**ORIGINAL PAPER**

# Development of a diagnostic support system for distal humerus fracture using artificial intelligence

Aashay Kekatpure[1] · Aditya Kekatpure[1] · Sanjay Deshpande[2] · Sandeep Srivastava[2]

## Abstract

**Purpose** AI has shown promise in automating and improving various tasks, including medical image analysis. Distal humerus fractures are a critical clinical concern that requires early diagnosis and treatment to avoid complications. The standard diagnostic method involves X-ray imaging, but subtle fractures can be missed, leading to delayed or incorrect diagnoses. Deep learning, a subset of artificial intelligence, has demonstrated the ability to automate medical image analysis tasks, potentially improving fracture identification accuracy and reducing the need for additional and cost-intensive imaging modalities (Schwarz et al. 2023). This study aims to develop a deep learning–based diagnostic support system for distal humerus fractures using conventional X-ray images. The primary objective of this study is to determine whether deep learning can provide reliable image-based fracture detection recommendations for distal humerus fractures.

**Methods** Between March 2017 and March 2022, our tertiary hospital's PACS data were evaluated for conventional radiography images of the anteroposterior (AP) and lateral elbow for suspected traumatic distal humerus fractures. The data set consisted of 4931 images of patients seven years and older, after excluding paediatric images below seven years due to the absence of ossification centres. Two senior orthopaedic surgeons with 12 + years of experience reviewed and labelled the images as fractured or normal. The data set was split into training sets (79.88%) and validation tests (20.1%). Image pre-processing was performed by cropping the images to $224 \times 224$ pixels around the capitellum, and the deep learning algorithm architecture used was ResNet18.

**Results** The deep learning model demonstrated an accuracy of 69.14% in the validation test set, with a specificity of 95.89% and a positive predictive value (PPV) of 99.47%. However, the sensitivity was 61.49%, indicating that the model had a relatively high false negative rate. ROC analysis showed an AUC of 0.787 when deep learning AI was the reference and an AUC of 0.580 when the most senior orthopaedic surgeon was the reference. The performance of the model was compared with that of other orthopaedic surgeons of varying experience levels, showing varying levels of diagnostic precision.

**Conclusion** The developed deep learning–based diagnostic support system shows potential for accurately diagnosing distal humerus fractures using AP and lateral elbow radiographs. The model's specificity and PPV indicate its ability to mark out occult lesions and has a high false positive rate. Further research and validation are necessary to improve the sensitivity and diagnostic accuracy of the model for practical clinical implementation.

**Keywords** Artificial intelligence · Diagnostic support system · Distal humerus fractures · Orthopaedic imaging · Deep learning

## Introduction

The professor emeritus of computer science at Stanford University and the originator of the term "artificial intelligence" is John McCarthy [2]. Artificial intelligence (AI) was described for the first time in 1956; however, due to higher data amounts, smarter algorithms, and advancements in computer power and storage, AI is becoming increasingly prevalent today [3]. In the 1960s, the US Department of

✉ Aashay Kekatpure
dr.aashayk@gmail.com

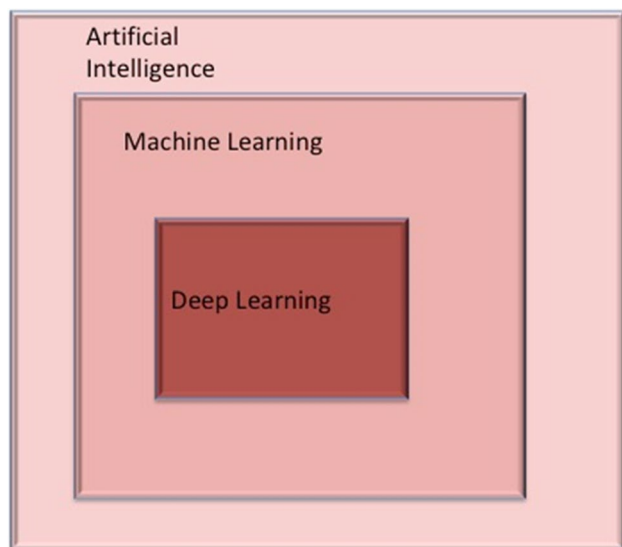1 Department of Orthopaedics, DMMC, Nagpur, India

2 Department of Orthopaedics, JNMC, Wardha, India

Defence grew interested in this type of work and began teaching computers to mimic core human reasoning. In the 1970s, for instance, the Defence Advanced Research Studies Agency (DARPA) [4] conducted street mapping projects. DARPA created intelligent personal assistants in 2003, long before Siri, Alexa, and Cortana became popular names. This ground-breaking study paved the way for today's automation and formal reasoning in computers, such as decision support systems and intelligent search engines, which may be designed to complement and enhance human abilities [5]. To aid readers unfamiliar with AI and machine learning (ML), these fields encompass the development of algorithms and statistical models that enable computers to perform tasks without explicit instructions. AI and machine learning are increasingly integral in health care, offering novel approaches to diagnosis and treatment (Fig. 1). Machine learning in orthopaedics can be likened to a highly skilled assistant who learns from each case, becoming increasingly proficient in predicting outcomes and diagnosing conditions. Just as a surgeon accumulates experience over years of practice and can make more accurate decisions, machine learning algorithms improve their diagnostic capabilities with each new data point. This continuous learning process enables AI systems to assist in accurately diagnosing complex cases, much like a seasoned surgeon's ability to identify subtle nuances in patient conditions. This analogy encapsulates the essence of AI and ML's role in enhancing the precision and efficacy of orthopaedic care.

