Artificial Intelligence Application in Bone Fracture Detection

Ahmed AlGhaithi, Sultan Al Maskari

Division of Orthopedic, Department of Surgery, Sultan Qaboos University Hospital, Muscat, Oman

Abstract

The interest of researchers, clinicians, and industry in artificial intelligence (AI) continues to grow, especially with recent deep-learning (DL) advances. Recent published reports have shown the utility of DL for bone fracture diagnosis in the radiological examination. It is important for practicing physicians to recognize the current scope of DL as it may impact the clinical practices in the near future. This article will give an insight to the practicing clinician of the current advances in AI fracture diagnosis by reviewing the current literature on this participant. Electronic databases were searched for relevant articles relating to AI applications in bone fracture detection. We included all published work in PubMed, Medline, and Cross-references, which satisfied the inclusion criteria. The search identified 104 references. Of those, 13 articles were eligible for the analysis. AI advancements in fracture imaging applications can be divided into the categories of fracture detection, classification, segmentation, and noninterpretive tasks. Despite the potential work presented in the literature, there are many challenges in the form of clinical translation and its widespread uses. These challenges range from the proof of safety to clearance from the regulatory agencies.

Keywords: Artificial Intelligence, convolutional neural networking, deep learning, fracture imaging, machine learning, musculoskeletal

Introduction

Bone fractures are among the most common causes of emergency department visits. Diagnostic errors often occur due to misinterpretation of radiological examination, which may lead to the delayed treatment and poor outcomes.[1] The analysis of causes of fracture diagnostic inaccuracies has found them to be multifactorial, including physician factors, image quality, insufficient clinical information, fracture type, and polytrauma.[2] Four out of five diagnostic errors in an emergency settings are due to physician factors, yet radiographs are often interpreted by clinicians who lack the required specialized expertise.[3] Even with an experienced radiologist, physician fatigue and error may increase during a long busy day, increasing the risk of missing a subtle fracture.[4] Thus, a model that can offer assistance to physicians presenting second opinions through highlighting concerning areas in imaging examination may produce more efficient interpretation, standardize quality, and decrease errors. With recent advances in deep learning (DL) and computer vision, artificial intelligence (AI) may play a significant role in this field.

AI is a powerful technology that has demonstrated good potential at radiographic image interpretation. While earlier levels of AI performance were subhuman, modern versions are able to match or even surpass humans' performance.[5] AI has also shown promising results in complex diagnostics in other medical specialties such as ophthalmology, dermatology, and pathology.[6] The aim of this article is to explore the potential of utilizing AI in fracture diagnosis by reviewing the current literature on this subject.

Technical Aspects

AI, machine learning (ML), DL, and convolutional neural networking (CNN) are terminology, which often used interchangeably [Figure 1]. AI refers to any skill where a machine performs tasks that mimic human intelligence. ML is a subfield of AI that enables a machine to learn and improve from the experience independently of human action. DL is a more specialized subfield of ML, which can analyze more data sets transforming the inputs of an algorithm into outputs using

Address for correspondence: Dr. Ahmed AlGhaithi,
P. O. Box 478 P.C 130, Muscat, Sultanate of Oman.
E-mail: a.alghaithi@squ.edu.om

Received: 26-11-2020 **Revised:** 18-01-2021
Accepted: 26-01-2021 **Published Online:** 20-02-2021

Access this article online

Quick Response Code:	
	Website: www.journalmsr.com
	DOI: 10.4103/jmsr.jmsr_132_20

How to cite this article: AlGhaithi A, Al Maskari S. Artificial intelligence application in bone fracture detection. J Musculoskelet Surg Res 2021;5:4-9.

the sophisticated computational models such as deep neural networks. CNN is evolutional computational technique of DL, which can impact the key areas of medicine such as medical imaging.[7] CNN is built of computational units called nodes, which are analogous to biological brain neurons. Each node takes one or more weighted input connections and performs mathematical operations resulting in outputs that can pass to other connected nodes.

Material and Data Source

Online databases (PubMed and MEDLINE) search was carried to find the literature related to AI use in fracture diagnosis. The search was carried accordance to preferred reporting items for systematic reviews and meta-analyses statement. Keywords included "artificial intelligence," "deep learning," "machine learning," and "fracture." Searches were conducted on April 1, 2020, yielding a total of 104 articles from the two databases, without applying any restriction on language or date of publication [Figure 2]. An independent reviewer performed screening of articles' titles and abstracts in the first reviewing stage, in addition to the titles and abstracts of crossover references. The following inclusion criteria were used: all levels of evidence and studies on humans. We did not place restrictions on the target population, the outcome of the disease of interest, or the intended context for using the model. We excluded from the search nontraumatic musculoskeletal pathologies and conferences abstracts due to incomplete data presentation.

