches double digits.⁹ Missed diagnosis is particularly more common in patients with multiple injuries and cases of road traffic accidents.

Advances in AI enormously benefit the field of radiology and imaging. MSK fractures and bone/soft tissue injury are to significantly benefit from these developments. Current research and the role of AI through multiple clinical research papers clearly show that a physician or imaging expert aided by AI fares far better than the unaided, and this is the crucial aspect of how AI will be deployed in radiological workflows in the near future.¹⁰

Establishing AI in radiology workflow helps to benefit developed and developing countries even though the problem statements in both these situations are different. In developed economies, the issue is regarding insurance and malpractices, which can arise if occult fractures and subtle lesions are missed. While in developing economies, there is a core issue of lack of adequately trained radiology resources who can screen for such disease conditions quickly. As a result, there is a significant advantage of deploying AI in workflows for quick and instant triage. Triage solutions are an innovative, sustainable, and scalable way for ensuring equitable distribution of healthcare technology across the world. The technology allows for various models and ensures that the solution can be deployed across various geographic.¹¹

In the region where there is no access to the internet, the simple tool of fracture triage can be employed even on an edge device solution (next to a digital X-Ray machine or a CT scanner). At the same time, hospitals and clinics which are working in an agile manner with a skeletal information technology (IT) and support staff can use adopt the cloud AI system that harnesses central processing unit (CPU) and graphic processing unit processing from the cloud enabling scalable allocation of resources.¹²

This review will explore the current applicability of AI in MSK trauma, the developments in the past leading up to the current. We also provide comments on how the dynamics of the radiologists reporting room would change with integrating AI and other inter-linked technologies.

2. Methods

An eclectic search of the PubMed and EMBASE database was performed using numerous combinations of the keywords, “artificial intelligence”, “deep learning”, “machine learning”, “musculo-skeletal trauma”, “law”, “CT”, “MRI”, “USG”, “radiomics”, “fracture”, “osteoporosis”, “future directions”.

3. AI- the what, how and why?

AI is a field that amalgamates computer science and robust datasets to allow problem-solving. AI further is grouped based on its extent as narrow AI- “Artificial Narrow Intelligence (ANI) or broader AI. ANI can be trained to perform specific tasks while the

broad AI, also known as general AI, would have self-awareness and would therefore permit it to solve problems, learn and plan and apply it to the future. All the AI applications in the present-day world are ANI. General AI still remains an entirely theoretical concept, but with the ongoing active research in the field, one day, the concept of super, self-thinking AI will be feasible. Till then, there is much scope for further improvisation of ANI to allow it to supersede tasks that would previously require human input. AI encompasses the fields of machine learning (ML) and deep learning (DL), both of which are concerned with creating AI algorithms to create expert systems that make predictions based on accurate data.¹³

Arthur Samuel coined the word “Machine Learning” in 1952 and described it as the ability of a computer to learn without being explicitly trained to do so. In 1956, John McCarthy coined the term artificial intelligence at the first ever such conference. In the 1980s the neural networks came into being and soon gained popularity.^{14,15} ML can be further classified into four types, based on the learning styles-supervised, unsupervised, semi-supervised, and reinforcement, amongst which supervised and unsupervised are the more commonly used ones; the differences between them have been depicted in [fig. 1](#).¹⁶

Before we move further it is important to understand the multitude of algorithms that exist today based on various learning styles. For easier understanding, we have grouped them based on their similarities in [fig. 2](#).¹⁷

A simplified explanation of how the various AI algorithms work using the various models. Regression is focused on iteratively refining a model's relationship between variables using a measure of inaccuracy in the model's predictions. Instance-based learning models function by building up a database of example data and comparing new data to the database using a similarity measure. As a result, instance-based approaches are also known as winner-take-all and memory-based learning methods. The representation of the stored instances and the similarity measures used between instances are both highlighted. Regularisation models is a modification to another approach (usually regression methods) that penalises models for their complexity, favouring simpler models that are also better at generalising. Decision tree approaches create a model of decisions based on the actual values of data characteristics. In tree architectures, decisions fork until a forecast choice is made for a specific record. For classification and regression problems, decision trees are trained on data. Decision trees are popular in machine learning because they are often fast and accurate. Bayesian algorithms make use of the Bayesian theorem to solve the problems of classification and regression. Clustering, like regression, identifies the problem type and procedure type.^{17,18}

Artificial Neural Networks (ANNs) are models based on the structure and/or function of biological neural networks. They are a sort of pattern matching that's often used for regression and classification problems. However, they are actually a colossal subfield comprising various methods and variations for a wide range of applications. DL models are essentially extensions of the ANNs, concerned with building much more complicated models using multiple layers. The “Deep” in DL essentially implies the presence of three or more layers of the neural network, “nodes”, including both input and output. It is essentially scalable ML, as it can significantly reduce the necessity of human intervention, something which was needed in the early models of ANN and classical ML models.^{13,18} A subtype of DL is the convolutional neural network (CNN), which utilises images for processing. It differentiates between items in an image by allotting their biases and weights. It then makes use of relevant filters to understand the spatial and temporal dependencies.¹⁶ Ensemble methods are models that are made up of multiple individually weaker models that are trained

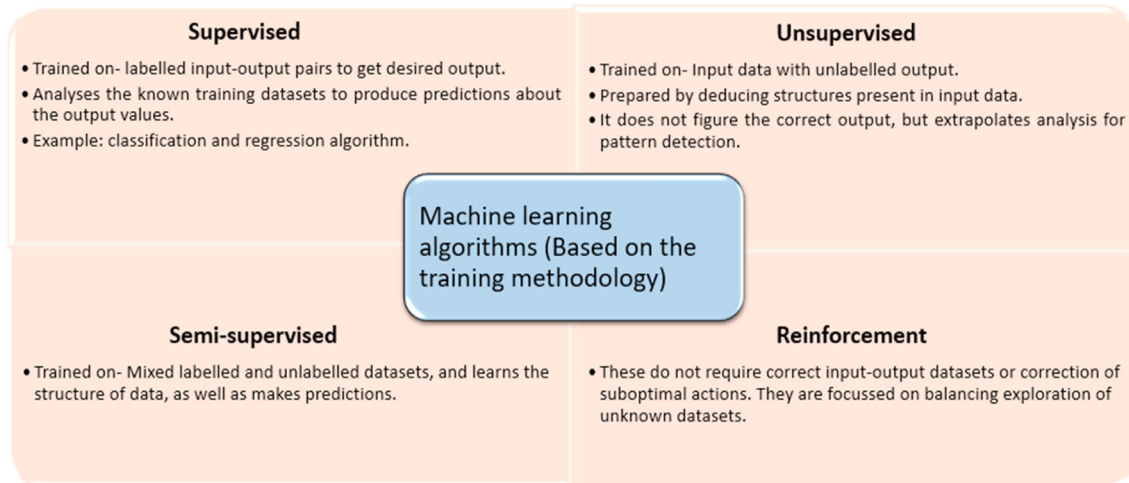


Fig. 1. The numerous machine learning algorithms over the period of time have been grouped under four headers as depicted, with each category being distinguished from the other based on its specific learning style and the features centric to it.