Distal humerus fractures are an important clinical and public health concern around the world. Avoidance of consequences such as neurovascular sequelae and deformities depends on early diagnosis and treatment of



**Fig. 1** Artificial intelligence and its subsets machine learning and deep learning

the distal humerus fracture. Accurate diagnosis of distal humerus fractures is crucial due to possible complications such as joint stiffness, which can severely impact patient outcomes, including increased mortality risk. The most common method to diagnose a fracture is by X-ray imaging, which accounts for about 6% of all imaging referrals from our tertiary public hospital emergency department. For an accurate diagnosis and treatment plan for the distal humerus fracture, anteroposterior (AP) and lateral elbow radiographs are required. Most distal humerus fractures exhibit a distinct fracture line or bone displacement. Occult fractures, on the other hand, are missed on radiographs. If the fracture line or displacement is minimal or not visible on conventional radiography, the diagnosis of a distal humerus fracture requires a high level of expertise. Only a third of these patients have additional imaging, such as computed tomography (CT) or magnetic resonance imaging (MRI), to reduce the risk of misdiagnosis. Not only does this increase diagnostic costs and resource use, but without access to these advanced imaging modalities (for example in far flung and remote locations), delayed or missed diagnoses are more likely to result in poor patient outcomes, such as increased morbidity, elbow dysfunction, dependency rate, length of hospitalisation and cost of care, as well as employment loss.

Advances in deep learning in medical image analysis have resulted in automated systems that can perform as well as human professionals in a variety of medical tasks. Deep learning is a vertical of AI to enable CNN models to recognise patterns that can be used to distinguish between groups of photos, such as images with and without a specific disease [6]. The evaluation of the distal humerus fracture using X-ray studies based on AI (deep learning) that are highly sensitive and specific can help with a more precise and earlier diagnosis. AI in all probabilities can also minimise the need for costly CT and MRI tests, improving service efficiency and expanding access to highly accurate distal humerus fracture identification in underserved areas [7]. Given the acknowledged difference in diagnostic confidence between human specialists of various levels of experience, automation could also increase reproducibility. We investigate the use of CNNs in deep learning for fracture identification and report this study where a deep learning system was used for X-ray evaluation [8]. In recent years, deep learning using convolutional neural networks (CNN) has become increasingly popular in medical imaging [9]. Recent research has shown that deep learning has the ability to automate the identification and classification of anomalies in a variety of medical imaging modalities [10]. The multi-view technique [11] is particularly promising among different proposed models [12]. The goal of this study was to create a CNN-based deep learning system that uses AP and lateral elbow radiographs to diagnose distal humerus fractures

using conventional radiography and to evaluate its practicality and diagnostic performance.

The literature search revealed that there is a need to establish a validated data set for the training and development of a distal humerus fracture diagnostic algorithm. There is a need to increase awareness among orthopaedic surgeons about proper data labelling and collection. There is a need to develop an external validation set for the assessment of the deep learning algorithm. Therefore, we need to answer the question of whether the deep learning–based diagnostic support system (AI) can diagnose a distal humerus fracture.

a) Aim

   The aim of this study was to determine if deep learning provides reasonable image-based fracture detection recommendations for distal end humerus fractures.

b) Objectives

   - Develop a deep learning–based diagnostic support system for X-ray distal Humerus using deep learning( AI).
   - Establishing the data set.
   - Validation of the output.
   - Testing of the algorithm.

c) Type or nature of study

   Phase 1 study.

d) Sample size

   We collected 4931 images in total for better modelling training and testing (Table 1).

e) Variables

   The difference in the experience of the reviewers ranged from 1 to 12 years.

## Material with their standardisation

Between March 2017 and March 2022, we evaluated PACS data in our tertiary hospital for conventional radiograph images of AP, lateral or both elbows for suspected traumatic distal humerus fracture. Institutional ethics board certification was taken for conducting the study. We excluded images from the paediatric age group under seven years as the centre of ossification of the capitellum, radial head, internal epicondyle and the trochlea does not appear before seven years of age. Radiographs were collected from the

seven and older age group. The total number of images collected was 4931. The casting material was off while taking the image. As some of the images were outpatient once the exact time since the injury was difficult to mention, the average duration was zero to nine days. There were 20% more male compared to female patients.

After anonymisation, the elbow images were retrieved in digital imaging and communications (DICOM) format. Two senior orthopaedic surgeons with 12 + years of experience who were not part of the review group reviewed and labelled the images. Each patient's image was classified as fractured or normal. Each patient's image was classified as fractured (1) or normal (0). All images were viewed in PACS without magnification. Orthopaedic surgeons were told to the group images based only on the above instructions.

From the original data set from our institute, we used 4931 images. Based on the labels, the data for model development were sequentially split into training (3944 pictures, 79.88%) and a validation test set (987 images, 20.1%). All data sets were well balanced for the evaluation of the model.

