



Detecting Distal Radial Fractures from Wrist Radiographs Using a Deep Convolutional Neural Network with an Accuracy Comparable to Hand Orthopedic Surgeons

Takeshi Suzuki¹ · Satoshi Maki^{2,3} · Takahiro Yamazaki² · Hiromasa Wakita² · Yasunari Toguchi² · Manato Horii² · Tomonori Yamauchi⁴ · Kouji Kawamura⁴ · Masaaki Aramomi⁴ · Hiroshi Sugiyama⁴ · Yusuke Matsuura² · Takeshi Yamashita⁵ · Sumihisa Orita^{2,3} · Seiji Ohtori²

Received: 6 November 2020 / Revised: 27 August 2021 / Accepted: 16 September 2021 / Published online: 15 December 2021
© Society for Imaging Informatics in Medicine 2021

Abstract

In recent years, fracture image diagnosis using a convolutional neural network (CNN) has been reported. The purpose of the present study was to evaluate the ability of CNN to diagnose distal radius fractures (DRFs) using frontal and lateral wrist radiographs. We included 503 cases of DRF diagnosed by plain radiographs and 289 cases without fracture. We implemented the CNN model using Keras and Tensorflow. Frontal and lateral views of wrist radiographs were manually cropped and trained separately. Fine-tuning was performed using EfficientNets. The diagnostic ability of CNN was evaluated using 150 images with and without fractures from anteroposterior and lateral radiographs. The CNN model diagnosed DRF based on three views: frontal view, lateral view, and both frontal and lateral view. We determined the sensitivity, specificity, and accuracy of the CNN model, plotted a receiver operating characteristic (ROC) curve, and calculated the area under the ROC curve (AUC). We further compared performances between the CNN and three hand orthopedic surgeons. EfficientNet-B2 in the frontal view and EfficientNet-B4 in the lateral view showed highest accuracy on the validation dataset, and these models were used for combined views. The accuracy, sensitivity, and specificity of the CNN based on both anteroposterior and lateral radiographs were 99.3, 98.7, and 100, respectively. The accuracy of the CNN was equal to or better than that of three orthopedic surgeons. The AUC of the CNN on the combined views was 0.993. The CNN model exhibited high accuracy in the diagnosis of distal radius fracture with a plain radiograph.

Keywords Distal radial fractures · Convolutional neural network · Deep learning · Radiograph

Introduction

Distal radial fractures (DRF) comprise approximately 20% of all fractures in the adult population [1]. Plain radiographs remain the standard diagnostic approach for detecting DRFs [2, 3]. DRFs are common fractures and non-orthopedic surgeons could be the primary physician to assess fractures in an outpatient clinic or emergency room, where the fracture may be overlooked. Missed fractures can lead to delay in treatment, malunion, and osteoarthritis. Therefore, a more accurate and efficient fracture detection method is of interest.

In recent years, a deep learning technique called convolutional neural networks (CNNs) has received much attention across various areas including diagnostic imaging in medicine. There are increasing numbers of studies that utilize CNNs in medical image analysis in certain fields, including dermatology for skin lesion identification [4],

✉ Satoshi Maki
satoshimaki@gmail.com

¹ Department of Orthopedic Surgery, Tonosho Hospital, Chiba, Japan

² Department of Orthopedic Surgery, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuou-ku, Chiba 260-8670, Japan

³ Center for Frontier Medical Engineering, Chiba University, Chiba, Japan

⁴ Department of Orthopedic Surgery, Asahi General Hospital, Chiba, Japan

⁵ Department of Orthopaedic Surgery, Oyumino Central Hospital, Chiba, Japan

ophthalmology for the detection of diabetic retinopathy [5], and in radiology for interpreting chest X-ray images for tuberculosis [6]. However, there remain few studies using CNNs in the field of trauma and orthopedics. To date, there are studies using CNNs for radiographic diagnosis of hip fractures [7–10], distal radius fractures [11–14], proximal humeral fractures [15], ankle fractures [16] and hand, wrist, and ankle fractures [17].