Results

The search terms, as described above, identified 216 references [Figure 2]. After duplicate removal, 104 articles titles and abstracts were screened. Of these 19 full-text articles were assessed independent by both authors for analysis eligibility, finally 13 studies satisfied all the inclusion and exclusion criteria. A complete list of included published work is provided in Table 1. The application of AI in fracture imaging can be classified into four major categories: Pathology detection (e.g., calcaneus fracture), segmentation (which means automated segmentation of the region of interest whereby the irrelevant pixels are cropped out and would not influence the training process e.g., cropping out soft tissue), classification (e.g., calcaneal fracture classification), noninterpretive (e.g., image-quality improvement from under-sampled magnetic resonance imaging or low-dose computed tomography [CT]).[5]

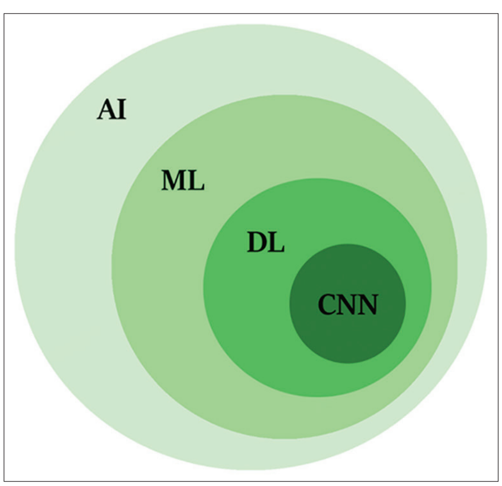

Figure 1: Shows the relationships of artificial intelligence, machine learning, deep learning, and convolutional neural network

Upper Limbs Fractures

The rate of missing a fracture between the upper and lower extremity is almost analogous. Upper limb fractures most likely to be missed are elbow (6%), hand (5.4%), wrist (4.2%), and shoulder (1.9%).[8] Kim and MacKinnon trained a model using 1112 images of wrist radiographs, then they added additional 100 images for final testing and analysis (comprising 50 fractures and 50 normal). The area under the curve (AUC) was 0.954, with a diagnostic sensitivity of 90% and 88% specificity.[9] Lindsey *et al.* developed another CNN model for detecting wrist fractures using 135,409 radiographs and was able to improve the sensitivity of clinicians' image reading from 88% unaided to 94% aided, and and by doing so, misinterpretation improved by 53%.[10] Olczak *et al.* designed an algorithm for distal radius fractures and tested it on hand and wrist radiographs. They compared the network performance with two experienced orthopedic surgeons and showed a high

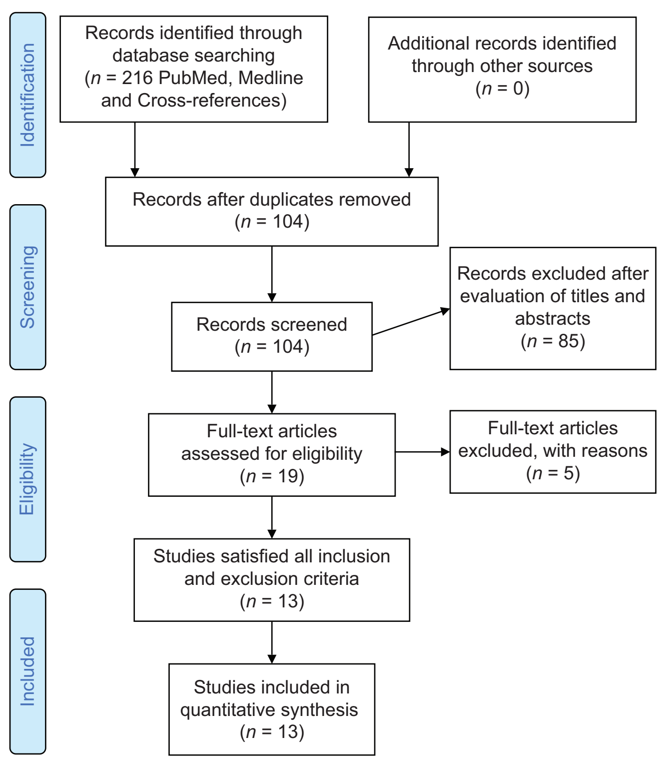

Figure 2: Preferred reporting items for systematic reviews and meta-analyses flow diagram for study selection