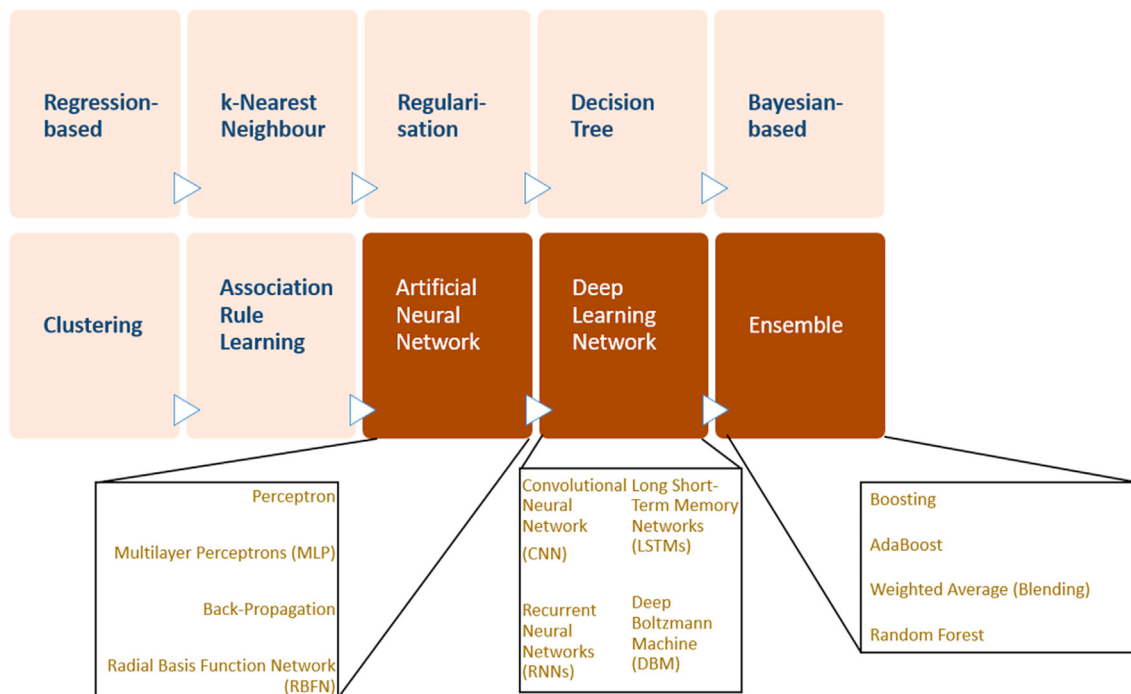


Fig. 2. An illustration depicting the various AI algorithms grouped together based on their similarities. The three major algorithmic networks which have recently shown incremental usage have been highlighted in the brown boxes, with a lens view into the various subcategories of algorithms that exist within the larger network.

independently and then integrated in some way to create a final prediction. This is a very potent group of tactics, and as a result, it is extremely popular.¹⁸

4. State of the art developments with a snapshot of the past decade

Within the past decade, the applications of AI have increased exponentially, and there has been a corresponding yearly increment in research targeting use cases for development in the field of musculoskeletal radiology. The main factor contributing to the rapid growth is that with each model, there is increased learning in the field and also, there is an increased availability of data for

training, validation and testing. Earlier applications were restricted to the identification of bone tumours, assessing mineral density in bones, recognition of trabecular patterns in long bones.¹⁹ Also, the modality exploited primarily in the past was radiography with lesser developments in the field of CT and MRI. However, the past five years particularly have shown an increase in the number of DL algorithms being developed to detect fractures on radiographs and CT; detect injury to the meniscus, ligament tear, bone marrow edema on MRI. A crucial step for this is the ability of AI to segment the region of interest, as depicted in fig. 3.

Another significant breakthrough has come in the field of ultrasonography (USG), which was erstwhile neglected. USG has gained significant traction as it provides high intra tissue contrast

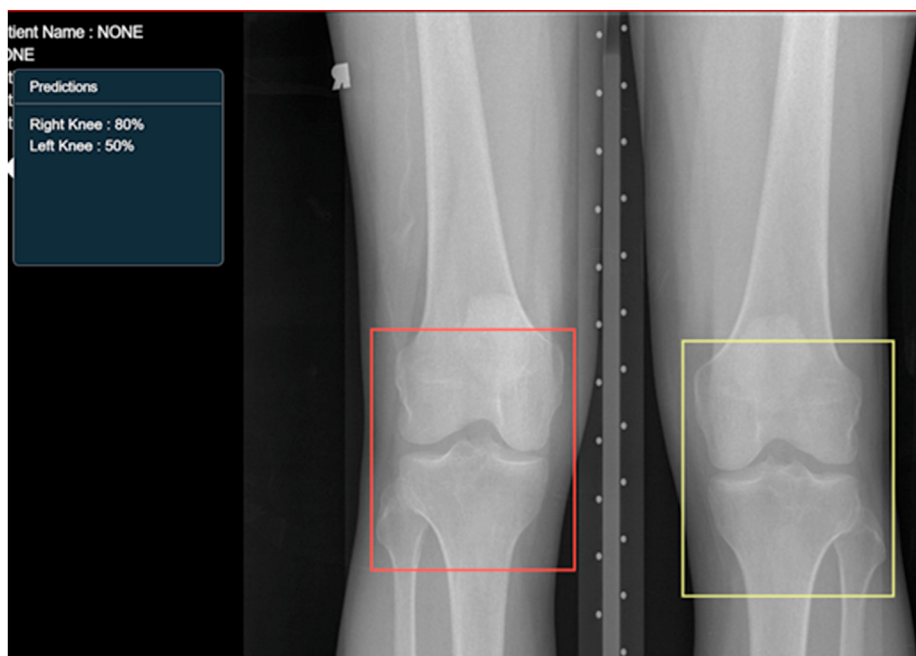


Fig. 3. A snapshot of how a scan on an AI-enabled platform looks - the AI automatically segments the region of interest. For example: the red bounding box demarcates the right knee, while the yellow bounding box demarcates the left knee. Right, and left knee segmentation has been done by training the AI model to identify the location of fibula and DICOM metadata. On the top left corner, a text box shows the AI confidence to recognise if the finding is correct. Such a system is not used as a stand-alone but as a force multiplier by having specific algorithms to identify regions of interest. Other algorithms can build upon this and further identify the abnormality, with the development and addition of such algorithms being a continuous step. [Image courtesy: DeepTek Inc.™ segmentation algorithm.](#)

resolution and dynamic assessment for evaluating small structures, ligaments, muscles and tendons. In MSK imaging, this high contrast resolution in the superficial 3–5 cm zone is critical for assessment of soft tissue such as tendon, ligament and bursae. These have been made possible due to extensive research and developments in investment powerful yet compact and light-weight transducers, Power Doppler and ultrasound elastography. The result of this improvement has a direct impact on the quality of data acquisition. Better quality data results in good output from resultant computer-aided detection (CAD) algorithms.

The key to developing robust point solutions is access to good quality annotated datasets. To improve data availability, several data libraries have come up, like Kaggle, Google dataset search, UCI machine repository, and the natural language processing database, Datahub and google fusion tables.²⁰ These make the data accessible to researchers and data science experts. Kaggle follows an interesting model of incentivising the development of algorithms by coming up with competitions that support innovations. Data managers like CKAN, Quandl and DataMarket exist, which providing users with a platform.²¹

5. DL algorithms in the realm of MSK trauma

5.1. Radiograph based DL algorithms

The key element of MSK imaging is the pivotal role plain radiograph plays in assessment. It is still relevant and the key modality to start with while dealing with MSK trauma; while many specialities have neglected them for the sake of cross-sectional imaging, radiographs still hold an enviable position in MSK imaging. The signs of injury on MSK radiographs can be as subtle as blurring of the fat planes, the fullness of the fat pad and indistinct subtle fracture lines, which can be seen on all or one of the view (antero-posterior/postero-anterior, oblique, lateral). The other end

of the spectrum are complex fractures with prominent fracture lines that are unlikely to be missed. Radiographs in two planes provide a wealth of information to aid clinical diagnosis and help set the appropriate clinical management. A radiograph with no abnormal findings also has critical value in patient management. All of this makes skeletal radiographs an indispensable tool, particularly in emergencies of the MSK system.

The majority of the algorithms have focussed on identifying fractures of a single anatomical region, and most of them have utilised methods based on CNN, like Kim DH et al.²² who developed a DL system for detection of wrist radiographs which surpassed other computation methods of automated fracture detection based on detection of edge, features and segmentation. Gale W et al.²³ developed a DenseNet based architecture to detect hip fractures from frontal pelvic radiographs, and their system managed to beat expert human radiologists.