## Image pre-processing

Each DICOM file was converted to a PNG file and cropped to $224 \times 224$ pixels around the capitellum.

On AP and lateral elbow radiographs, we designated the centre of the capitellum. If the distal humerus was not visible in the cropped picture, the cropping was manually changed. Each image was saved as an array in.npz format.

The deep learning algorithm architecture used is ResNet18 [13]. ResNet18 is a 72-layer architecture with 18 deep layers used for building the model (Fig. 2). The deep learning algorithm was ResNet18, a Python-based high-level deep learning library, implemented on top of PyTorch Lightning, running on a computer equipped with an Intel Core i5 CPU. The GPU utilised was the GTX 1560 Ti via Nvidia CUDA 11. We built an image classification model that used two identical $224 \times 224$ images as input. Every image was fed into ResNet18 models (Fig. 3). The model produces output as 1 and 0 for the images (Fig. 4). The architecture of this network helps in computing large amounts of convolutional layers to function efficiently.

After training on the training set, the model was subjected to evaluation by the validation test set and it gave an out as 0 being negative (normal) and 1 being positive (fracture).

The validation test data set was reviewed by four certified orthopaedic trauma surgeons practising in our institute with variable years of experience (1 to more than 12 years). Reviewers reviewed the DICOM files; clinical information was hidden.

On a two-point scale of 0 and 1, similar to the model output, the reviewers labelled them 0 and 1, 0 being negative

**Table 1** Data set

| Data distribution: |
| --- |
| Total images = 4931 |
| Train units = 3944 |
| Units of test/validation = 987 |

**Fig. 2** Pre-processing if the image and the workflow
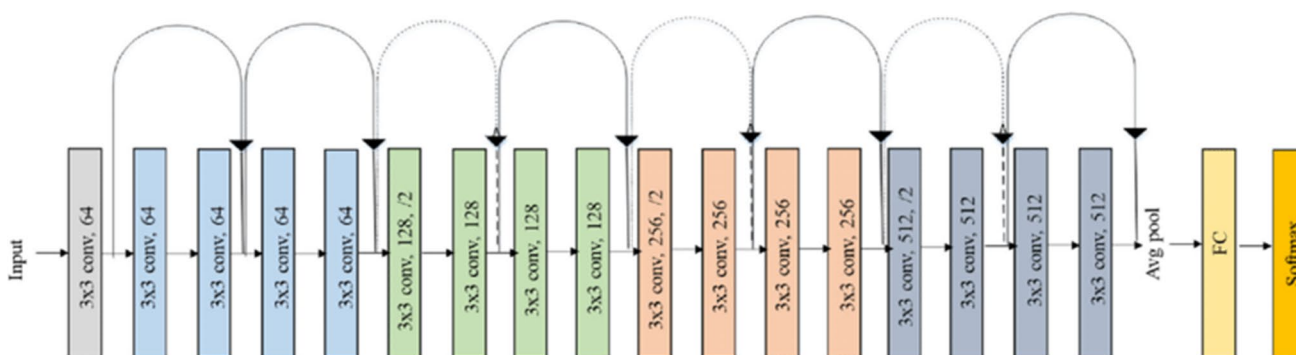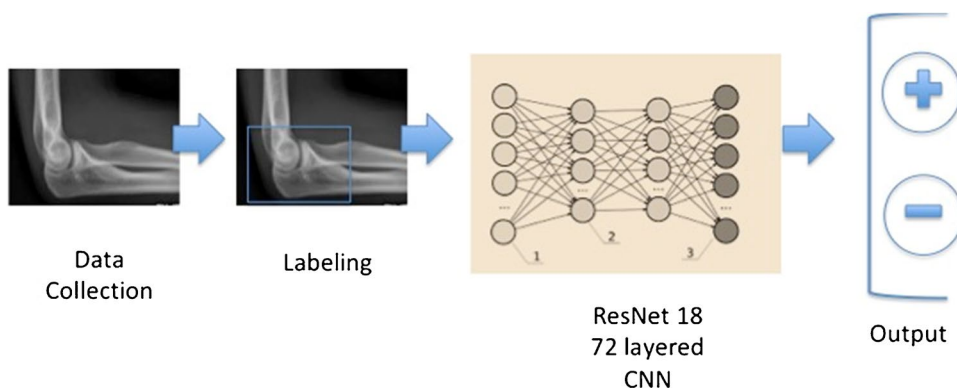


**Fig. 3** The architecture of ResNet18 model. Image source: researchgate.net

(no fracture) and 1 being positive (fracture) (Table 2). The decision of the most senior orthopaedic surgeon was considered the ground truth for the calculation of the ROC performance among orthopaedic surgeons.

## Inclusion and exclusion criteria

- All AP and posterior elbow radiographs above seven years of age.
- X-rays of the elbows with diseased bone, tumour, infection, artefacts and poor quality.

## Analysis of result and discussion

### Data analysis in tabular form

Deep learning and processing information:

1. The input layer and output layers are modified to suit the project (Fig. 5).
2. "Conv1" and "fc" are modified. Conv1 is modified to have one input channel instead of three input channels. Fc is modified to output a single feature.