Table 1: Classification of artificial intelligence application in view of body part fracture

Reference	Anatomic area	Module purpose	Modality	Compared to human expert performance	Performance (metric)
Kim *et al*. 2018	Wrist	Diagnosis	Radiographs	No	Provided proof of concept in fracture detection on plain radiographs 0.95 (AUC), 90% sensitivity and 88% specificity
Olczak *et al*. 2017	Hand/wrist/ankle	Diagnosis	Radiographs	Yes	Performance in detecting fractures from hand/wrist/ankles radiograph sensitivity of 90% and specificity of 88%accuracy of 83% versus. radiologists, 82%
Lindsey *et al*. 2018	Wrist	Diagnosis	Radiographs	Yes	Improved clinicians image reading sensitivity from 88% unaided compared to 94% aided
Chung *et al*. 2018	Proximal humerus	Diagnosis and classification (Neer)	Radiographs	Yes	Diagnosis accuracy of 96%, 99% sensitivity, 97% specificity Classification accuracy range between 65% and 86%, sensitivity 88% to 97%, specificity 83% to 94% (dependent on the type)
Rayan *et al*. 2019	Pediatrics elbow fractures	Diagnosis	Radiographs	No	The model accuracy was 88% with sensitivity of 91% and specificity of 84%
Urakawa *et al*. 2019	Intertrochanteric hip fractures	Diagnosis	Radiographs	Yes	Convolutional neural network outperformed orthopedic surgeons at detecting, accuracies of 96% versus 92%, specificities of 97% versus 97%. 57 and sensitivities of 94% versus 88%
Cheng *et al*. 2019	Hip fracture	Diagnosis	Radiographs	No	Accuracy of 91%, a sensitivity of 98%, AUC of 0.98
Adams *et al*. 2019	Neck of femur	Diagnosis	Radiographs	Yes	Accuracy of 91%, AUC 0.98 Performing junior's physician increased from 87.6% to 90.5%
Balaji *et al*. 2020	Femur diaphyseal	Diagnosis	Radiographs	No	Accuracy of 90.69% with 86.66% sensitivity and 92.33% specificity
Kitamura *et al*. 2019	Ankle	Diagnosis	Radiographs	No	Model with multiple views shown improved accuracy in fracture detection of 81% compared with single view of 76%
Pranata *et al*. 2019	Calcaneus	Classification (Sander)	CT	No	Sanders classification system model accuracy 98%
Rahmaniar *et al*. 2019	Calcaneus	Classification (Sander)	CT	Yes	An accuracy of 86% with computational performance of 133 frame per second
Burns *et al.* 2017	Spine	Diagnosis	CT	No	Model which detect, localize, classify the fractures and measure bone density vertebral bodies employing more lumbar and thoracic CT images. Attained sensitivity was 95.7%
Tomita *et al*. 2018	Spine	Diagnosis	CT	No	Model which detect osteoporotic vertebral fractures achieved an accuracy of 89.2%
Muehlematter *et al*. 2019	Spine		CT	No	Accuracy of classifying of unstable/stable vertebrae was low with AUC 0.53

AUC: Area under the curve, CT: Computed tomography

detection rate with a sensitivity of 90% and specificity of 88%.[11] They did not specify the type of fractures or grade of difficulty of fracture detection.

Chung *et al.* trained a CNN model to detect the fractures of proximal humerus and classify the type of fracture (four parts Neer's classification) on a dataset of 1891 anteroposterior shoulder radiographs. The model showed a high throughput precision of 96% and a mean AUC of 1.00 compared to specialists, with a sensitivity of 99% and a specificity of 97%. However, the task of classifying the fracture was more challenging; the reported accuracy was ranging from 65% to 85%. The model showed superior performance accuracy compared to general physicians and orthopedic surgeons and almost similar performance to specialized shoulder surgeons.[12]

Rayan *et al.* developed a model with a multi-view approach, which mimics the human radiologist when reviewing multiple images of acute pediatric elbow fractures. They used 21,456 radiographic studies containing 58,817 elbow radiographs. The model accuracy was 88%, with a sensitivity of 91% and specificity of 84%.[13]