The purpose of the present study was to evaluate the performance of CNNs in detecting DRFs on plain radiographs. We investigated the usefulness of multiple views for detecting DRFs from anteroposterior (AP) and lateral plain wrist radiographs using deep CNNs. We also compared the diagnostic capability of CNNs with that of orthopedic surgeons.

Materials and Methods

Patients

This study was performed in line with the principles of the Declaration of Helsinki. This study was approved by the institutional review committees of the four institutions involved. The requirement for consent was waived because the study was a retrospective analysis (IRB No. 3329). Our retrospective study included all consecutive patients with DRF who were surgically or conservatively treated. The patient's data was taken from those attending Tonosho Hospital between November 2012 and May 2019, Asahi General Hospital between January 2013 and June 2019, Chiba

University Hospital between April 2014 and January 2020, and the Oyumino Central Hospital between March 2014 and January 2019. Fractures were diagnosed mainly using radiographs. If the fracture was unclear, we also reviewed the clinical course and computed tomography. Diagnoses were confirmed by two board certified orthopedic surgeons. There were 961 patients with wrist radiographs. We excluded a total of 169 patients: 152 with a plaster cast, 12 with a metal implant in the radius, two with an old fracture, two with an intravenous catheter, and one who had screw holes after implant removal. Ultimately, we included 503 patients with DRF and 289 patients without fracture (Fig. 1). The wrist radiographs without fracture were taken from patients with carpal tunnel syndrome or suspected fractures that were diagnosed as sprains. Except for patients in the validation dataset, patients do not always have both AP and lateral views. No multiple images more than one AP and one lateral view were obtained from a single patient.

Radiograph Dataset/Image Preprocessing for Deep Learning

We identified 905 AP wrist radiographs (452 with DRF and 453 without fracture) and 728 lateral wrist radiographs (420 with DRF and 308 without fracture) (Fig. 1). The radiographic images stored in the Digital Imaging and Communications in Medicine (DICOM) server were extracted from the Picture Archiving and Communication System (PACS) in Joint Photographic Experts Group (JPEG) format. The compression level of the jpeg image was set to 100.

The matrix sizes of the identified the region of interest (ROIs) ranged from 1024 to 2505 pixels in width and from 1024 to 3015 pixels in height, respectively. An orthopedic surgeon (TS, 6 years of experience) cropped the smallest region, including the carpometacarpal joint and distal 1/5 of the forearm, on both AP and lateral wrist radiographs. The images were manually cropped into a square in which the distal radius was centered (Fig. 2). Images were pre-processed using Preview (Apple, Cupertino, CA) to generate images for CNN training. For the validation dataset, we held out each of 75 images with and without fractures from AP and lateral radiographs. On AP radiographs, 377 fracture views and 378 views without fracture were used for training. On lateral radiographs, 345 fracture views and 233 views without fracture were used for training (Fig. 1). Training was performed using AP and lateral views separately.

Model Construction for CNN

We used Python programming language, version 3.6.7 (<https://www.python.org>) to implement the classification model as CNN with Keras, version 2.2.4 and Tensorflow, version 1.14.0 (<https://www.tensorflow.org>) as frameworks.