Hyperparameters: Learning rate = 0.0004, Optimizer = Adam, Loss function = BCEWithLogitsLoss.

3. Deep learning library used = PyTorch Lightning
4. The GPU used is GTX 1560 Ti through Nvidia CUDA 11
5. Epochs trained: 50

Result:
Accuracy = 69.14%
Table 3 shows the confusion matrix of the AI model.

The probability of an accurate diagnosis by orthopaedician 1 or AI (deep learning CNN) was significant whether orthopaedician 1 was the reference (AUC 0.580 (95% CI, 0.544–0.617), $p < 0.001$) (Figure 1a) or AI (deep learning CNN) was the reference (AUC 0.787 (95% CI, 0.747–0.827), $p < 0.001$) (Graph 1b).

Graph 2 a, b, and c show the ROC curves for the diagnosis of DHF by different orthopaedicians having a varying range of experience in reference to AI (deep learning CNN).

There was a significant probability of diagnosing by the first, second, and third orthopaedicians, as indicated by the AUC of 0.713, 0.775, and 0.679 ($p < 0.0001$ for all comparisons), respectively (Table 4).
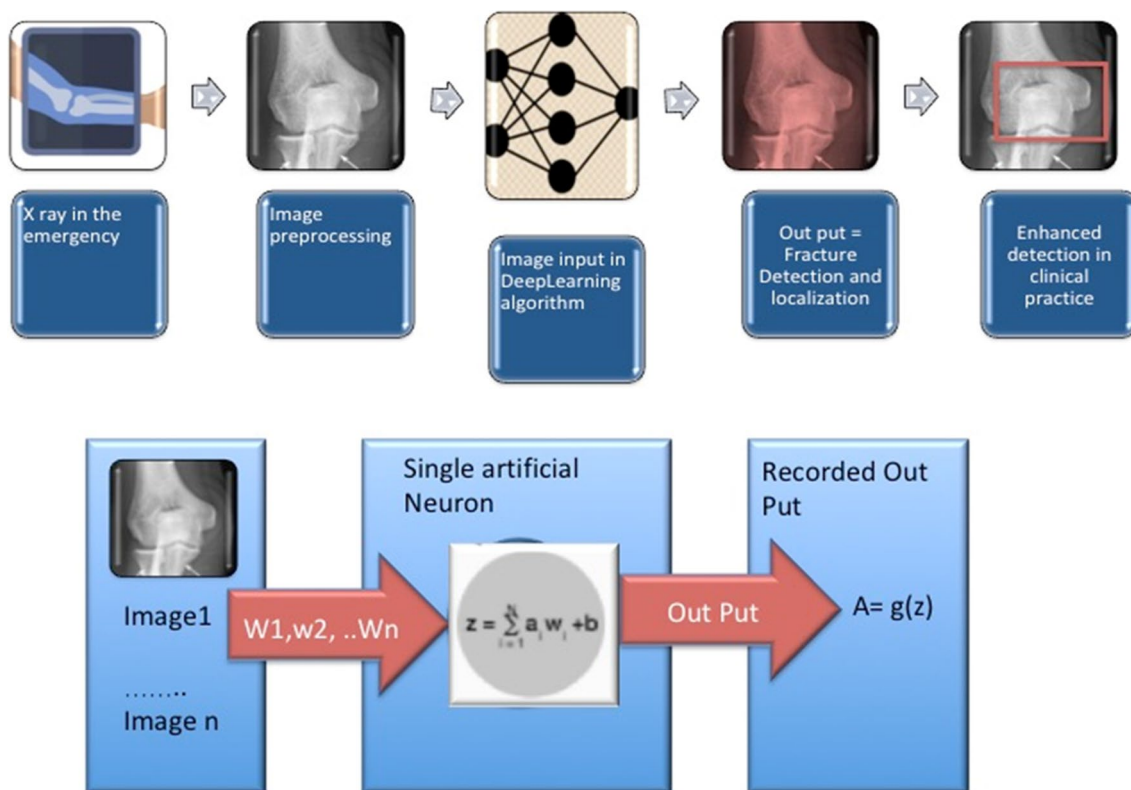
**Fig. 4** Pictorial representation of the functioning of a single AI neuron

**Table 2** Output generated by the AI model

| A. Labelling: |
| --- |
| Negative cases are labelled 0 |
| Positive cases are labelled as 1 |

Similar results were observed compared to the investigator findings as reference (Graph 1) where the first orthopaedician (AUC 0.823, 95% CI 0.790–0.856, $p < 0.0001$) and second (AUC 0.863, 95% CI 0.832–0.894, $p < 0.0001$) orthopaedician had a significantly higher probability of accurate diagnosis at elbow. The third orthopaedician had lower chances of an accurate

**Fig. 5** Workflow of the ResNet model



Diagonal segments are produced by ties.