Vertebral fractures are a stronger determinant of future fractures and are linked to a greater risk of death, persistent back pain, kyphotic deformity, immobility, and a loss of self-esteem. The overarching aims in managing osteoporosis are early detection and adequate therapy. Chen HY et al. developed a ResNeXt architecture as the backbone and further utilised transfer learning as it helps in effective learning of parameters by the system even if it is trained on unrelated datasets. They utilised abdominal radiographs as their dataset to identify vertebral fractures and arrived at an area under the curve of 0.72, which was though lesser than the parameters of trained radiologists, orthopaedic surgeons and physicians, serves to elaborate the extent to which AI systems have developed. Despite having a lower accuracy, the system can identify regions that may have fractures and, in the process, save the clinicians time in evaluation of radiograph.²⁴

An example of a DL algorithm at work identifying rib fractures is shown in [fig. 4](#). Another DL algorithm successfully identifying the region of pneumothorax on a chest radiograph is elaborated in [fig.](#)

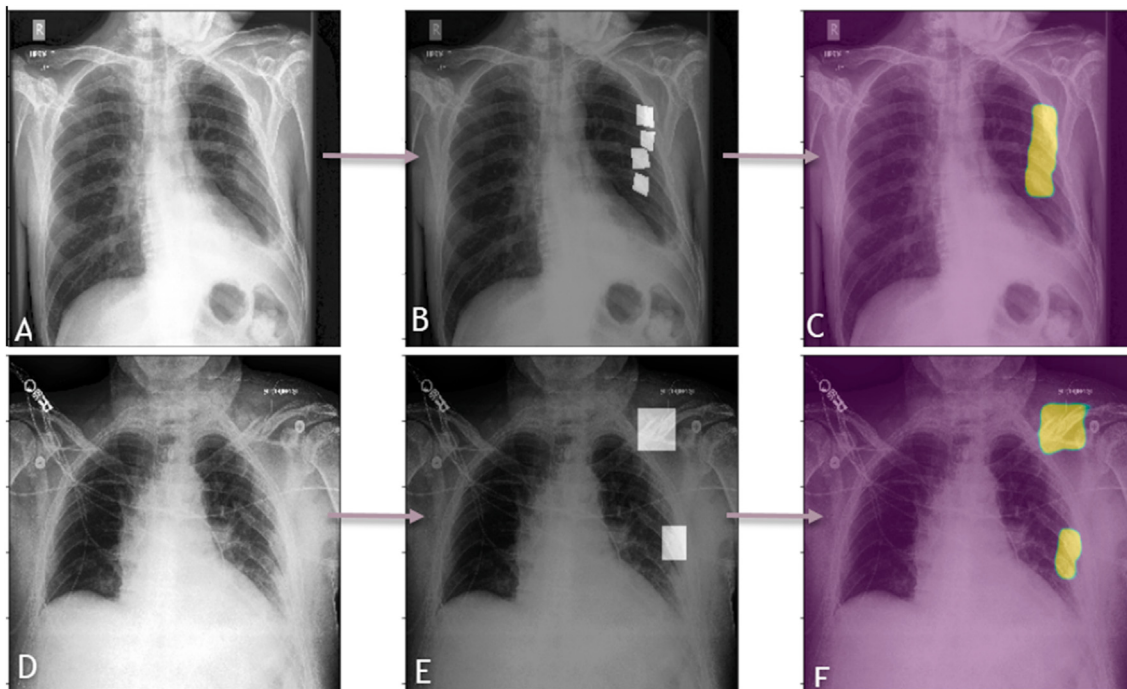


Fig. 4. A panel of chest radiographs (A-C and D-F) where a DL algorithm has been used to identify the presence or absence of rib fractures. The images on the left side (A and D) are the raw images, that is the input data where we can see that there are multiple rib fractures in A and rib fractures with a clavicular fracture in B. The center images (B and E) are the same regions annotated by a radiologist. The images on the right side (C and F) are the fracture regions as identified by the DL algorithms, as can be seen by the masks. Image courtesy: DeepTek Inc.™ segmentation algorithm.

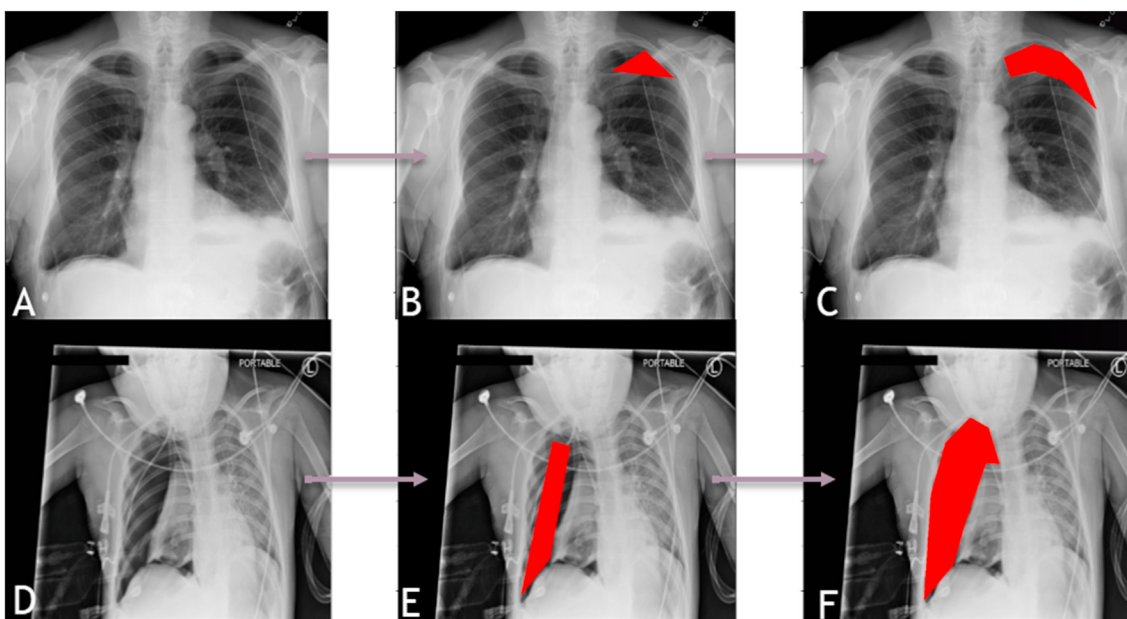


Fig. 5. A panel of chest radiographs (A-C and D-F) where a DL algorithm has been used to identify the presence of pneumothorax which can be a complication of rib fracture. The images on the left side (A and D) are the raw images, where we can see the presence of massive pneumothorax displacing the left lung in A and milder pneumothorax in the right hemithorax in D. The same regions have been annotated by a Board-Certified Radiologist in the images in the center (B and E). The images on the right side (C and F) are the involved regions as identified by the DL algorithm and marked using the red masks. Image courtesy: DeepTek Inc.™ segmentation algorithm.

5. In the above literature, we reviewed how point algorithms assessed radiographs and detected pathologies like fractures. Now we discuss how multiple algorithms work together cohesively give more information, and helping assist experts make their process

much easier. The AI here may utilise different approaches to arrive at the fracture results, for example, the AI may use other signs like shape, texture features, edge detection, transfer learning.²² Another focus area would be the generalisability and scope of the algorithm. That is, a single solution should be able to identify the region and

Upload Date	TAT	Chat	Triage	Report	Red X	Person	Library
		Chat	Green	Report icons	Red X	Person	-
		Chat	Red	Report icons	Red X	Person	-
		Chat	Green	Report icons	Red X	Person	+
		Chat	Yellow	Report icons	Red X	Person	+
		Chat	Yellow	Report icons	Red X	Person	+
		Chat	Green	Report icons	Red X	Person	+
		Chat	Yellow	Report icons	Red X	Person	+
		Chat	Green	Report icons	Red X	Person	+
		Chat	Red	Report icons	Red X	Person	+
		Chat	Red	Report icons	Red X	Person	+

Fig. 6. A snapshot of a PACS enabled with RAG AI (Red, Amber and Green AI Criticality Legend) and other tools to ease the workflow. On the right side is the data and time of each scan; next to it are the RAG Criticality Legend assigned by AI to each scan, depicting the urgency as determined by the AI algorithm. Red- Need immediate reporting and dispatch; amber-need priority reporting, but maybe acceptable to delay slightly; Green-routine reporting scan. *Image courtesy:* DeepTek Inc.™ segmentation algorithm.

assess for the presence of a fracture. At the same time, its applicability should be global, that is accuracy is not affected by the ethnic and geographic origin of the patient. Such an approach was successfully attempted by Ma et al.,²⁵ who developed a seemingly two-step approach. Drawing inspiration from how human radiologists first identify the bone involved and then zeroes down on the exact region of fracture, they developed this DL model. The first algorithm called Faster R-CNN fulfils the role of detection and identifies the anatomical location (from amongst the 20 bones) under review. Subsequently, bony regions are extracted by bounding boxes to identify the bone. For the latter part of region-wise fracture classification, they used a Crack-sensitive CNN, using the Schmid filter. They obtained an accuracy and F- score of over 90%.