Lower Limbs Fractures

Hip fractures constitute 20% of patients admitted to orthopedic surgery, while the incidence of occult fractures on radiographs ranges from 4% to 9%.[14] Urakawa *et al.* developed CNN to study intertrochanteric hip fractures in a total of 3346 hip images (1773 fractured and 1573 nonfractured hip images). His model was compared to the performance of five orthopedic surgeons and showed accuracy of 96% versus 92%, specificities of 97% versus 57% and sensitivities of 94% versus 88%.[15] Cheng *et al.* developed CNN algorithm, which was pretrained using 25,505 limb radiographs. Achieved algorithm accuracy for diagnosing hip fracture is 91%, sensitivity is 98%. The performance achieved has a low false-negative rate of 2%, which make it a good screening tool.[16] Adams *et al.* developed a model to detect the neck of femur fracture with an accuracy of 91% and AUC 0.98.[17] Balaji *et al.* developed CNN to diagnose femur diaphyseal fractures. The model was developed using 175 radiographs (100 normal and 75 fractured). Then trained to classify the type of diaphyseal femur fracture, namely transverse, spiral, and comminuted. The achieved highest accuracy of 90.7% with 86.6% sensitivity and 92.3% specificity.[18]

Missed ankle and foot fractures are common, especially in trauma patients. Some reports estimated missed diagnosis due to different reasons in the initial contact may reach up to 44%, of which 66% were due to radiological misdiagnosis.[19] This is why researchers tried to train models for this purpose. Kitamura *et al.* developed CNN of a small number of ankle radiographs (298 normal and 298 fractured ankles). The model was trained to detect ankle fractures, where ankle fracture was defined as proximal forefoot, midfoot, hind foot, distal tibia, or distal fibula. The model with multiple views has shown improved accuracy in fracture detection from 76% to 81%.[20] Pranata *et al.* proposed two types of CNN algorithms for the classification of calcaneal fractures using CT images using the Sanders classification system. The proposed algorithm exhibited 98% accuracy, which makes it a viable tool for future use in computer-assisted diagnosis.[21] Rahmaniar and Wang developed a computer-aided method for calcaneal fracture detection in CT. Sanders system was also used for fracture classification, where calcaneus fragments were detected and marked by color segmentation. The achieved performance accuracy was high (86%), with a computational performance of 133 frames per second.[22]

Spine Fractures

The incidence of misdiagnosed spine fractures varies among studies and ranges from 19.5% to 45%.[23] Burns *et al*. was able to detect, localize, classify vertebral spine fractures as well as measure bone density of vertebral bodies using lumbar and thoracic CT images. Achieved sensitivity was 95.7% and a false-positive rate of 0.29 per patient for compression fractures detection and localization.[24] Tomita *et al.* developed CNN to extract radiological features of osteoporotic vertebral fractures in CT scan. The model was trained using 1432 CT scans, comprised of 10,546 sagittal views, and achieved an accuracy of 89.2%. The product algorithm was then tested on 128 spine CT scans and an accuracy of 90.8% was achieved.[25] Muehlematter *et al.* proposed algorithms to detect vertebrae at risk of fracture using 58 CT scans of patients with acquired fractures due to vertebral insufficiency. One hundred and twenty items (60 stable vertebrae and 60 unstable vertebrae) were included in the study. However, the grading accuracy of unstable/stable vertebrae was low with AUC of 0.5.[26]

Discussion

The efficacy of AI compared to human's intelligence is emerging as an effective tool to address the current blemishes of human errors. The AI current status of the technology can be described by Gartner's hype cycle [Figure 3], which defines how a technology, or an innovation progresses through its life cycle from concept to widespread adoption.[27] The cycle consists of five phases: The first phase is a "technology trigger" where only technology is envisioned, followed by a "peak of inflated expectations phase," where the technology profile is raised with successful and unsuccessful trials. Then, it is followed by the "trough of disillusionment phase" at which defects in the technology cause disappointment in its effectiveness, followed by the "slope of enlightenment" as companies begin to test it in their own environments. The final phase is the "plateau of productivity," where technology is available in the market.[25] AI in medical applications, and fracture detection specifically, is still in the early phases of this cycle and fall at the peak of the inflated expectation phase as more reports continue to demonstrate the efficiency of AI in detecting fractures.[7] Currently, the work published in the field of orthopedic traumatology to date is small collective initiatives, trying to get proof of concept rather than applying technology.