Diagonal segments are produced by ties.
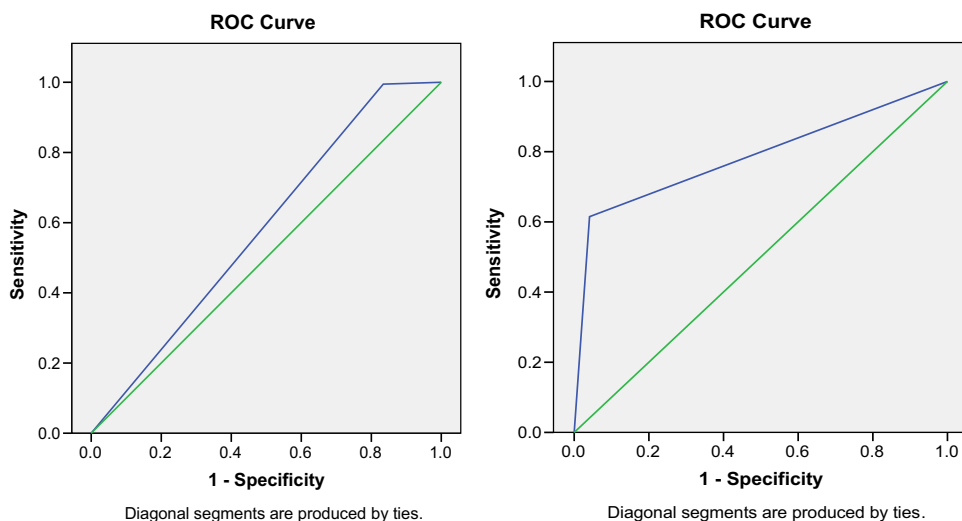
**Table 3** Confusion matrix of the AI model

|  | Predicted yes | Predicted no |
|---|---|---|
| Actual yes | 562 | 3 |
| Actual no | 352 | 70 |

diagnosis in relation to an investigator (AUC 0.545, 95% CI 0.500–0.590, $p = 0.053$).

## Statistical analysis

On the validation test set, a receiver operating curve (ROC) analysis (25) was performed and the area under the receiver operating curve (AUC) was calculated (Graph 3). To investigate the diagnostic performance of the model, we compared the performance of the deep learning model with the three orthopaedic surgeons based on the ROC analysis on the validation set. We constructed $2 \times 2$ confusion matrices [14] for the deep learning algorithm to evaluate the diagnostic performance of the model in the validation data set. The comparison between orthopaedic surgeons was also done using the ROC curve considering the input of the senior orthopaedic surgeons as the ground truth. We calculated the sensitivity, specificity, PPV, NPV, and accuracy of the model using the confusion matrix (Table 5). All data were analysed using SPSS (version 15; SPSS Inc., 233 South Wacker Drive, 11th floor, Chicago, IL 60606–6412).

## Discussion

We built a CNN-based deep neural network model for the identification of distal humerus fractures based on AP and lateral elbow radiographs and compared its detection accuracy with that of orthopaedic surgeons. The probability of an accurate diagnosis by the investigator or AI (deep learning CNN) was significant regardless of whether the investigator was the reference (AUC 0.580 (95% CI 0.544–0.617), $p$ 0.001) or whether the AI (deep learning CNN) was the reference (AUC 0.787 (95% CI 0.747–0.827), $p$ 0.001). ROC curves show the diagnosis of distal humerus fracture by other orthopaedic surgeons with different levels of expertise relative to AI (deep learning CNN). The likelihood of identifying a distal humerus fracture by the second, third and fourth orthopaedists was substantial, as evidenced by AUC values of 0.713, 0.775, and 0.679, respectively ($p$ 0.0001 for all comparisons). Compared to orthopaedician 1 findings as reference (Graph 1), both the first orthopaedician (AUC 0.823, 95% CI 0.790–0.856, $p$ 0.0001) and second orthopaedician (AUC 0.863, 95% CI 0.832–0.894, $p$ 0.0001) had a considerably higher probability of making an appropriate diagnosis in the elbow. Orthopaedician 4 had a lower likelihood

of making a correct diagnosis than orthopaedician 1 (AUC 0.545, 95% CI 0.500–0.590, $p = 0.053$). In the validity section, our model had high AUC values (0.787). The algorithm developed showed high sensitivity and low NPV, for it even exhibited 95% specificity to the test set. The high specificity of our models is an indication that it can mark out occult lesions and have a high false positive rate which we originally expected. Despite the fact that the built deep learning model demonstrated excellent specificity, PPV, and NPV compared to orthopaedic surgeons, less precision and sensitivity showed that there are still improvements to be made in the developed framework.

Rayner et al. [15] demonstrated that artificial intelligence might improve the precision of diagnosis for proximal femoral fractures. The performance of this model was maintained with an AUC of 0980 (0931–1000). However, preclinical analysis revealed hurdles to safe deployment, such as a significant change in the operating point of the model during test validation and an increased error rate in cases with atypical bones. During additional tests, it revealed some limitations. The AUC of our model (AUC 0.787 (95% CI 0.747–0.827), $p$ 0.001) when AI was the reference was 0.787 (0.747–0.827). On the test set, our model generated a satisfactory result.

Ren et al. [16] developed and evaluated a two-stage deep convolutional neural network method to identify triquetral avulsion and Segond fractures. The cross-validated area under the receiver operating characteristic curve values for the two-stage system were 0.95 and 0.990, while the one-stage classifier obtained 0.86 ($p = 0.0086$), whereas the studied model is a CNN with a single stage. The AUC (AI deep learning CNN) produced by the model in the initial test set served as the standard (AUC 0.787 (95% CI 0.747–0.827), $p$ 0.001).