Algorithms need to attain more maturity to allow classification of fractures as per existing classification systems as this will have a role in assessing the patient prognosis and identifying treatment alternatives. DL can play a huge role as many of the classification systems are tedious and take time to interpret; this time could instead be better utilised in patient assessment. A DL model developed by Oczak et al.²⁶ to classify the ankle fractures based on the 2018 Arbeitsgemeinschaft für Osteosynthesefragen (AO)/Orthopedic Trauma Association (OTA) over a dataset of 4941 patients reached an average AUC of 0.90. A study by Li YC et al. to classify vertebral fractures based on the Genant classification using plain lateral radiographs from 941 patients reached an AUC of 0.919, 0.989 and 0.990 for grades-1, 2 and 3 respectively. The results were particularly good in patients suffering from osteoporosis.²⁷ Another study was performed by Tobler P et al. to classify distal radial fractures based on displacement, intra- and extra-articular, multi fragmented, metal implant in-situ was performed using a ResNet18 D-CNN architecture, exploiting 15,775 frontal and lateral

radiographs of the distal radius.²⁸ While the model is successful in automatically detecting fracture, the classification of the type of distal radius fracture is variable in terms of accuracy. It is lower in cases of involvement of joint and fragment displacement. Thus, while it cannot serve as a stand-alone, it can still be used as an assisting tool.

There has been a growing focus on the need for larger, high-quality datasets; researchers have come up with multiple methods to refine data by improved processing, and augmentation of data amongst others. In the very recent past, another domain that has come under the lens of innovators is synthetic images. Chedid N et al. have developed an ML algorithm that can convert diagrams of fractures into realistic radiographs. Towards this end, they used a pix2pixHD algorithm and a new human-guided post generation refinement phase. Multiple bodily components, such as the humerus, wrist, and fingers, could be convincingly replicated by their fine-tuned networks. Other features of appearance, such as over-or underexposure, and orientation, vary across the synthetic photos. Based on the results of the visual Turing test, the physicians had an overall sensitivity of 49.63% in differentiating real from synthesized ones, something which would be expected by a random error of chance. The sensitivity of the system trained on both types of data was 93.33% for transverse fractures without fracture fixation hardware and was 75.5% for any fracture without fracture fixation hardware. In contrast, the system trained on real data alone had a poor sensitivity of 73.3% and 67.2% for transverse fractures and any kind of fracture without any type of fracture fixation hardware, respectively. However, there was a corresponding decrease in accuracy with the addition of synthetic data from 82.1% to 81.7%, a marginal decrease.²⁹

5.2. CT based DL algorithms

While there is a vast amount of data now available for research of deep learning in radiographs, the corresponding data for the application of CT is not comparable. However, recently, more and more focus is being placed on this area.

In a study by Jin L et al.³⁰ to develop a DL model to classify rib fractures from CT on a dataset of 7473 annotated traumatic rib fractures, they used a 3D U-Net architecture. The model had a sensitivity of 92% and a DICE score of 72.5% on testing. Pranata YD et al.,³¹ built a DL model based on ResNet and another on VGG- CNN to identify calcaneal fractures. Subsequently, a bone fracture detection algorithm incorporating the Speeded-Up Robust Features (SURF) method, Canny edge detection, and contour tracing was utilised to precisely localise the region. These CNNs had a high accuracy bordering 98% and are hence feasible for deployment.

In an audit performed in the UK and published in BMJ, one of the most common causes of patients presenting to the emergency department is trauma.³² Trauma is one of the most frequent conditions for which CT is advised, particularly in an emergency department. In a retrospective, electronic chart review performed at the Penn State Hershey Medical Center on an image set of 81,201 images performed during emergency hours revealed that the most common modality associated with radiological discrepancy during “off hours” was CT and that fractures were the most commonly missed finding. One contributory factor to this event is that most scans during emergencies are read-only by residents.³³ It is, therefore, worthwhile to develop DL algorithms that can provide a second opinion to the resident on call and minimise the chances of scan interpretation error.

5.3. MRI based DL algorithms

As was the case in CT, MRI too has been receiving a lot of

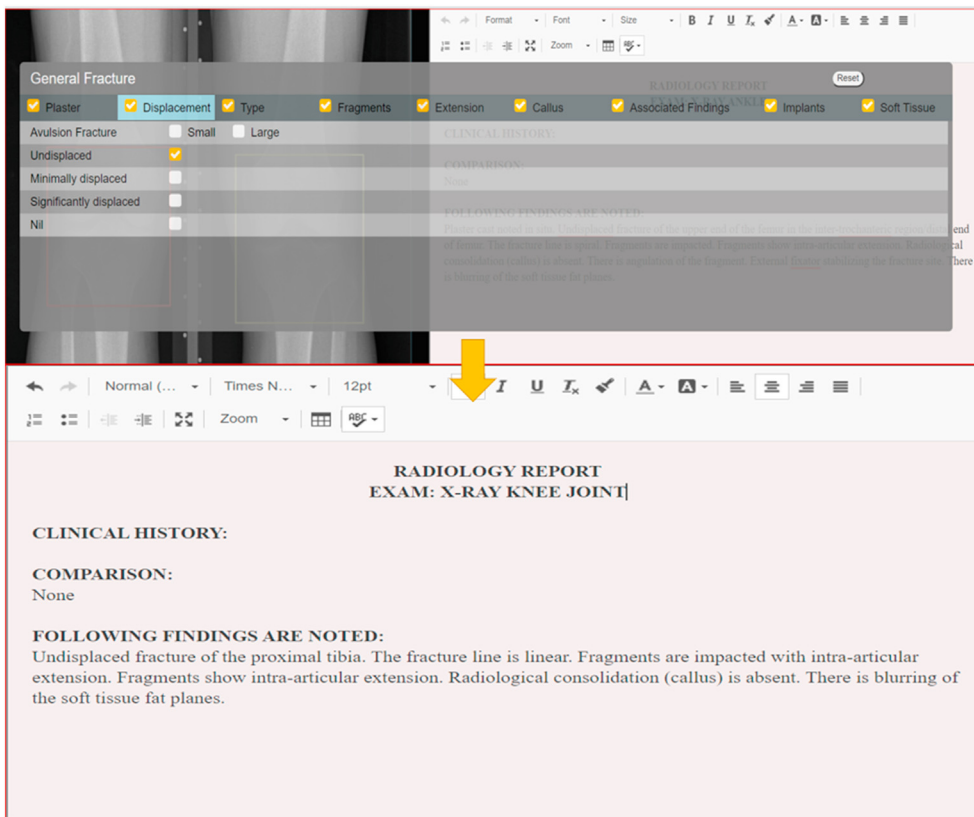


Fig. 7. A sequential representation of a smart reporting tool as a microservice deployed in an AI-enabled PACS environment. The radiologists look at the scan with a region of interest identified by AI, allowing for quicker scan reads. After a look at the segmented regions, they can simply check through the list under various headers (top image), choosing all that is appropriate for the case. As a result, all points of significance are covered by all radiologists, ensuring standardisation. The output report is generated as shown (bottom image). This human-in the loop model with AI doing the groundwork is the highlight of such a system. *Image courtesy:* DeepTek Inc.™ segmentation algorithm.

attention by researchers keen to exploit the extremely good resolution of MRI for soft tissue to detect pathologies that cannot be assessed on either radiographs or CT, and to connect this capability with advancements in the field of AI. Amongst all the joints, involvement of the knee and hip remains the most common sites and these have been the primary areas of focus in most studies.³⁴ Some of these are listed in [table-1](#).