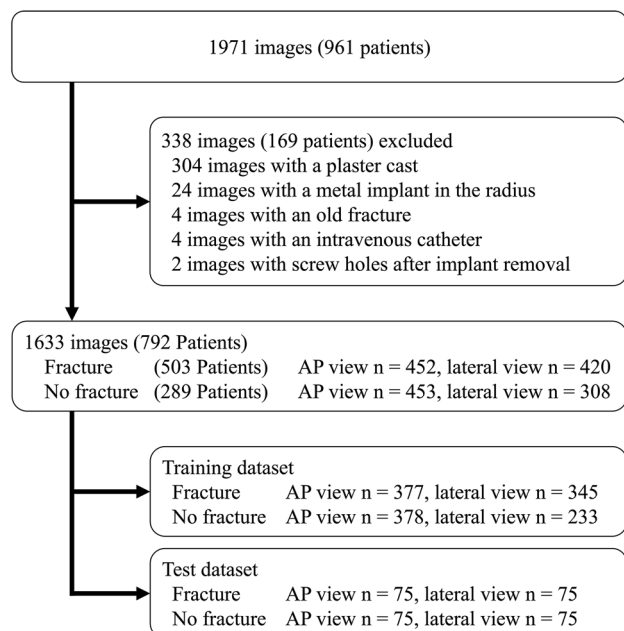


Fig. 1 Data flow diagram from images to model training. AP, anteroposterior



Fig. 2 Image preprocessing for the convolutional neural network model training and validation. We cropped the image to a minimum region that included the carpometacarpal joint and distal 1/5 forearm on both the anteroposterior (a) and lateral (b) wrist radiographs

We used four different models from the family of EfficientNets [18], which had been already trained using images with ImageNet. EfficientNets is a family of image classification models developed based on AutoML and compound scaling. A simple, yet highly effective composite scaling method is proposed to scale up mobile-sized baseline networks to improve performance while maintaining efficiency. It has fewer model parameters and is more accurate and efficient than existing convolutional networks. Pretrained on ImageNet, an EfficientNet-B2, B3, B4, and B5 convolutional neural network with a single fully connected 2-class classification layer was used. The input images were scaled to 260×260 pixels for EfficientNet-B2, 300×300 for EfficientNet-B3, 380×380 for EfficientNet-B4, and 456×456 for EfficientNet-B5. Then we applied transfer learning to the model using the dataset of radiographs of DRF and wrists without fracture. The network was trained for at a learning rate of 0.1 in 100 epochs. If there was no improvement, the learning rate decreased. Model training convergence was monitored using cross-entropy loss. Image augmentation was conducted using ImageDataGenerator (<https://keras.io/preprocessing/image/>) by a rotation angle range of 20° , width shift range of 0.2, height shift range of 0.2, brightness range of 0.3–1.0, and a horizontal flip in 50%. Furthermore, we constructed separate models for AP and lateral radiographic views. Training and validation of the CNN were performed using a computer with a GeForce RTX 2060

graphics processing unit (Nvidia, Santa Clara, CA), a Core i7-9750 central processing unit (Intel, Santa Clara, CA), and 16 GB of random access memory.

Performance Evaluation

We evaluated the diagnostic ability of the trained CNN model using a validation dataset. This validation dataset was prepared separately from the training dataset.

The diagnosis of DRF was made for each view: (1) AP wrist radiographs alone, (2) lateral wrist radiographs alone, and (3) both AP and lateral wrist radiographs. We evaluated the CNN performance in these three radiographic views. When diagnosing the fracture based on both AP and lateral views, the best performing model for each view was selected and a definitive diagnosis was made based on the averaged probability of the diagnosis in AP and lateral radiographs. This allowed for a comprehensive diagnosis based on both AP and lateral radiographs rather than a single view, closely mimicking the way in which a clinician makes a diagnosis. We made the final decision based on the optimal cutoff point for the DRF probability score.

Image Assessment by Orthopedic Surgeons

Three hand orthopedic surgeons (7, 8, and 10 years of experience as orthopedic surgeon, respectively) interpreted the AP and lateral wrist radiographs. The radiographs were the same validation dataset as used for evaluation of the CNN. They reviewed jpeg-format images at the same resolution as the original DICOM images, using the same area as used for CNN training. This was to ensure consistent conditions between the CNN and clinicians. However, these conditions differ from those used in actual clinical settings. At the time of interpretation, the surgeons were blinded to the mechanism of injury and patient age.