The objective of integrating AI into the clinical practice is to augment the workflow at clinical environment rather than replacing the workforce. Thus, with the evaluation of new computing platforms and the development of new algorithm models, the new generation of AI is anticipated to advance the quality of workflow in several ways namely improving the experience of care, the diagnoses, minimizing the errors, improving time management, and reducing costs.[5] One of the greatest challenges, which can be improved by AI is accurate

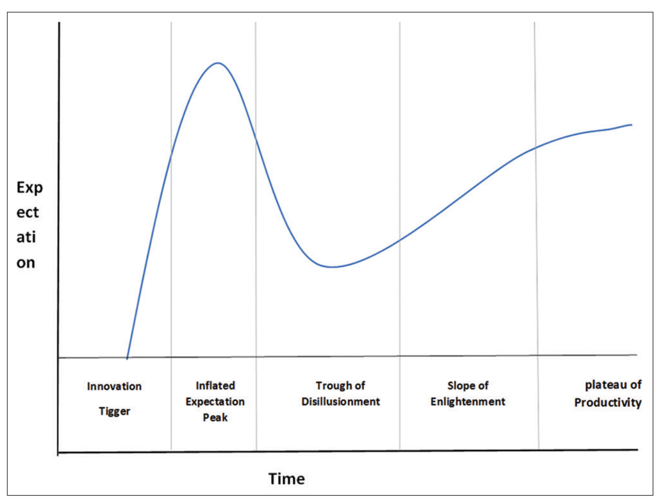

Figure 3: Gartner's hype cycle provides a graphic illustration of the maturity and deployment of technologies and applications

radiological diagnosis, especially in an emergency setting by inexperienced or exhausted clinicians. Therefore, the aid of AI in the fracture detection is more important in augmenting workflow compared to segmentation or classification.[9] For example, assisting AI in diagnosing difficult fractures such as elbow fracture in children will have a greater impact on the treatment outcomes compared to classifying the type of fracture.

By integrating AI into the clinical setting, AI is expected to provide clinicians with better clinical insights needed to reduce the errors and improve the quality of task interpretation. Another important aspect where AI can play a major role is official reporting systems after office hours. AI should support a reporting system for an examination performed in hospitals where the radiologist is not attending in person.[7]

The expectation from the latest AI tools is to demonstrate the state-of-the-art results. It should improve workload and increase daily productivity by replacing the manual retrieval of image data from a database to suggest a comparison with new images, or even for audits and clinical studies. Moreover, AI should drive efficient worklist prioritization in the work environment, communicating the important image analysis and ensure automatic assignment to the most appropriate available physician.[4]

Limitations and Challenges of Artificial Intelligence in the Clinical Setting

AI remains far from independently operating in a clinical setting. In the face of many successful implementations of AI models, application limitations must be recognized. Published works are of an experimental nature and are not incorporated into daily clinical practice, which may show the feasibility and efficacy of proposed diagnostic models. Added to that even, the published works are challenging to be reproduced, because most training data sets and codes are rarely published. Moreover, the proposed models need to be integrated within clinical information software as well as Picture Archiving and Communications Systems in order to be useful. However, until now, very limited data present this type of integration.[5] Moreover, the safety demonstration of these models to regulatory agencies is an important step for clinical translation and widespread uses. However, there is no denying that AI is making rapid progress and great improvements.[4,6]

In general, new generations of DL and in particular CNN have successfully demonstrated to be more accurate and rapidly developed with innovative results than earlier generations. These approaches are now diagnostically accurate and are predicted to outperform human experts in the future. It would also potentially give a more precise diagnosis to patients. In general, to be able to interpret and use artificial intelligence correctly, physicians must have a clear understanding of the tools used on AI. Taking in account the challenges standing in the way of clinical translation and widespread uses. These challenges range from proof of safety to clearance from regulatory agencies.

Conclusion

Several AI models demonstrated certain performance at the expert level. Although the comprehensive interpretation of the image has not been achieved yet, it is too early to consider AI operating independently in a clinical setting. However, with the current technology, AI has the potential to be considered to augment the efficiency of clinical workflow.

Ethical approval

The authors confirm that this review had been prepared in accordance to COPE roles and regulation. Given the nature of the review, IRB review was not required.

Financial support and sponsorship

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

There are no conflicts of interest.