Gang H et al.⁵² developed an AI model to analyse the efficacy of knee ligament trauma repair. In a study by Liu F et al., the DL model approached the output of a radiologist when detecting only the anterior cruciate ligament of the knee.⁵³ Other researchers like Astuto B et al.,⁵⁴ Bien N et al.⁴⁸ took a broader approach. They tried building a model which could perform multiple functions like identify ACL tears, bone marrow edema, meniscus tear, and cartilage abnormality. While the results were sub-par to multiple radiologists, they provide a direction for future growth by serving as an essential proof of concept. A study by Kim M et al. to detect rotator cuff tears using the DL algorithm had a very high AUV of 0.98 with a diagnostic accuracy of 87%. Their study is amongst the first ones to utilise DL for this purpose. The dataset used by them of 2447 patients has been made publicly available for development and testing by other researchers.⁴⁶

5.4. USG based DL algorithms

Improved accuracy in assessing musculoskeletal problems has occurred subsequent to advancements in USG imaging technology.^{55,56} There are a certain group of pathologies that are best detected on USG, particularly those needing dynamic manoeuvres-

shoulder impingement syndrome, rupture/subluxation of the long head of biceps tendon, tendons around fracture hardware, and costal cartilage fracture.⁵⁷ They can be utilised to identify the site while administering nerve blocks. One such system was developed by Shin Y et al.,⁵⁶ who developed a DL algorithm for femoral nerve region segmentation utilising U-net. The model obtained a satisfactory performance with the intersection over union over the test set equal to 0.638. The segmentation accuracy over the test set was 83.9%.

The dependency on subjective evaluations of displayed pictures, as well as differences in image acquisition and equipment employed among studies, have slowed the adoption of US compared to magnetic resonance imaging (MRI). Computer-assisted diagnostic (CAD) solutions have acquired a prominent role in radiology to assist in overcoming these issues, as well as the ambiguity with which musculoskeletal illnesses may appear on US imaging. These systems give a quantitative analysis and a second opinion, allowing radiologists to swiftly make correct and consistent image judgments.⁵⁶⁻⁵⁸

Classical CAD systems suffer from various limitations pertaining to the algorithm and due to bias. DL-based systems are better at being more robust and more generalisable. Within the vast scope of AI, its utility in USG is mainly linked to classification, segmentation and diagnosis. Its applicability for image augmentation and enhancement still remains unexplored.⁵⁹

A detailed compilation of certain important experiments performed in the field of MSK applying AI to all three modalities in the last five years is listed in [table-1](#).

Table-1
Ongoing developments in the field of AI in MSK, in the recent years (5-years; 2017–2021).

Serial No.	Year	Modality	Aim (task assisted)	MSK tissue	AI Approach	Dataset used	Performance Output	Obvious limitations	Reference
1.	2021	Radiograph	Application of deep learning algorithm to detect and visualize vertebral fractures on plain frontal radiographs	Bone	ImageNet convolutional neural network (CNN)	1306	Area under curve = 0.72 Sensitivity = 73% Specificity = 73%	-Small dataset	Chen HY et al. ³⁵
2.	2020	Radiograph	Assessment of a deep-learning system for fracture detection in musculoskeletal radiographs	Bone	Ensemble of CNNs	7,15,343	Overall AUC = 0.974; Sensitivity = 95.2%; Specificity = 81.3%; PPV = 47.4% NPV = 99.7%	over-represented infrequently acquired regions	Jones RM et al. ³⁶
3.	2020	Radiograph	Bone fracture detection through the two-stage system of Crack-Sensitive Convolutional Neural Network	Bone	Double CNN models in sequence- FastNet, followed by CrackNet	3053	Accuracy = 0.91; precision = 0.89; recall = 0.90; F-measure = 0.90	-Small dataset	Ma Y et al. ²⁵
4.	2019	Radiograph	Classify hip fracture, patient traits and hospital process variables	Bone	CNN's	23,602	The fracture was predicted moderately well from the image (AUC = 0.78) and better when combining image features with patient data (AUC = 0.86)	-Absence of a reliable gold standard. -Limited label accuracy. -Limited accuracy of covariate data. -Pre-processing reduces image resolution.	Badgeley M et al. ³⁷
5.	2018	Radiographs	Deep neural network improves fracture detection by clinicians (all extremities for pretraining but wrist radiographs for final training, validation and testing)	Bone	CNN	1,32,345	The average clinician's sensitivity was 80.8% (95% CI, 76.7–84.1%) unaided and 91.5% (95% CI, 89.3–92.9%) aided, and specificity was 87.5% (95% CI, 85.3–89.5%) unaided and 93.9% (95% CI, 92.9–94.9%) aided.	-Single Institute study -Ground truth is subject to the experience of the radiologist	Lindsey R et al. ³⁸
6.	2018	Radiograph	The ability of a deep learning algorithm to detect and classify proximal humerus fractures using AP shoulder radiographs.	Bone	CNN	1891	Sensitivity = 0.99 Specificity = 0.97; Youden index = 0.97; Area under curve = 1.0	-Neer classification was used, which is only moderately reliable. -Cannot be applied to clinics	Chung SW et al. ³⁹
7.	2017	Radiograph	Automated deep learning system to detect hip fractures from frontal pelvic x-rays	Bone	Regression-based CNN	53,000	The area under the ROC curve of 0.994	Small labelled dataset	Gale W et al. ²³
8.	2017	Radiograph	Automated fracture detection on plain radiographs (wrist radiographs).	Bone	Inception V3 Network-CNN	11,112	The area under the ROC curve 0.954	-Ground truth was a radiologist (human) -Small labelled dataset.	Kim DH et al. ⁴⁰
9.	2017	Radiographs	Automatic Classification of Proximal Femur Fractures	Bone	Attention Models-Spatial transformer	1000	High sensitivity and specificity	-Small dataset (Single institution study)	Kazi et al. ⁴¹
10.	2021	CT	A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance.	Bone	CNN	8529	-Increased detection recall and classification accuracy (0.922 and 0.863) compared with the radiologists alone (0.812 vs. 0.850). -The radiologists achieved a higher precision rate, recall rate, and F1-score for fracture detection when using the deep learning model, at 0.943, 0.978, and 0.960,	NA	Meng XH et al. ⁴²
11.	2020	CT	A multiscale Deep Learning Method for Quantitative Visualization of Traumatic Hemoperitoneum at CT: Assessment of Feasibility and Comparison with Subjective Categorical Estimation.	Bone	3D- U-Net	130	Mean DSC for the multiscale algorithm was 0.61 ± 0.15 compared with 0.32 ± 0.16 for the 3D U-Net method and 0.52 ± 0.17. AUCs for automated volume measurement and categorical estimation were 0.86 and 0.77, respectively (P = .004). An optimal cutoff of	-Single institution study	Dreizin D et al. ⁴³

Table-1 (continued)