Statistical Methods

We conducted all statistical analyses using JMP (version 12.0.1; SAS Institute, Cary, NC, USA). Continuous variables were described as means and standard deviations (SD), and categorical variables as frequencies and percentages. A Student *t* test was used to compare continuous variables and a Pearson chi-square test was used to compare categorical variables between the groups. We considered $P < 0.05$ to be significant in two-sided tests of statistical inference. Based on the predictions, we calculated the percentages of true positives, true negatives, false positives, and false negatives. A receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC) were used to evaluate the performance of the CNN. Then we calculated the sensitivity, specificity, and accuracy for the CNN and the

Table 1 The demographic data of the wrist radiograph dataset

	Distal radius fracture	Non-fracture	<i>P</i> -value
<i>n</i> (patients)	503	289	
Age, mean (SD)	63.7 ± 16.4	67.6 ± 18.1	0.0026
Sex (M/F)	128/375	109/180	0.0004

three hand orthopedic surgeons. Sensitivity, specificity, and accuracy were determined from optimal thresholds using the highest Youden index (sensitivity + 1, specificity − 1) in ROC analysis. The sensitivity, specificity, and accuracy of the diagnostic ability between CNN and the hand orthopedic surgeons were compared using a McNemar test.

Results

Demographic Data of the Included Patients

The demographic data of the patients in the wrist radiograph dataset are shown in Table 1. In total, we included 503 patients with DRF and 289 patients without fracture.

We enrolled 256 with DRF and 205 without fracture from Asahi General Hospital, and 189 DRF patients only from Oyumino Central Hospital, 33 patients with DRF and 59 without fracture from Tonosho Hospital, and 25 patients with DRF and 25 without fracture from Chiba University Hospital. The proportion of women in the DRF group was significantly higher, as consistent with a previous study [12]. The mean age of the patients was significantly higher in the DRF group. We identified 905 AP wrist radiographs (452 with DRF and 453 without

fracture) and 728 lateral wrist radiographs (420 with DRF and 308 without fracture).

Performance of the CNN Compared to the Hand Orthopedic Surgeons

The diagnostic performance of the four CNN models on AP and lateral views is shown in Table 2. EfficientNet-B2 in the AP view and EfficientNet-B4 in the lateral view showed the best AUC, respectively. The diagnostic capability of DRF in AP view by EfficientNet-B2, lateral view by EfficientNet-B4, and both views by an ensemble of both models, exhibited excellent diagnostic performance with an AUC of 0.995 (95% CI 0.971–0.999), 0.993 (95% CI 0.955–0.999) and 0.993 (95% CI 0.949–0.999). The ROC curve of the ensemble model probability compared with three hand orthopedic surgeons is shown in Fig. 3. Table 3 shows the sensitivity, specificity, and accuracy of the CNN model and the three hand orthopedic surgeons on the two views at the optimal cutoff point. The CNN model showed a high performance of 99.3% accuracy, 98.7% sensitivity, and 100% specificity on the two views. The accuracy of the CNN was equal to or better than that of the three hand orthopedic surgeons. The sensitivity and specificity of the CNN tended to be superior to that of the three hand orthopedic surgeons, although most comparisons were not significantly different.

For reference, we show representative wrist radiographs from two patients that CNN correctly diagnosed, but orthopedic surgeons misdiagnosed, and radiographs

Table 2 The accuracy, sensitivity, and specificity of the CNN models. Ensemble model consists of two models that produced the best outputs together, which in this case was the EfficientNet B2 for AP