Authors contributions

AAG contributed to developing the project idea, searched the literature and interpretation of the results and preparation and revising the manuscript; SAM contributed in developing the idea and critically revising the manuscript. All authors have critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

References

1. Pinto A, Reginelli A, Pinto F, Lo Re G, Midiri F, Muzj C, *et al*. Errors in imaging patients in the emergency setting. Br J Radiol 2016;89(1061):20150914.
2. Pinto A, Berritto D, Russo A, Riccitiello F, Caruso M, Belfiore MP, *et al*. Traumatic fractures in adults: Missed diagnosis on plain radiographs in the Emergency Department. Acta Biomed 2018;89:111-23.
3. Hallas P, Ellingsen T. Errors in fracture diagnoses in the emergency department–characteristics of patients and diurnal variation. BMC Emerg Med 2006;6:4.
4. Krupinski EA, Schartz KM, Van Tassell MS, Madsen MT, Caldwell RT, Berbaum KS. Effect of fatigue on reading computed tomography examination of the multiply injured patient. J Med Imaging [Internet]. 2017 Jul [cited 2020 Aug 10];4(3).
5. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. Skeletal Radiol 2020;49:183-97.
6. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Family Med Prim Care 2019;8:2328-31.
7. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, *et al*. A guide to deep learning in healthcare. Nat Med 2019;25:24-9.
8. Tyson S, Hatem SF. Easily missed fractures of the upper extremity. Radiol Clin North Am 2015;53:717-36, viii.
9. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: Transfer learning from deep convolutional neural networks. Clin Radiol 2018;73:439-45.
10. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer, Sicular S, *et al*. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A. 2018;115(45):11591-11596. doi:10.1073/

pnas.1806905115.
11. Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, *et al*. Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop 2017;88:581-6.
12. Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP, *et al*. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 2018;89:468-73.
13. Rayan JC, Reddy N, Kan JH, Zhang W, Annapragada A. Binomial classification of pediatric elbow fractures using a deep learning multiview approach emulating radiologist decision making. Radiol Artif Intell 2019;1:e180015.
14. Yu JS. Easily missed fractures in the lower extremity. Radiol Clin North Am 2015;53:737-55, viii.
15. Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiol 2019;48:239-44.
16. Cheng CT, Ho TY, Lee TY, Chang CC, Chou CC, Chen CC, *et al*. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. Eur Radiol 2019;29:5469-77.
17. Adams M, Chen W, Holcdorf D, McCusker MW, Howe PD, Gaillard F. Computer vs. human: Deep learning versus perceptual training for the detection of neck of femur fractures. J Med Imaging Radiat Oncol 2019;63:27-32.
18. Balaji GN, Subashini TS, Madhavi P, Bhavani CH, Manikandarajan A. Computer-Aided Detection and Diagnosis of Diaphyseal Femur Fracture. In: Satapathy SC, Bhateja V, Mohanty JR, Udgata SK, editors. Smart Intelligent Computing and Applications. Singapore: Springer; 2020. p. 549-59.
19. Ahrberg AB, Leimcke B, Tiemann AH, Josten C, Fakler JK. Missed foot fractures in polytrauma patients: A retrospective cohort study. Patient Saf Surg 2014;8:10.
20. Kitamura G, Chung CY, Moore BE 2nd. Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, *de novo* training, and multiview incorporation. J Digit Imaging 2019;32:672-7.
21. Pranata YD, Wang KC, Wang JC, Idram I, Lai JY, Liu JW, *et al*. Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. Comput Methods Programs Biomed 2019;171:27-37.
22. Rahmaniar W, Wang WJ. Real-time automated segmentation and classification of calcaneal fractures in CT images. Appl Sci 2019;9:3011.
23. Aso-Escario J, Sebastián C, Aso-Vizán A, Martínez-Quiñones JV, Consolini F, Arregui R. Delay in diagnosis of thoracolumbar fractures. Orthop Rev (Pavia) 2019;11:7774.
24. Burns JE, Yao J, Summers RM. Vertebral body compression fractures and bone density: Automated detection and classification on CT images. Radiology 2017;284:788-97.
25. Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. Comput Biol Med. 2018;98:8-15.
26. Muehlematter UJ, Mannil M, Becker AS, Vokinger KN, Finkenstaedt T, Osterhoff G, *et al*. Vertebral body insufficiency fractures: Detection of vertebrae at risk on standard CT images using texture analysis and machine learning. Eur Radiol 2019;29:2207-17.
27. Hype Cycle Research Methodology. Gartner. Available from: https://www.gartner.com/en/research/methodologies/gartner. [Last accessed on 2020 Apr 25].