Serial No.	Year	Modality	Aim (task assisted)	MSK tissue	AI Approach	Dataset used	Performance Output	Obvious limitations	Reference
12.	2020	CT	Automatic Detection and Classification of Rib Fractures on Thoracic CT Using Convolutional Neural Network: Accuracy and Feasibility.	Bone	Faster R-CNN and YOLOv3	1079	278.9 mL yielded Accuracy = 84%, Sensitivity = 82%, Specificity = 93%, PPV = 86%, NPV = 83%. The precision of the five radiologists improved from 80.3% to 91.1%, and the sensitivity increased from 62.4% to 86.3% with artificial intelligence-assisted diagnosis. On average, the diagnosis time of the radiologists was reduced by 73.9 s.	-The current model cannot show the anatomical location of the rib fractures (right or left, number of ribs, anatomical name of fractured rib) -Small validation test set	Zhou QQ et al. ⁴⁴
13.	2018	CT	An automatic system that can detect incidental osteoporotic vertebral fractures in chest, abdomen, and pelvis.	Bone	ResNet34 model for feature extraction; Long-short term memory model	1432	These results indicate that our CNN/LSTM approach has high efficacy for diagnosing OVF and its performance is on par with practicing radiologists. Mean area under the curve = 0.98	-Single institution study, therefore generalisability is arguable. -Single label for entire model, therefore chances of confounding.	Tomita N et al. ⁴⁵
14.	2020	MRI	MRI-based Diagnosis of Rotator Cuff Tears using Deep Learning and Weighted Linear Combinations	Muscle, tendons	Base model = VGG-16	2492	Mean area under the curve = 0.98	-Single Institution study	Kim M et al. ⁴⁶
15.	2019	MRI	Deep Learning Algorithm in Detecting Osteonecrosis of the Femoral Head on MRI	Bone	ResNet-CNN	1892 hips (1037 diseased and 855 normal)	Sensitivity and specificity for the external test set were 84.8% and 91.3% for the DL algorithm. Sensitivity and specificity for the geographic external test set were 75.2% and 97.2% for the DL algorithm. Higher than less experienced radiologist, and comparable to the experienced radiologist.	-Ideal testing environment. -Slight selection bias. -Difficult to know if the performance of the model will be hindered by other diseases affecting the trabecular pattern.	Chee CG et al. ⁴⁷
16.	2018	MRI	Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet	Ligament	MRNet (CNN)	1370	In detecting abnormalities, ACL tears, and meniscal tears, this model achieved area under the ROC (AUC) values of 0.937 (95% CI 0.895, 0.980), 0.965 (95% CI 0.938, 0.993), and 0.847 (95% CI 0.780, 0.914), respectively, on the internal validation set.	-Performance was sub-par as compared to radiologists.	Bien N et al. ⁴⁸
17.	2018	MRI	Super-resolution musculoskeletal MRI using deep learning	NA (Scan quality)	3D- CNN "DeepResolve"	124 double echo in steady-state (DESS) data sets with 0.7-mm slice thickness and tested on 17 patients.	Significantly better structural similarity, peak signal to noise ratio, and root mean square error than tricubic interpolation, Fourier interpolation, and sparse-coding super-resolution for all down-sampling factors.	It did not match the image quality of the high-resolution ground-truth images, but it outperformed other resolution enhancement methods.	Chaudhari AS et al. ⁴⁹
18.	2018	USG	investigation into the feasibility of using deep learning methods for developing arbitrary full spatial resolution regression analysis of B-mode ultrasound images of human skeletal muscle.	Muscle	Feature engineering (Wavelet), convolutional neural networks (CNN), residual convolutional neural networks (ResNet) and deconvolutional neural networks	8	Deconvolutional Neural Network > CNN/ ResNet > Wavelet	None stated.	Cunningham R et al. ⁵⁰
19.	2017	USG	Ultrasound aided vertebral level	Bone	Deep CNN, Random Forest	19	DL method outperformed the		Baka N et al. ⁵¹

(continued on next page)

Table-1 (continued)

Serial No.	Year	Modality	Aim (task assisted)	MSK tissue	AI Approach	Dataset used	Performance Output	Obvious limitations	Reference
			localization for lumbar surgery				Random Forest on the test dataset (F-measure of 0.90 vs 0.83)	Semi-automatic (therefore, user dependent)	

6. AI in radiology workflow and image processing

6.1. Prioritisation of the worklist

AI in radiology workflows is likely to make workflows smarter and faster. Automatic study list reprioritisation is one such feature; this can convert existing Picture Archiving and Communication System (PACS) to smart PACS. Stat studies with critical findings, even if they happen later in the day, can be automatically reprioritised by AI on the top of the worklist. This principle can be applied to fracture, bleed, vessel rupture and pneumothorax. Many at times there are findings of significance that are detected incidentally; if triaged timely, this can help ER physicians and imaging experts to prioritise the studies. AI-enabled features plugged in existing PACS can fit well into this role by allowing automatic triage, often with provision for notification to the expert on what to expect, saving up on essential time. An example of such a system employed for deployment at tertiary care centres is provided in [fig. 6](#).

6.2. Accelerating imaging

Many of the conditions mentioned above describe AI tools that work on scans post-acquisition. However, AI even has a role during image acquisition; AI-enabled techniques which can shorten image acquisition can benefit studies of trauma. This is particularly true for MRI studies where scan times are pretty high. One such system called the FastAI model was developed based on a neural network working on MRI scans. The system reduces scan time and thereby minimizes motion related artefacts; reduces incidences of a repeat scan.⁶⁰

6.3. Preoperative planning

The DL based segmentation algorithms may have a role to play in automatic segmentation. A study by Zeng G. et al.,⁶¹ performed automatic segmentation of MRI-based 3D models utilising DL, which was at par in terms of accuracy with the CT-based 3D models for patients with hip diseases of childbearing age. Thus, DL enabled improved usage of MRI, which permits radiation-free and patient-tailored preoperative simulation and surgical planning of periacetabular osteotomy in cases suffering from developmental dysplasia of the hip.

6.4. Smart reporting tools for comprehensive fracture reporting

The number of radiological examinations being performed each year continues to increase each year with the US alone accounting for >1 billion examinations each year. The early 2000s saw a growth of over 70% to the 1990s level and subsequently a four-year growth of nearly 7%. During the same period, there has been a decline in medical insurance-related reimbursements putting pressure on existing radiologists to improve their work output, deal with larger case volume while working under the same time constraints. In an experiment performed by Sokolovskaya et al.⁶² to judge the impact of hastening the speed of reporting by radiologists, the results were adverse with an increase of misinterpretation by 16% from the

baseline error rate of 10% in five radiologists. Clearly, with this new decade, the pressures will only increase, and so will the associated risk of missed diagnosis, misinterpretation. This added pressure can be offset by making use of technological advancements, which can help with rapid reporting while ensuring standardised reporting as demonstrated by the Smart Reporting Tool. The road further calls for incorporating AI into the routine reporting of radiologists and assistance for clinicians as a part of quality assurance. To make this possible, the system needs to be integrated with existing PACS so that it can actually serve the end-user. The benefits of smart reporting tool can be potentially multi-fold with the key aspect serving structured quantified and standardized reporting. An example of the smart reporting tool is showcased in [fig. 7](#).

6.5. Bone fragility assessment

A combination of well-trained algorithms can improve the assessment of bone strength and thus quantify a parameter which is directly associated with a risk of future fractures. Most scope lies in the usage of trabecular bone architecture and ML-based automated segmentation of structural assessment of the bone on MRI scans.⁵⁵

6.6. Radiomics

This is a yet emerging field which centres around quantify various characteristics from image and using them for data mining and pattern identification. Currently they are being primarily tested on detecting characteristics of the neoplastic bony or soft tissue mass. The characteristics are not apparent to eye and thus difficult to interpret by humans, such as analysis of texture, histograms and image-voxel relationships. These characteristics have an impact on the treatment approach and are hence of extreme relevance clinically. Since, the field is still novel, the applications are likely to grow multi-fold and research is already underway to establish the extent of its usage to quantify bone mineral density loss and thus quantify the risk of future fractures. The envelop can be expanded to assess the likelihood of fracture union based on various CT/MRI characteristics.^{10,40,55}

7. Challenges-

7.1. Dataset availability

Challenges that impede the development of AI include the availability of large, diverse, labelled and curated datasets. Most of the available datasets are for more common conditions (limited types of fractures, little joints), while the data for the infrequent locations remain relatively negligible. A likely cause for such data imbalance is an overall lack of incentivisation and support. While many researchers are working on a variety of DL systems, they need big data to test and refine their systems to churn out performance which can be at par or can even supersede experts. As elaborated in this review article earlier and also a quick glance at [table-1](#), the issue of the majority of the DL systems being developed on data from a single institute is blatantly visible. So, despite having high functioning systems, their applicability and generalisability to the

larger population are still questionable. This calls for an initiative by the MSK societies to promote multi-institutional studies either for AI development or for pooling in data from multiple hospitals and making this accessible to the general public and developers.

To partially offset the current problem, other methods of obtaining data, like creating synthetic images too are being exploited. Perhaps, the most significant development of recent times in this field was the development of the general adversarial network (GAN). It consists of two types of neural networks—a generative network G and a discriminative network D.⁶³ They have a symbiotic function; that is, while G functions to synthesize images, D functions to discriminate these synthesized images from actual images. An essential development towards this end was the development of GANs by Chuquicusma et al. to create fake lung cancer nodules and incorporate these into real CT images and test them on radiologists. These fake lung nodules turned out to be difficult to differentiate from real ones.⁶⁴ The importance lies in being able to produce more data to train the deep classifiers and thus improvise their accuracy.