Model	Views	Accuracy	Sensitivity	Specificity	AUC	95% CI
EfficientNet B2	AP	98.7 (95.3–99.6)	98.7 (92.8–99.8)	98.7 (92.8–99.8)	0.995	0.971–0.999
	Lateral	93.3 (88.2–96.3)	92.0 (83.6–96.3)	94.7 (87.1–97.9)	0.976	0.940–0.990
EfficientNet B3	AP	96.7 (92.4–98.6)	93.3 (85.3–97.1)	100 (95.1–100)	0.993	0.980–0.998
	Lateral	93.3 (88.2–96.3)	94.7 (87.1–97.9)	92.0 (83.6–96.3)	0.982	0.958–0.992
EfficientNet B4	AP	96.7 (92.4–98.6)	97.3 (90.8–99.3)	96.0 (88.9–98.6)	0.989	0.961–0.997
	Lateral	98.7 (95.3–99.6)	98.7 (92.8–99.8)	98.7 (92.8–99.8)	0.993	0.955–0.999
EfficientNet B5	AP	98.7 (95.3–99.6)	98.7 (92.8–99.8)	98.7 (92.8–99.8)	0.995	0.968–0.999
	Lateral	96.0 (91.5–98.2)	97.3 (90.8–99.3)	94.7 (87.1–97.9)	0.987	0.954–0.997
Ensemble	AP+Lateral	99.3 (96.3–99.9)	98.7 (92.8–99.8)	100 (95.1–100)	0.993	0.949–0.999

and the EfficientNet B4 for lateral radiographs. Data in parentheses are the 95% confidence interval. CNN, convolutional neural network; AP, anteroposterior

misdiagnosed by both CNN and the hand orthopedic surgeons (Fig. 4).

Discussion

The present study demonstrated the high performance of deep learning CNNs to distinguish DRF from normal wrists. The CNN model exhibited better accuracy than that of orthopedic hand orthopedic surgeons.

For the detection of DRF in plain radiographs, the CNN model showed excellent performance with an accuracy of 99.3% and sensitivity of 98.7%, specificity of 100%, and AUC of 0.993 based on both AP and lateral views. The accuracy of the CNN was better than that of the hand orthopedic surgeons. We used 755 front wrist X-rays and 578 lateral wrist radiographs for training, and 150 front and lateral wrist radiographs for verification. Automated detection of the DRFs on plain radiographs has been reported in several articles. Three studies applied CNN to classify DRF and without fracture. Olczak et al. evaluated openly available deep learning networks, achieving 83% accuracy in fracture detection in hand, wrist, and ankle radiographs [17]. Kim and MacKinnon reported that the CNN using Inception-v3 based on 1389 lateral wrist radiographs exhibited an accuracy of 89% and an AUC of 0.954 [11]. Blüthgen et al. used a closed-source framework of CNN based on 624 combined AP and lateral wrist radiographs to identify DRFs in AP X-ray images achieving a sensitivity of 78%, specificity of 82%, and AUC of 0.89 [14]. These diagnostic parameters were comparable to those of radiologists. By contrast, there are studies that used Faster R-CNN to detect and locate DRFs. Yahalomi et al. described object detection by CNN based on 95 AP wrist radiographs to detect wrist fractures in AP X-ray images achieving an accuracy of 96% and a mean average precision (mAP) of 0.87 [13]. However, they did not investigate the difference between the performance of the CNN and that of experts. Thian et al. used the Faster R-CNN based on 7356 AP and lateral wrist radiographs to identify and locate distal radius fractures achieving a sensitivity of 98.1%, specificity of 72.9%, and AUC of 0.895 [19]. Gan et al. used R-CNN to extract the distal part of radius and used Inception-v4 to distinguish DRF and wrists without fracture based on 2040 AP radiographs with an accuracy of 93% and an AUC of 0.96 [12]. These diagnostic performances were better than those of radiologists and similar to those of orthopedic surgeons. The object detection network is more informative compared

to classification networks as it provides location of the fracture. Although the object detection network requires a larger annotated training dataset to guarantee its performance, using a larger dataset is time consuming, as when detecting a minimally displaced fracture. Comparing the performance of experts with the CNN is desirable as a reference to estimate the difficulties in using the validation dataset. In the present study, we achieved a comparable or better diagnostic capability using the CNN than that achieved by hand orthopedic surgeons, based on a modestly sized dataset.