Choi et al. [17] sought to develop and evaluate a dual input convolutional neural network for paediatric supracondylar fracture on conventional radiography. The model has a sensitivity of 93.9%, specificity of 92.2%, positive predictive value of 80.5% and a negative predictive value of 97.8%. Our ResNet18 model has a sensitivity of 61.48796499, a specificity of 95.89041096, a positive predictive value (PPV) of 99.46902655 and a negative predictive value (NPV) of 16.58767773, as well as an accuracy of 69.14660832. A single input model was used.

In the examination of musculoskeletal imaging in general and distal humerus fracture in particular, a comparison with the contralateral side is necessary, since the alignment of the bones and the sign of the fat pad can appear differently depending on the radiographic position [18]. As the suggested model lacked a contralateral comparison method, it is believed to produce more false-positive results than the review group. That is a limitation which other studies have also encountered. Usually, in cases of

**Graph 1** ROC curve for the diagnosis of DHF. **a** AI (deep learning model) compared to the reference performance of orthopaedician 1, denoted by the green line. AUC 0.580 (95% CI, 0.544–0.617), $p < 0.001$. **b** Orthopaedician 1 compared to AI, represented by the green line. AUC 0.787 (95% CI, 0.747–0.827), $p < 0.001$
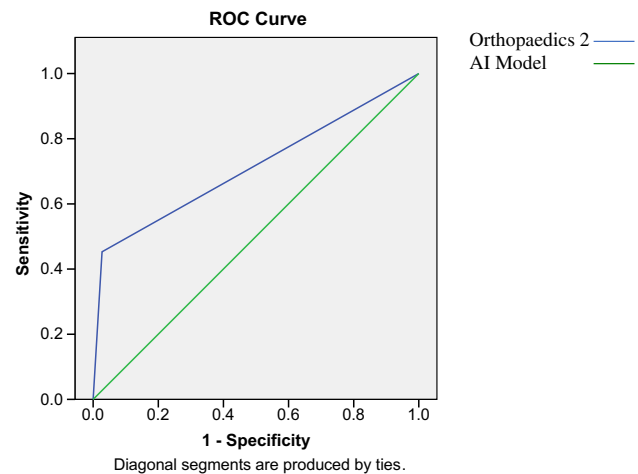
a Orthopedician 2 (>12 years of experience)



b Orthopaedician 3 (>10 years of experience)



c Orthopedician 4 (>1 year of experience post residency .)



clinical dilemma, further investigations are made to arrive at a diagnosis [19, 20].

The decision-making of radiological tests would be especially helpful in a high-volume emergency department with clinical demand for rapid reading, but typically limited human resources. The results of our study show that deep learning can play a role in the evaluation of elbow injuries in the emergency room. It may also decrease the usage of need of CT scan when used as diagnostic support system along with the clinical judgement of an orthopaedic surgeon. Using elbow radiography, the developed AI model may enable sensitive automated identification of a distal humerus fracture, the most prevalent kind of elbow fracture. In a testing scenario, a single pair of radiographs was analysed in roughly 1.6 s after 30 to 40 s of prototype loading. When the paradigm is used in clinical practice, this timescale is acceptable.

An objective test, more often based on external validation in other areas of medicine, is a critical tool to prevent the unreasonably high diagnostic performance of a predictive model caused by skewed data sets [21]. This is, as far as we know, the first study in distal humerus fracture orthopaedic radiology to examine the diagnostic performance of a deep learning model using a test set. The reasonably good AUC values in the test set show that our CNN-based deep learning approach has application promise. Combining AP and lateral elbow radiographs as input, our model applied an approach to diagnostic imaging similar to that of a person. Furthermore, we used a training strategy similar to the way orthopaedic surgeons learn, rather than a multi-step growth procedure that begins with training on single view images. However, contralateral images were not examined in our model, despite the fact that their interpretation usually provides orthopaedists with vital clinical information. The diagnostic performance of the model may be improved in the future by comparing it with healthy contralateral sides.