7.2. Domain shift

One major challenge posed by the presently developed ML and DL models is the problem of domain shift. Domain shift implies that a model does not perform well in a setting other than the one in which it was trained and validated. The model is not robust enough to be generalisable to data from different hospitals and as a result its real-world utility is poor. The problem of domain shift is faced by ML models in all fields and the primary cause of this problem is that the ML models makes spurious associations. To minimise such a development, it is important that in the training phase, data should be collected prospectively from multiple sources and cleaned before being run through the system. Secondly, the model should focus on medically relevant features and not other confounding variables. Towards this end the clinician and computer science experts need to work closely in sync to pre-emptively identify the possible confounding variables and come up with ways to neutralise them.⁶⁵

7.3. The law and the prospect of liability

A lot of the current scenarios fall in the gray zone as law is not very clearly defined on a number of situations pertaining to usage of AI and diagnosis reliant on feedback from AI.

Though algorithms mention the scope for false positive and false negative clearly, one cannot miss the fact that algorithm can miss a relevant finding on the scan for which it is designed. This can potentially impact patient management. The way AI is going to influence medical imaging is by aiding experts quickly find relevant findings in images saving them time and effort. AI is unlikely to compete with experts and here is where there is significant scope for augmenting imaging practises. The existing laws recommend that clinicians who do rely on interpretation of the scan by AI should be individually capable of making the diagnosis, such that AI provides more of assistance and is not the sole authority to dish out diagnosis.⁶⁶

A significant number of tools used clinically are developed by organization or institutes need to invest significantly in the algorithmic design and development process. The algorithmic performance and design and bias needs proper assessment. The key challenges of AI is its “Black-Box” nature as a result clinician may have no way of assessing how the algorithm came to a conclusion and hence, no reasonable way of questioning the algorithms approach.⁶⁷ To ensure that algorithms performance can be validated the solution should be bench marked on a set of test dataset.

Such an appropriate system can allow transparent assessment of algorithms in a blind manner. This can help improve confidence of clinician in the AI solution at the disposal and ensuring the chances of error are minimal and all scenarios where the algorithm is likely to err or fail is documented.

While creating such independent test datasets by global bodies focussed on particular disease or condition. There should be funding of these bodies through Government support or non-Governmental organization to ensure that there is transparent assessment and there is focus on explainability of results. The European Union recently passed into law Article-22 (4) which mandates explainability of the results; other nations too have taken note and are likely to follow-suit.⁶⁸ The most basic of steps would be to have in place bounding boxes and heat maps, which clearly demarcate the region on the radiograph or CT suspected to have fractures, neoplasia. Different colours can be used to highlight different features of the relevant disease, such a display makes it easy for clinician to then use their clinical experience to assess the same.

8. Conclusion

AI in MSK trauma is a new and upcoming field with a lot of promise. Many of the solutions are point solutions that can detect fractures, complications like bleed, pneumothorax, blurring of fat planes. Many studies show that imaging experts and clinicians assisted with AI perform better than those without assistance. The key to AI adoption for trauma is based on how it's incorporated in radiology workflows. The adoption of AI in radiology practises likely to reduce burnout and minimise errors. The tool has immense potential and likely to be as good as experts in the near future as the AI will use information from multiple sources to triage and recommend a diagnosis based on key clinical information it may have in hand which the expert may not have time to assimilate during the time of review of the image or study, which the AI tool has already assimilated. The way this tool is likely to work is to be a virtual assistant to the experts empowering them with the right information and at the right time enabling faster analysis.

Conflicts of interest to declare

•Dr Amit Kharat is a professor in the Department of Radiology at Dr D.Y. Patil Medical College, Hospital and Research Center, DPU, Pune, India and Co-founder of DeepTek Inc.

•Dr Rajesh Botchu and Dr Harun Gupta, both contributed to the article and are also on the editorial board of the journal.

Funding source

No funding to disclose.

References

1. Kraaijvanger N, Rijpsma D, van Leeuwen H, Edwards M. Self-referrals in the emergency department: reasons why patients attend the emergency department without consulting a general practitioner first—a questionnaire study. *Int J Emerg Med.* 2015 Dec;8(1):1–6.
2. Riggan CN, Morris TR, Soslowsky LJ. *Tendinopathy II: Etiology, Pathology, and Healing of Tendon Injury and Disease. In Tendon Regeneration.* Academic Press; 2015 Jan 1:149–183.
3. *Musculoskeletal Conditions (who.Int).* 2021. Accessed on 21st June.
4. Firestein Gary S. *Chapter-1, Kelley and Firestein's Textbook of Rheumatology.* tenth ed. Elsevier Science; 2017.
5. Stepnick LS. The frequency of bone disease. *Bone health and osteoporosis: A Report of the Surgeon General.* 2004:68–87.
6. Riggs BL, Melton Iii LJ. The worldwide problem of osteoporosis: insights afforded by epidemiology. *Bone.* 1995 Nov 1;17(5):S505–S511.
7. Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A.