The present study has also shown the importance of using radiographs taken from two directions when diagnosing fracture using the CNN model. Acquiring two radiographs taken orthogonally makes it easier to assess the relative position of two fractured bones [20]. On occasion, fractures are diagnosed using only a single view, and as a consequence the reviewer may misdiagnose DRF as without fracture in the absence of the additional view. The diagnostic sensitivity and specificity of DRF were improved by ensemble decision making. We used a CNN model trained on radiographs taken from two directions. The use of radiograph images taken from two directions for training also served to reduce the number of images required for diagnostic accuracy equal to or better than that for previously reported CNN models. However, to our knowledge, there are only a few reports of the accuracy of CNN fracture diagnosis using radiograph images obtained from multiple directions [14, 16, 19].

There are several limitations to our study. First, the present study was based on a binary classification: DRF or wrist without fracture. The CNN could not localize the pathological region. For less experienced clinicians, it is difficult to trust broad classification labels of such “black-box” models, although there is Grad-CAM [21], which is a heatmap visualization for a given class label. Second, our dataset of 1633 radiographs is considered small. To solve this problem, we used fine-tuning and data augmentation. As a result, we achieved a high accuracy of 99% despite the small sample size. Third, assessment of the diagnostic performance of the deep learning models was based on only adult wrist radiographs. Pediatric wrist radiographs have growth plates that can mimic the appearance of a fracture. We did not have enough pediatric radiographs without fracture to evaluate performance for pediatric DRF. Fourth, it was disadvantageous for hand orthopedic surgeons to make a judgment based on a small cropped image, especially without information that would be available in clinical settings. The diagnostic accuracy of the surgeons would improve if they were provided clinical information, such as the location of pain.

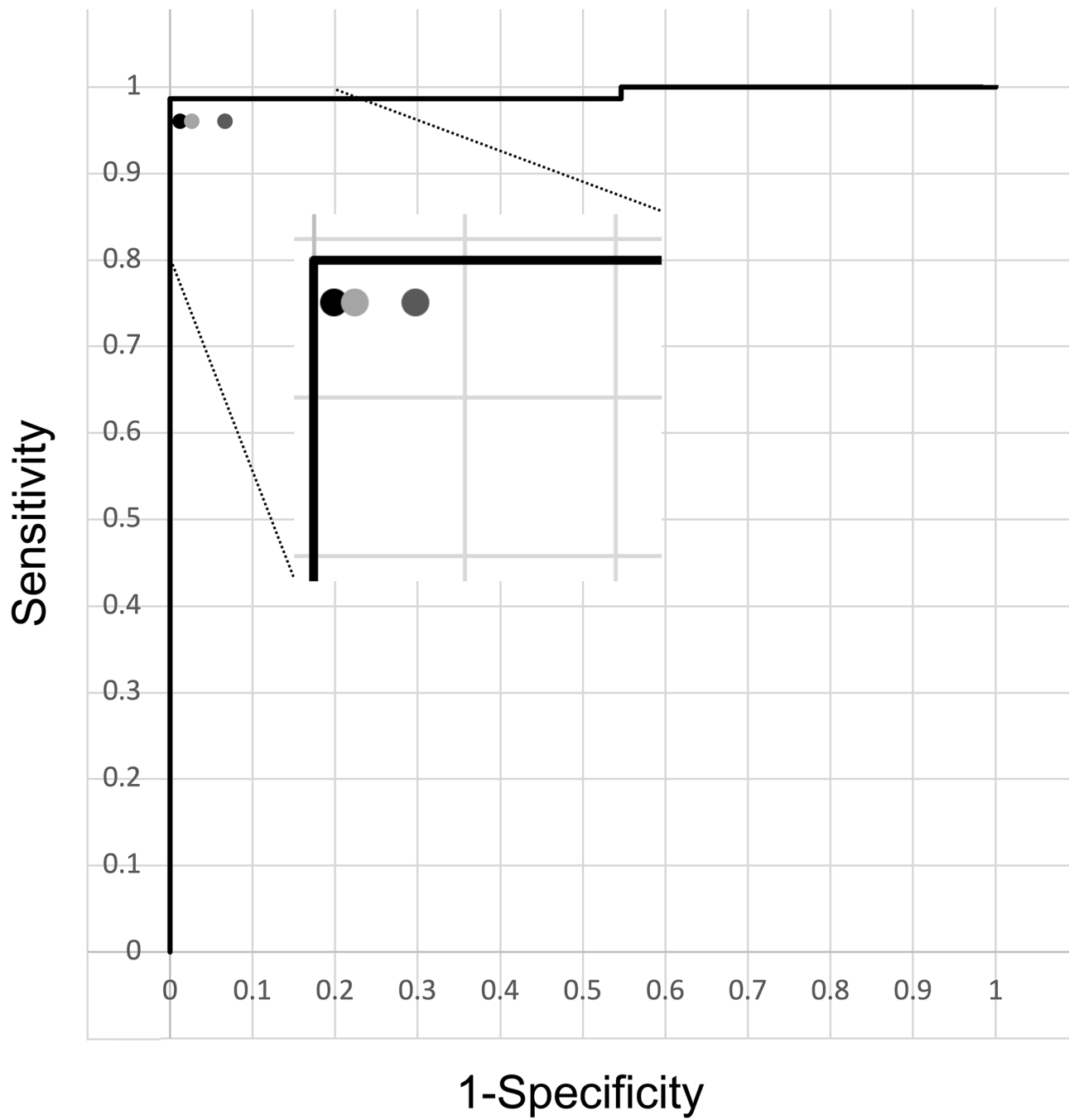


Fig. 3 Receiver operating characteristic curves of the convolutional neural network model based on both anteroposterior and lateral wrist radiographs. Performances of the hand orthopedic surgeons 1, 2, and

3 are also shown as the black circle, dark gray circle, and light gray circle symbols, respectively

Table 3 Comparison of accuracy, sensitivity, and specificity between the ensemble CNN model and the three hand orthopedic surgeons based on frontal and lateral wrist radiographs. Data in parentheses are the 95% confidence interval. CNN, convolutional neural network

	Accuracy	<i>P</i> value (compared with CNN)	Sensitivity	<i>P</i> value (compared with CNN)	Specificity	<i>P</i> value (compared with CNN)
CNN (ensemble)	99.3 (96.3–99.9)	-	98.7 (92.8–99.8)	-	100 (95.1–100)	-
Hand orthopedic surgeon						
1	97.3 (93.3–99.0)	0.083	96.0 (88.9–98.6)	0.083	98.7 (92.8–99.8)	0.317
2	94.7 (89.8–97.2)	0.008	96.0 (88.9–98.6)	0.083	93.3 (85.3–97.1)	0.025
3	96.7 (92.4–98.6)	0.046	96.0 (88.9–98.6)	0.083	97.3 (90.8–99.3)	0.157

Fig. 4 Representative X-ray images of wrist fractures. The AP (a) and lateral (b) radiographs of a distal radius fracture, which 2 of the 3 hand orthopedic surgeons misdiagnosed as without fracture, while the CNN diagnosed the fracture correctly. The AP (c) and lateral (d) radiographs of a distal radius fracture, which 2 of the 3 hand orthopedic surgeons diagnosed correctly and the CNN misdiagnosed as without fracture. The fracture is clear on oblique view (e), although this view is only for the reference and not used for validation. AP, anteroposterior; CNN, convolutional neural network



Conclusion

In the present study, we showed the feasibility of a CNN for accurate detection of DRF on wrist radiographs and showed comparable or better diagnostic capabilities than hand orthopedic surgeons under the same conditions. Compared with using a single view, we found the diagnostic ability of the CNN improved using frontal and lateral views.

References

- Nellans KW, Kowalski E, Chung KC: The epidemiology of distal radius fractures. *Hand Clin* 28:113–125, 2012
- Mauffrey C, Stacey S, York P J, Ziran B H, Archdeacon M T: Radiographic evaluation of acetabular fractures: review and update on methodology. *J Am Acad Orthop Surg* 26(3): 83–93, 2018
- Waeber D, Madsen M L, Rolfing J H D, Borris L C, Henriksen M, Nagel L, Thorninger R: Distal radius fractures are difficult to classify. *Injury* 49 (Suppl .1): S29–S32, 2018
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542: 115–118, 2017
- Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL Webster DR: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316: 2402–10, 2016
- Lakhani P, Sundaram B: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284: 574–82, 2017
- Adams M, Gaillard F, Chen W, Holcdorf D, Mccusker MW, Howe PDL, Gillard F: Computer vs human: deep learning versus perceptual training for the detection of neck of femur fractures. *J Med Imaging Radiat Oncol* 63(1): 27–32, 2018
- Badgeley MA, Zech JR, Oakden-rayner L, Glicksberg BS, Liu M, Gale W, McConnell M V, Percha B, Snyder TM, Dudley JT:

- Deep learning predicts hip fracture using confounding patient and healthcare variables. *NPJ Digit Med* 2(1); 1–10, 2019
9. Cheng C, Ho T, Lee T, Chang C, Chou C, Chen C: Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol* 29(10); 5469–5477, 2019
 10. 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* 48(2); 239–244, 2019
 11. Kim DH, MacKinnon T: Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* 73(5); 439–45, 2018
 12. Gan K, Xu D, Lin Y, Shen Y, Zhang T, Hu K, Zhou K, Bi M, Pan L, Wu W, Liu Y: Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. *Acta Orthop* 90(4); 394–400, 2019
 13. Yahalomi E, Chernofsky M, Werman M: Detection of distal radius fractures trained by a small set of X-ray images and Faster R-CNN. *Adv Intell Syst Comput* 997; 971–81, 2019
 14. Blüthgen C, Becker AS, Martini IV De, Meier A, Martini K, Frauenfelder T: Detection and localization of distal radius fractures: deep learning system versus radiologists. *Eur J Radiol* 126; 108925, 2020
 15. Chung SW, Han SS, Lee JW, Oh K, Kim NR, Yoon JP, Kim JY, Moon SH, Kwon J, Lee H, Noh Y, Kim Y, Chung SW, Han SS, Lee JW, Oh K, Kim NR, Yoon JP: Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop* 89(4); 468–473, 2018
 16. Kitamura G, Chung CY, Moore BE: Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, de novo training, and multiview incorporation. *J Digit Imaging* 32(4); 672–677, 2019
 17. Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, Sköldenberg O, Gordon M: Artificial intelligence for analyzing orthopedic trauma radiographs: deep learning algorithms—are they on par with humans for diagnosing fractures? *Acta Orthop* 88(6); 581–586, 2017
 18. Tan M, Le QV: Efficientnet: rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning* [Internet]. 2019. Available from: arXiv:1905.11946. Accessed 18 June 2020
 19. Thian YL, Li Y, Jagmohan P, Sia D, Chan VEY, Tan RT: Convolutional neural networks for automated fracture detection and localization on wrist radiographs. *Radiol Artif Intell* 1(1); e180001, 2019
 20. van der Plaats GJ: *Medical X-ray technique*, Eindhoven, The Netherlands: Macmillan International Higher Education, 1969
 21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-cam: visual explanations from deep networks via gradient-based localization, in 2016 IEEE International Conference on Computer Vision; 618–626, 2016. arXiv:1610.02391. Accessed 18 June 2020

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.