Our investigation has certain limitations. First, nondistal humerus fractures and dislocations were not considered new types of elbow injuries. Our primary objective was to conduct a feasibility study on the use of deep learning in distal humerus fractures. We believed that instances of abnormalities other than distal humerus fractures would be too rare to learn from our rather small data set. Second, we labelled the data according to orthopaedic physicians' consensus. Ideal reference standards consisting of clinical outcome follow-up or advanced imaging modalities, such as CT and MRI, are
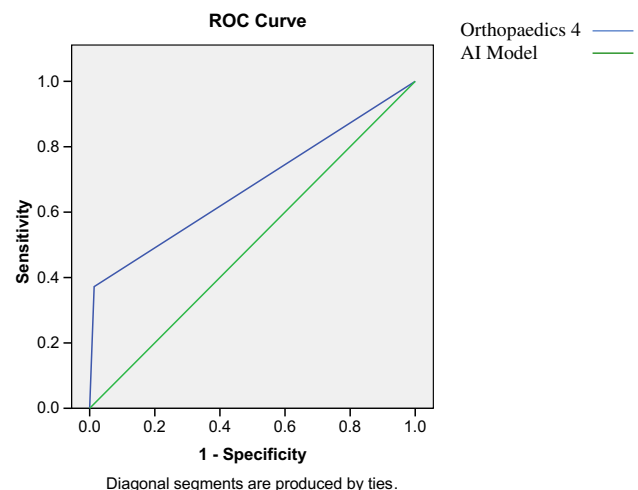
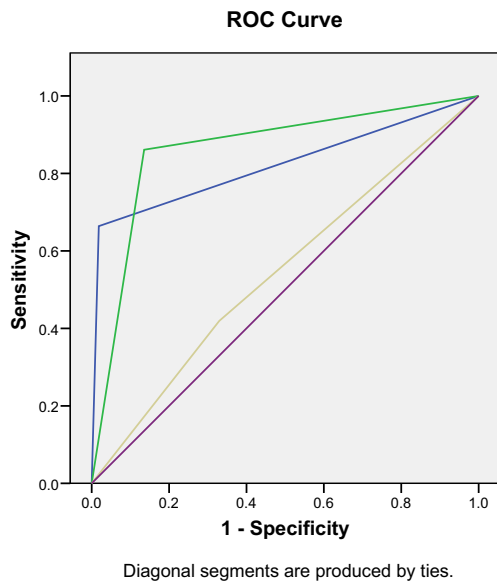**ROC Curve**



Diagonal segments are produced by ties.
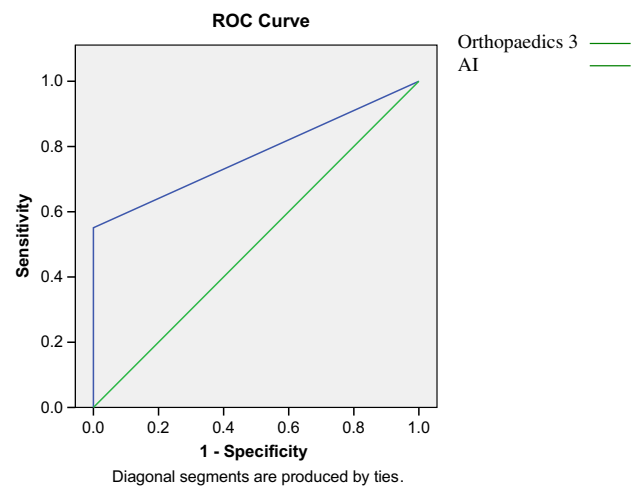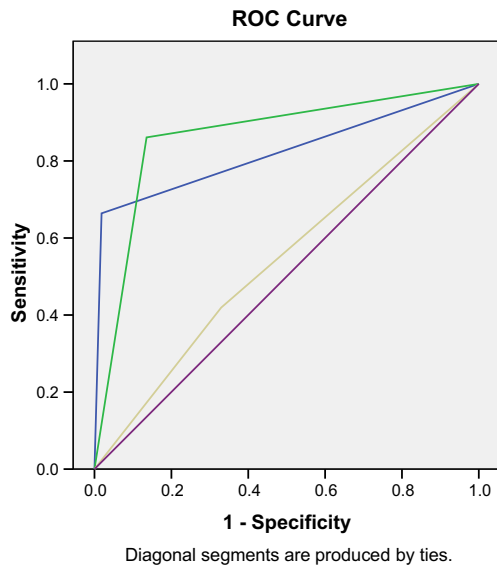
**Graph 2** ROC curve for the diagnosis of DHF by orthopaedicians with variable years of experience compared to the AI model

**Table 4** Area under the curve

| Orthopaedician | AUC | 95% confidence interval | *P*-value |
| --- | --- | --- | --- |
| 1 | 0.823 | 0.790–0.856 | < 0.0001 |
| 2 | 0.863 | 0.832–0.894 | < 0.0001 |
| 3 | 0.545 | 0.500–0.590 | 0.053 |

**ROC Curve**



Diagonal segments are produced by ties.

**Graph 3** Graph depicting the ROC curves of the Orthopedics Surgeons

**Table 5** Statistical parameters of the AI (deep learnings) model based on the confusion matrix

| 1 | **Sensitivity** | 61.48796499 |
| --- | --- | --- |
| 2 | **Specificity** | 95.89041096 |
| 3 | **PPV** | 99.46902655 |
| 4 | **NPV** | 16.58767773 |
| 5 | **Accuracy** | 69.14660832 |

difficult to obtain or not conducted at all. In actual clinical practice, an orthopaedic surgeon's initial assessment determines whether a patient is discharged from the emergency room or needs more evaluation or treatment. In essence, this was a diagnostic assessment research with minimal data. Further prospective multi-institutional research is required to verify the performance and therapeutic utility of artificial intelligence employing deep learning.

We created an artificial intelligence model capable of evaluating both AP and lateral distal humerus radiographs and providing a diagnosis of distal humerus fractures comparable to that performed by orthopaedic surgeons in various phases of training. Our work demonstrates the feasibility and clinical validity of an AI algorithm in the investigation of distal humerus fractures.

## Conclusion

To conclude, we developed a deep-learning model that is capable of analysing both AP and lateral elbow radiographs and delivering a diagnosis of distal humerus fractures reasonably comparable to those of orthopaedic surgeons in various stages of training. Our research has showed the specificity and positive predictive value of a deep learning system for distal humerus fracture examinations.

## Declarations

# References

1. Schwarz GM, Simon S, Mitterer JA et al (2023) Can an artificial intelligence powered software reliably assess pelvic radiographs? Int Orthop 47:945–953. https://doi.org/10.1007/s00264-023-05722-z

2. Beyaz S (2020) A brief history of artificial intelligence and robotic surgery in orthopedics & traumatology and future expectations. Jt Dis Relat Surg 31:653–655. https://doi.org/10.5606/ehc.2020.75300

3. Cesario A, D'Oria M, Calvani R, Picca A, Pietragalla A, Lorusso D, Daniele G, Lohmeyer FM, Boldrini L, Valentini V, Bernabei R (2021) The role of artificial intelligence in managing multimorbidity and cancer. J Personal Med 11(4):314. https://doi.org/10.3390/jpm11040314

4. You J (2015) Artificial intelligence. DARPA sets out to automate research. Science 347(6221):465. https://doi.org/10.1126/science.347.6221.465

5. ( 2016)Step aside, Olympics: here's the Cybathlon. Nature 536(7614):5. https://doi.org/10.1038/536005b

6. Olczak J, Fahlberg N, Maki A et al (2017) Artificial intelligence for analyzing orthopedic trauma radiographs: deep learning algorithms—are they on par with humans for diagnosing fractures? Acta Orthop 88:581–586. https://doi.org/10.1080/17453674.2017.1344459

7. Almog YA, Rai A, Zhang P et al (2020) Deep learning with electronic health records for short-term fracture risk identification: crystal bone algorithm development and validation. J Med Internet Res 22:e22550. https://doi.org/10.2196/22550

8. Kandel I, Castelli M (2021) Improving convolutional neural networks performance for image classification using test time augmentation: a case study using MURA dataset. Health Inf Sci Syst 9:33. https://doi.org/10.1007/s13755-021-00163-7

9. Monchka BA, Kimelman D, Lix LM, Leslie WD (2021) Feasibility of a generalized convolutional neural network for automated identification of vertebral compression fractures: the Manitoba Bone Mineral Density Registry. Bone 150:116017. https://doi.org/10.1016/j.bone.2021.116017

10. Aghnia Farda N, Lai J-Y, Wang J-C et al (2021) Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques. Injury 52:616–624. https://doi.org/10.1016/j.injury.2020.09.010

11. Yang S, Yin B, Cao W et al (2020) Diagnostic accuracy of deep learning in orthopaedic fractures: a systematic review and meta-analysis. Clin Radiol 75:713.e17-713.e28. https://doi.org/10.1016/j.crad.2020.05.021

12. Alsinan AZ, Patel VM, Hacihaliloglu I (2019) Automatic segmentation of bone surfaces from ultrasound using a filter-layer-guided CNN. Int J Comput Assist Radiol Surg 14:775–783. https://doi.org/10.1007/s11548-019-01934-0

13. Zhou Q, Zhu W, Li F et al (2022) Transfer learning of the ResNet-18 and DenseNet-121 model used to diagnose intracranial hemorrhage in CT scanning. Curr Pharm Des 28(4):287–295. https://doi.org/10.2174/1381612827666211213143357

14 Tötsch N, Hoffmann D (2021) Classifier uncertainty: evidence, potential impact, and probabilistic treatment. PeerJ Comput Sci 7:e398. https://doi.org/10.7717/peerj-cs.398

15. Oakden-Rayner L, Gale W, Bonham TA et al (2022) Validation and algorithmic audit of a deep learning system for the detection of proximal femoral fractures in patients in the emergency department: a diagnostic accuracy study. Lancet Digit Health 4:e351–e358. https://doi.org/10.1016/S2589-7500(22)00004-8

16. Ren M, Yi PH (2022) Deep learning detection of subtle fractures using staged algorithms to mimic radiologist search pattern. Skeletal Radiol 51(2):345–353. https://doi.org/10.1007/s00256-021-03739-2

17. Choi JW, Cho YJ, Lee S et al (2020) Using a dual-input convolutional neural network for automated detection of pediatric supracondylar fracture on conventional radiography. Invest Radiol 55:101–110. https://doi.org/10.1097/RLI.0000000000000615

18 Adaş M, Bayraktar MK, Tonbul M et al (2014) The role of simple elbow dislocations in cubitus valgus development in children. Int Orthop 38(4):797–802. https://doi.org/10.1007/s00264-013-2199-4

19 Magee LC, Baghdadi S, Gohel S, Sankar WN (2021) Complex fracture-dislocations of the elbow in the pediatric population. J Pediatr Orthop 41(6):e470–e474. https://doi.org/10.1097/BPO.0000000000001817

20. Nich C, Behr J, Crenn V et al (2022) Applications of artificial intelligence and machine learning for the hip and knee surgeon: current state and implications for the future. Int Orthop (SICOT) 46:937–944. https://doi.org/10.1007/s00264-022-05346-9

21 Ramspek CL, Jager KJ, Dekker FW (2020) External validation of prognostic models: what, why, how, when and where? Clin Kidney J 14(1):49–58. https://doi.org/10.1093/ckj/sfaa188