- Incidence and economic burden of osteoporosis-related fractures in the United States, 2005–2025. *J Bone Miner Res.* 2007 Mar 1;22(3):465–475.
8. Houshian S, Larsen MS, Holm C. Missed injuries in a level I trauma center. *Journal of Trauma and Acute Care Surgery.* 2002 Apr 1;52(4):715–719.
 9. Pfeifer R, Pape HC. Missed injuries in trauma patients: a literature review. *Patient Saf Surg.* 2008 Dec;2(1):1–6.
 10. Owoyemi A, Owoyemi J, Osiyemi A, Boyd A. Artificial intelligence for healthcare in Africa. *Frontiers in Digital Health.* 2020 Jul 7;2:6.
 11. Kharat A, Duddalwar V, Saoji K, et al. Role of Edge Device and Cloud Machine Learning in Point-Of-Care Solutions Using Imaging Diagnostics for Population Screening. arXiv preprint arXiv:2006.13808. 2020 Jun 18.
 12. Junaid S, Saeed A, Yang Z, Micic T, Botchu R. Developing an artificial intelligence project in your radiology department. *Indian J Musculoskeletal Radiol.* 2020;2(1):58–61.
 13. <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>. Accessed on 24th June, 2021.
 14. *A Brief History of Machine Learning - DATAVERSITY.* 2021. Accessed on 24th June.
 15. What Is Machine Learning? (cognizantsoftvision.Com). Accessed on 25th June, 2021.
 16. Mathew A, Amudha P, Sivakumari S. Deep Learning Techniques: An Overview. In International Conference on Advanced Machine Learning Technologies and Applications 2020 Feb 13 (pp. 599–608). Springer, Singapore.
 17. A Tour of Machine Learning Algorithms (machine learning mastery.Com). Accessed on 25th June, 2021.
 18. Kersting K. Machine learning and artificial intelligence: two fellow travelers on the quest for intelligent behavior in machines. *Frontiers in big Data.* 2018 Nov 19;1:6.
 19. Pankhania M. Artificial intelligence in musculoskeletal radiology: past, present, and future. *Indian Journal of Musculoskeletal Radiology* • Volume. 2020 Jul;2(2):89.
 20. Cabani A, Hammoudi K, Benhabiles H, Melkemi M. Masked Face-Net—A dataset of correctly/incorrectly masked face images in the context of COVID-19. *Smart Health.* 2021 Mar 1;19:100144.
 21. Roh Y, Heo G, Whang SE. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Trans Knowl Data Eng.* 2019 Oct 8.
 22. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol.* 2018 May 1;73(5):439–445.
 23. Gale W, Oakden-Rayner L, Carneiro G, Bradley AP, Palmer LJ. Detecting hip fractures with radiologist-level performance using deep neural networks. arXiv preprint arXiv:1711.06504. 2017 Nov 17.
 24. Chen HY, Hsu BW, Yin YK, et al. Application of deep learning algorithm to detect and visualise vertebral fractures on plain frontal radiographs. *PLoS One.* 2021 Jan 28;16(1):e0245992.
 25. Ma Y, Luo Y. Bone fracture detection through the two-stage system of crack-sensitive convolutional neural network. *Informatics in Medicine Unlocked.* 2021 Jan 1;2:100452.
 26. Olczak J, Emilson F, Razavian A, Antonsson T, Stark A, Gordon M. 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.* 2020 Oct 25;92(1):102–108.
 27. Li YC, Chen HH, Lu HH, Wu HT, Chang MC, Chou PH. Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists?. *Clinical Orthopaedics and Related Research*®. 2021 May 28;10–97.
 28. Tobler P, Cyriac J, Kovacs BK, et al. AI-based detection and classification of distal radius fractures using low-effort data labeling: evaluation of applicability and effect of training set size. *Eur Radiol.* 2021 Mar 19;1–9.
 29. Chedid N, Sadda P, Gonchigar A, et al. Synthesis of fracture radiographs with deep neural networks. *Health Inf Sci Syst.* 2020 Dec;8:1–0.
 30. Jin L, Yang J, Kuang K, et al. Deep-learning-assisted detection and segmentation of rib fractures from CT scans: development and validation of FracNet. *EBio-Medicine.* 2020 Dec 1;62:103106.
 31. Pranata YD, Wang KC, Wang JC, et al. Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. *Comput Methods Progr Biomed.* 2019 Apr 1;171:27–37.
 32. Armon K, Stephenson T, Gabriel V, et al. Determining the common medical presenting problems to an accident and emergency department. *Arch Dis Child.* 2001 May 1;84(5):390–392.
 33. Gergenti L, Olympia RP. Etiology and disposition associated with radiology discrepancies on emergency department patients. *The American journal of emergency medicine.* 2019 Nov 1;37(11):2015–2019.
 34. Kijowski R. Clinical cartilage imaging of the knee and hip joints. *Am J Roentgenol.* 2010 Sep;195(3):618–628.
 35. Chen HY, Hsu BW, Yin YK, et al. Application of deep learning algorithm to detect and visualise vertebral fractures on plain frontal radiographs. *PLoS One.* 2021 Jan 28;16(1):e0245992.
 36. Jones RM, Sharma A, Hotchkiss R, et al. Assessment of a deep-learning system for fracture detection in musculoskeletal radiographs. *NPJ digital medicine.* 2020 Oct 30;3(1):1–6.
 37. Badgeley MA, Zech JR, Oakden-Rayner L, et al. Deep learning predicts hip fracture using confounding patient and healthcare variables. *NPJ digital medicine.* 2019 Apr 30;2(1):1–0.
 38. Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci Unit States Am.* 2018 Nov 6;115(45):11591–11596.
 39. Chung SW, Han SS, Lee JW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop.* 2018 Jul 4;89(4):468–473.
 40. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol.* 2018 May 1;73(5):439–445.
 41. Kazi A, Albarqouni S, Sanchez AJ, et al. Automatic classification of proximal femur fractures based on attention models. In International Workshop on Machine Learning in Medical Imaging 2017 Sep 10 (pp. 70–78). Springer, Cham.
 42. Meng XH, Wang Z, Ma XL, Dong XM, Liu AE, Chen L. A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance. *Skeletal Radiol.* 2021 Feb 18:1–8.
 43. Dreizin D, Zhou Y, Fu S, et al. A multiscale deep learning method for quantitative visualisation of traumatic hemoepitomeum at CT: assessment of feasibility and comparison with subjective categorical estimation. *Radiology: Artif Intell.* 2020 Nov 11;2(6):e190220.
 44. Zhou QQ, Wang J, Tang W, et al. Automatic detection and classification of rib fractures on thoracic CT using convolutional neural network: accuracy and feasibility. *Korean J Radiol.* 2020 Jul;21(7):869.
 45. Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput Biol Med.* 2018 Jul 1;98:8–15.
 46. Kim M, Park HM, Kim JY, Kim SH, Hoeke S, De Neve W. MRI-based diagnosis of rotator cuff tears using deep learning and weighted linear combinations. In Machine Learning for Healthcare Conference 2020 Sep. 18 (pp. 292–308). PMLR.
 47. Chee CG, Kim Y, Kang Y, et al. Performance of a deep learning algorithm in detecting osteonecrosis of the femoral head on digital radiography: a comparison with assessments by radiologists. *Am J Roentgenol.* 2019 Jul;213(1):155–162.
 48. Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. *PLoS Med.* 2018 Nov 27;15(11):e1002699.
 49. Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med.* 2018 Nov;80(5):2139–2154.
 50. Cunningham R, Sánchez MB, May G, Loram I. Estimating full regional skeletal muscle fibre orientation from B-mode ultrasound images using convolutional, residual, and deconvolutional neural networks. *Journal of Imaging.* 2018 Feb;4(2):29.
 51. Baka N, Leenstra S, van Walsum T. Ultrasound aided vertebral level localisation for lumbar surgery. *IEEE Trans Med Imag.* 2017 Aug 10;36(10):2138–2147.
 52. Hong G, Zhang L, Kong X, Herbert L. Artificial intelligence image-assisted knee ligament trauma repair efficacy analysis and postoperative femoral nerve block Analgesia effect research. *World Neurosurgery.* 2021 May 1;149:492–501.
 53. Liu F, Guan B, Zhou Z, et al. Fully automated diagnosis of anterior cruciate ligament tears on knee MR images by using deep learning. *Radiology: Artif Intell.* 2019 May 8;1(3):180091.
 54. Astuto B, Flament I, K. Namiri N, et al. Automatic deep learning-assisted detection and grading of abnormalities in knee MRI studies. *Radiology: Artif Intell.* 2021 Jan 20;3(3):e200165.
 55. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *Am J Roentgenol.* 2019 Sep;213(3):506–513.
 56. Shin Y, Yang J, Lee YH, Kim S. Artificial intelligence in musculoskeletal ultrasound imaging. *Ultrasonography.* 2021 Jan;40(1):30.
 57. Henderson RE, Walker BF, Young KJ. The accuracy of diagnostic ultrasound imaging for musculoskeletal soft tissue pathology of the extremities: a comprehensive review of the literature. *Chiropr Man Ther.* 2015 Dec;23(1):1–29.
 58. Huang C, Zhou Y, Tan W, et al. Applying deep learning in recognising the femoral nerve block region on ultrasound images. *Ann Transl Med.* 2019 Sep;7(18).
 59. Burns JE, Yao J, Summers RM. Artificial intelligence in musculoskeletal imaging: a paradigm shift. *J Bone Miner Res.* 2020 Jan;35(1):28–35.
 60. Recht MP, Zbontar J, Sodickson DK, et al. Using deep learning to accelerate knee MRI at 3 T: results of an interchangeability study. *Am J Roentgenol.* 2020 Dec;215(6):1421–1429.
 61. Zeng G, Schmaranzer F, Degonda C, et al. MRI-based 3D models of the hip joint enables radiation-free computer-assisted planning of periacetabular osteotomy for treatment of hip dysplasia using deep learning for automatic segmentation. *European journal of radiology open.* 2021 Jan 1;8:100303.
 62. Sokolovskaya E, Shinde T, Ruchman RB, et al. The effect of faster reporting speed for imaging studies on the number of misses and interpretation errors: a pilot study. *J Am Coll Radiol.* 2015 Jul 1;12(7):683–688.
 63. Chedid N, Sadda P, Gonchigar A, et al. Synthesis of fracture radiographs with deep neural networks. *Health Inf Sci Syst.* 2020 Dec;8:1–0.
 64. Chuquicuma MJ, Hussein S, Burt J, Bagci U. How to fool radiologists with generative adversarial networks? A visual Turing test for lung cancer diagnosis. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) 2018 Apr 4 (pp. 240–244). IEEE.
 65. Machine-learning Models that Detect COVID-19 on Chest X-Rays Are Not Suitable for Clinical Use – Physics World. Accessed on 3rd July, 2021.
 66. Harvey HB, Gowda V. Clinical applications of AI in MSK imaging: a liability

- perspective. *Skeletal Radiol.* 2021 Apr 9:1-4.
67. Foss-Solbrekk K. Three routes to protecting AI systems and their algorithms under IP law: the good, the bad and the ugly. *J Intellect Property Law Pract.* 2021 Mar;16(3):247-258.
68. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf).