

Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments

Kaifeng GAN^{1,2}, Dingli XU², Yimu LIN³, Yandong SHEN^{1,2}, Ting ZHANG¹, Keqi HU¹, Ke ZHOU¹, Mingguang BI¹, Lingxiao PAN¹, Wei WU⁴, and Yunpeng LIU⁵

¹ Department of Orthopaedics, Ningbo Medical Center, Lihuili Hospital, Ningbo, 315000, China; ² School of Medicine, Ningbo University, Ningbo, 315000, China; ³ Department of Orthopaedics, Second Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325027, China; ⁴ Department of Orthopaedics, Second Hospital of Ningbo, Ningbo, 315000, China; ⁵ Faculty of Electronics & Computer, Zhejiang Wanli University, Ningbo, 315000, China

Correspondence: liuyujoy@163.com

Submitted 2018-12-11. Accepted 2019-03-09.

Background and purpose — Artificial intelligence has rapidly become a powerful method in image analysis with the use of convolutional neural networks (CNNs). We assessed the ability of a CNN, with a fast object detection algorithm previously identifying the regions of interest, to detect distal radius fractures (DRFs) on anterior–posterior (AP) wrist radiographs.

Patients and methods — 2,340 AP wrist radiographs from 2,340 patients were enrolled in this study. We trained the CNN to analyze wrist radiographs in the dataset. Feasibility of the object detection algorithm was evaluated by intersection of the union (IOU). The diagnostic performance of the network was measured by area under the receiver operating characteristics curve (AUC), accuracy, sensitivity, specificity, and Youden Index; the results were compared with those of medical professional groups.

Results — The object detection model achieved a high average IOU, and none of the IOUs had a value less than 0.5. The AUC of the CNN for this test was 0.96. The network had better performance in distinguishing images with DRFs from normal images compared with a group of radiologists in terms of the accuracy, sensitivity, specificity, and Youden Index. The network presented a similar diagnostic performance to that of the orthopedists in terms of these variables.

Interpretation — The network exhibited a diagnostic ability similar to that of the orthopedists and a performance superior to that of the radiologists in distinguishing AP wrist radiographs with DRFs from normal images under limited conditions. Further studies are required to determine the feasibility of applying our method as an auxiliary in clinical practice under extended conditions.

Conventional radiographs remain the primary diagnostic approach to detect distal radius fractures (DRFs) (Mauffrey et al. 2018, Waever et al. 2018). Non-orthopedic surgeons or young radiologists at emergency departments, where urgent decision-making is often required, are usually the first doctors to assess radiographs. Therefore, an accurate and efficient assistant technology in fracture detection is of interest.

Artificial intelligence (AI) is achieving remarkable progress in image interpretation (He et al. 2015, Russakovsky et al. 2015). Since 2012, deep learning, a branch of AI, has rapidly become a powerful method in image analysis with the use of convolutional neural networks (CNNs), which are well suited for analyzing image features (Russakovsky et al. 2015, Lakhani and Sundaram 2017). There are increasing numbers of experimental trials that apply deep learning in medical image analysis in certain fields, including the automated analysis of pulmonary tuberculosis (Lakhani and Sundaram 2017), lung nodules (Hua et al. 2015, Nishio et al. 2018), retinopathy (Ting et al. 2017), gastric cancer (Wang et al. 2018), and dermatological diseases (Li and Shen 2018, Yap et al. 2018, Fujisawa et al. 2019). In the field of traumatic orthopedics, a few studies (Olczak et al. 2017, Chung et al. 2018, Kim and MacKinnon 2018, Urakawa et al. 2019) investigated the experimental applications of deep learning to detect fractures on plain radiographs; all the CNNs adopted showed excellent performance, and some (Chung et al. 2018, Urakawa et al. 2019) had abilities superior to that of humans. To further validate the feasibility of AI as an automatic diagnostic model, we first evaluated the ability of a CNN, with a fast object detection algorithm previously identifying the regions of interest, to detect DRFs on AP wrist radiographs. Second, the diagnostic performances of CNNs were compared with those of radiologists and orthopedists.

Materials and methods

Design of study

With the dataset, a fast object detection algorithm based on deep learning was first trained to identify the distal radiuses on AP wrist radiographs as the regions of interest (ROIs). Second, we adopted this fast object detection algorithm, of which the feasibility had been verified by a validation process, to automatically annotate the ROIs on AP wrist radiographs in the training dataset and test dataset. The ROIs were extracted as images, with which a diagnostic CNN model was then trained and tested in detecting the DRFs. The diagnostic performances in terms of accuracies, sensitivities, specificities, and Youden Index were finally compared among the diagnostic CNN model, radiologists, and orthopedists.

Dataset

2 senior orthopedists with more than 10 years of orthopedic professional experience retrospectively reviewed 2,359 plain wrist radiographs with diagnostic reports from 2,359 adult patients (the inclusion and exclusion criteria for this study are given in the Supplementary data) who underwent radiological examinations at the Medical Center of Ningbo City, Lihuli Hospital, of the Ningbo University School of Medicine, between January 2010 and September 2017 to confirm that each case had an accurate diagnosis (with DRFs or without DRFs). A consensus was achieved in consultation with a third senior orthopedist with 22 years of orthopedic professional experience. For cases in which all 3 orthopedists did not agree, the corresponding wrist CT images were reviewed; CTs were available in most of these cases and a consensus on each case was reached after discussion. 19 controversial cases without CT exams were excluded from the study. 2,340 AP wrist radiographs (1,491 DRF cases and 849 normal wrists) from 2,340 adult patients were ultimately included in the final dataset.

Data preparation

Each plain AP wrist radiograph, originally stored as a Digital Imaging and Communications in Medicine (DICOM) file, was exported as a Joint Photographic Experts Group (JPEG) file with a matrix size of 600 x 800 pixels from the Picture Archiving and Communication System (PACS) by using eWorld Viewer (TomTaw Tech, Ningbo, China).

For further analyses, 1,491 images with DRFs and 849 images without DRFs (randomized with the Research Randomizer program, <http://www.randomizer.org>) were randomly divided into an original training dataset of 2,040 images (1,341 images with DRFs and 699 images without DRFs) and a test dataset of 300 images (150 images with DRFs and 150 images without DRFs).



Figure 1. A wrist radiograph was manually annotated with a red rectangle as the ground truth bound and automatically annotated with a blue rectangle as the candidate bound. The red rectangle and blue rectangle represent edges of the region of interest (ROI) detected by the orthopedists and edges of the ROI detected by Faster R-CNN, respectively.

Training the CNN models

The detailed experimental environment is described in the Appendix (see Supplementary data).

Training the Faster R-CNN (Region-based CNN)

Faster R-CNN (Ren et al. 2017) technology, one of the fast object detection algorithms based on deep learning, has excellent performance in locating the regions of interest (ROIs) on graphics. In this study, we trained and tested Faster R-CNN as an auxiliary algorithm to the diagnostic CNN model. The detailed training procedure of the Faster R-CNN model is shown in the Appendix.

Validation of the Faster R-CNN

A regression analysis (Mitra et al. 2018) was used to assess the training process of Faster R-CNN. The mean square error (MSE) (Kumar et al. 2018) was calculated to measure the loss of Faster R-CNN in the automatic annotation of the ROI.

The test dataset of 300 images, including 150 images with fractures and 150 images without fractures, was used to evaluate the capacity of the trained Faster R-CNN model in automatic annotation of the ROI on images. First, using LabelImg (<https://github.com/tzutalin/labelImg>), 2 orthopedists with more than 5 years of orthopedic professional experience annotated each image's ROI as the ground truth bounds (GTBs), in which the whole distal radius was definitely encased. Then, a candidate boundary (CB) on each image with a GTB was annotated as the automatically detected ROI by the trained Faster R-CNN (Figure 1). The matrix sizes of the identified ROIs ranged from 207 to 223 pixels in width and from 208 to 231 pixels in height, respectively. The Intersection of the Union (IOU) (Mitra et al. 2018) was calculated, as illustrated



Figure 2. The formula with which the Intersection of the Union (IOU) was calculated.

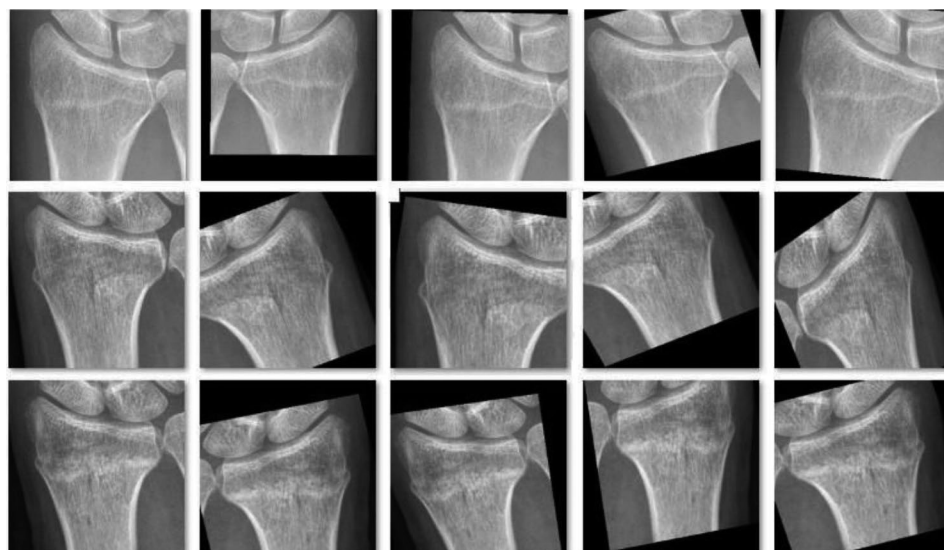


Figure 3. A typical example of the augmentation on 1 image from the annotated training dataset during the training course of Inception-v4.

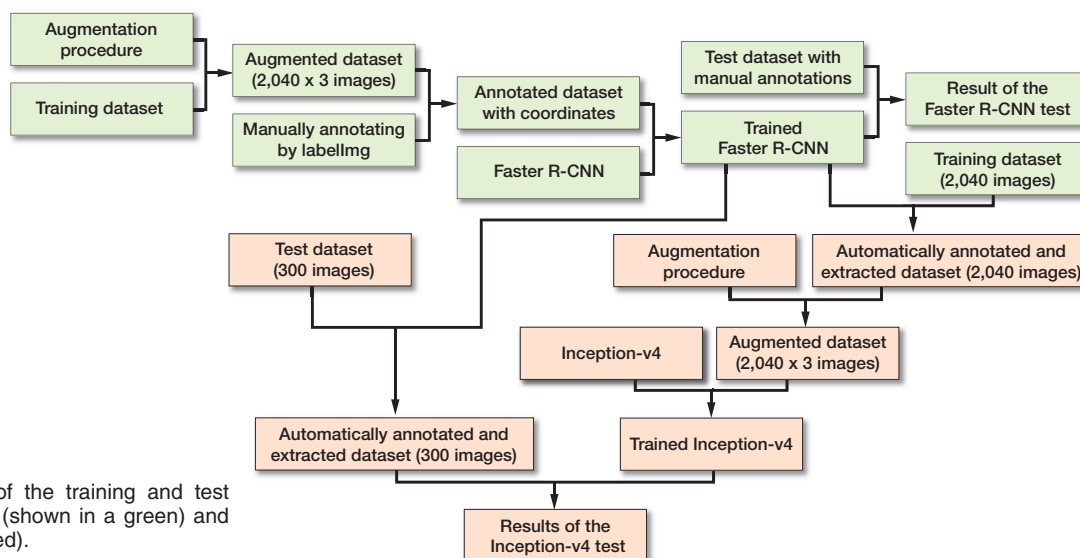


Figure 4. Flow diagram of the training and test courses of Faster R-CNN (shown in a green) and Inception-v4 (shown in a red).

in Figure 2, to statistically evaluate the trained Faster R-CNN, with a value greater than 0.5 indicating success in detecting the ROI on an image.

Training the diagnostic CNN model

We used Inception-v4 (Szegedy et al. 2017) as the diagnostic model, which has achieved state-of-the-art results in recent image classification contests.

In this study, only the images’ ROIs automatically annotated by Faster R-CNN were used as the recognition targets; after the ROI was extracted, the rest region on each initial image was discarded as unnecessary interference factors and noises. Since the areas where a DRF would occur were focused on, the Inception-v4 model’s training process of distinguishing images with fractures from normal images was much faster and more accurate than analyzing the entire image.

First, each initial image in the original training dataset, including 1,341 images with DRFs and 699 images without DRFs, was automatically annotated by the trained Faster R-CNN. The result of the annotations on 2,040 images was reviewed by 2 orthopedists, and each distal radius region was then confirmed to be appropriately encased in the bounds.

The ROIs extracted from all the annotated images were resized to 200 x 200 pixels, and stored as JPEG files, which were then augmented via the same technique as that used in the training of Faster R-CNN (Figure 3). Finally, there were 6,120 images in the data pool as the final training dataset for the Inception-v4 model, including 4,023 images with DRFs and 2,097 images without DRFs; 15% of the dataset was randomly selected into the validation dataset. The summary of the training course is illustrated in Figure 4. The detailed training procedure of the Inception-v4 model is shown in the Appendix.

Evaluation of the diagnostic performance of Inception-v4

First, each initial image in the original test dataset, including 150 images with DRFs and 150 images without DRFs, was automatically annotated by the trained Faster R-CNN. The result of the annotations on 300 images was reviewed by 2 orthopedists, and each distal radius region was then confirmed to be appropriately encased in the bounds. The ROIs extracted from the 300 annotated images were all resized to 200 x 200 pixels, and stored as JPEG files, consisting of a new test dataset.

The final analysis of the trained Inception-v4 model was performed using the new test dataset of 300 images to inspect its ability to discern images with fractures from the normal images. Each image was analyzed using the trained Inception-v4 model, which resulted in a score representing the likelihood that the image would be classified as “with a DRF” or “without a DRF.” This score had a continuous value between 0 and 1. The receiver operating characteristic (ROC) curve was generated using a Python script (<https://www.python.org>), and the AUC was determined.

Evaluation of the performance of the medical professionals

We set up a group of radiologists and a group of orthopedists to compare their results with those of the CNN to evaluate its diagnostic performance. The groups consisted of 3 radiologists who had at least 3 years of radiological professional experience and had passed the intermediate certificate exams and 3 orthopedists (none of whom participated in the validating process of review) with more than 5 years of orthopedic professional experience. The detailed procedure is described in the Appendix. Each image in the new test dataset was diagnosed as either “with a DRF” or “without a DRF.” In situations where disagreements arose in the same group regarding the diagnoses, the final decisions were made by a majority vote.

Comparison of the results of Inception-v4 and those of the medical professionals

After the ROC curve of Inception-v4 had been generated, the diagnostic cut-off at a threshold designed to maximize the Youden Index was set, and sensitivity, specificity, and accuracy of the Inception-v4 model were then calculated and statistically compared with these of the human groups.

Statistics

The SPSS software (version 22.0, IBM Corp, Armonk, NY, USA) was used to perform all statistical analyses. The demographic characteristics of all patients enrolled in this study are presented as mean (95% confidence intervals (CIs)) for age and count (percentage) for sex. P-values were derived from 1-way analysis of variance for age and chi-square tests for sex. The significance level was set at $p < 0.05$.

The IOU of Faster R-CNN and AUC of Inception-v4 were calculated and described in terms of the means and CIs. CIs of the distributions for the 4 kinds of outcomes and for the

Table 1. Demographic data of the whole dataset with 2,340 patients enrolled in this study

Factor	Patients with DRFs (n = 1,491)	Patients without DRFs (n = 849)	Total (n = 2,340)	Comparison (p-value)
Age, mean (CI)	48 (48–49)	48 (47–49)	48 (48–49)	0.4
Sex, n (%)				
Male	833 (56)	533 (63)	1,366 (58)	< 0.01
Female	658 (44)	316 (37)	974 (42)	

DRFs = distal radius fractures.

CI = 95% confidence interval.

P-values were derived from 1-way analysis of variance for age and chi-square tests for sex.

differences in the outcomes between the CNN model and each human group were computed via bootstrapping with 10,000 bootstraps. Comparisons between the CNN model and each human group were performed using a 1-way analysis of variance, followed by Dunnett’s test for multiple comparison with the significance level set at $p < 0.05$.

Ethics, funding, and potential conflicts of interest

The Ningbo Lihuili Hospital Ethics Committee approved the study (LH2018-039). Financial support for the study was from the Ningbo Natural Science Fund (No.2018A610164). All authors declare no conflicts of interest.

Results

Demographic data of the included patients

All of the patient radiographs (1,366 men and 974 women) were kept anonymous throughout this study. The patients’ mean age at the time they took the radiographs was 48 years (20–88). No statistically significant difference was found in age ($p = 0.4$) between the group of patients with fractures and group of patients without fractures, but there was significant difference in sex between the 2 groups ($p < 0.01$) (Table 1).

Performance of Faster R-CNN

The learning courses of Faster R-CNN in the final training and validation datasets are shown in the Appendix. In the test dataset, the average IOU value of Faster R-CNN was 0.87 (CI 0.86–0.87), and none of the IOU values was less than 0.5. 2 orthopedists reviewed each annotated CB, which was confirmed by encasing the whole distal radius on each image.

Performance of the Inception-v4 model

The learning courses of Inception-v4 in the final training and validation datasets are shown in the Appendix.

The ROC curve for the test output of Inception-v4 is plotted in Figure 5, and the AUC was 0.96 (CI 0.94–0.99). At the optimal cut-off point, the value of the threshold was 0.64.

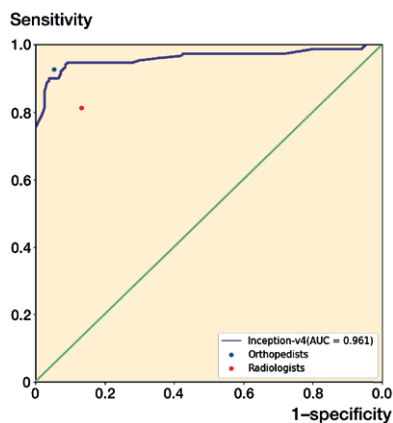


Figure 5. The receiver operating characteristic (ROC) curve for the test output of the Inception-v4 model. The dots on the plot represent the sensitivity and 1-specificity of the human groups (the blue dot represents the orthopedists' group; the red dot represents the radiologists' group). The sensitivity/1-specificity dot of the radiologists' group lies below the ROC curve of the Inception-v4 model, and the sensitivity/1-specificity dot of the orthopedists' group lies above the ROC curve of the Inception-v4 model.

Table 3. Performance differences in the outcomes between Inception-v4 and each human group

Factor	Differences between Inception-v4 and orthopedists Difference ^a (CI)	Differences between Inception-v4 and radiologists Difference ^b (CI)
Accuracy (%)	-1 (-5 to 3)	9 (3–15)
Sensitivity (%)	-3 (-9 to 4)	9 (1–16)
Specificity (%)	1 (-5 to 7)	9 (3–16)
Youden Index	-0.01 (-0.09 to 0.06)	0.18 (8–27)

CI = 95% confidence interval.

^a Difference = (mean of the outcome of Inception-v4) – (mean of the outcome of orthopedists).

^b Difference = (mean of the outcome of Inception-v4) – (mean of the outcome of radiologists).

Table 2. Diagnostic performance of the model and human groups

Factor	Inception-v4	Orthopedists	Radiologists	F-value	p-value
Accuracy (%) [CI]	279/300 (93) [90–96]	281/300 (94) [91–96]	252/300 (84) [80–88] ^a	10.19	< 0.001
Sensitivity (%) [CI]	135/150 (90) [85–95]	139/150 (93) [89–97]	122/150 (81) [75–87] ^a	5.07	0.007
Specificity (%) [CI]	144/150 (96) [93–99]	142/150 (95) [91–98]	130/150 (87) [81–92] ^a	4.82	0.009
Youden Index (CI)	0.86 (0.80–0.91)	0.87 (0.82–0.93)	0.68 (0.61–0.75) ^a	11.62	< 0.001

CI = 95% confidence interval.

^a Statistically significant in a comparison of Inception-v4 and each human group (results from Dunnett's test).

Comparison between the Inception-v4 model and human performance

The model showed a superior capacity compared with the radiologists' group to distinguish images with DRFs from normal images in terms of accuracy, sensitivity, specificity, and Youden Index. The CNN model presented a similar diagnostic capability to that of the orthopedists in terms of the outcomes (Tables 2 and 3).

Discussion

In our study, both deep learning models demonstrated an excellent ability to recognize image traits in wrist radiographs. The trained Faster R-CNN, which had a 100% success rate in automatically annotating the ROIs on images from the test dataset, acted as a valid auxiliary algorithm to the Inception-v4 model, which was trained to distinguish images with DRFs from normal images. The Inception-v4 model exhibited a similar diagnostic capability to that of the orthopedists and superior performance to that of the radiologists.

Previous studies investigating the feasibility of applying CNNs to detect fractures on radiographs showed promising results, consistent with those of our study. Kim and MacKinnon (2018) trained Inception-v3 to recognize wrist fractures on lateral wrist radiographs; their results showed that the value of AUC was 0.954 and the maximized values of the sensitivity and specificity were 0.9 and 0.88, respectively. Olczak et al. (2017) performed a study in which a Visual Geometry Group 16-layer (VGG_16) network was trained to detect fractures on hand, wrist, and ankle radiographs with an accuracy of 83%, similar to the performance of the radiologists (who had an accuracy of 82%). Chung et al. (2018) evaluated the ability of the Residual Network (ResNet) model to detect and classify proximal humerus fractures using shoulder radiographs. The CNN showed superior top-1 accuracy, an accuracy of 96%, which was greater than that of the orthopedists (93%). Urakawa et al. (2019) conducted a study in which they compared the capacities of the VGG_16 network and orthopedic surgeons in detecting intertrochanteric fractures on radiographs, revealing the diagnostic performance of the CNN; the CNN had an accuracy of 96%, which exceeded that of ortho-

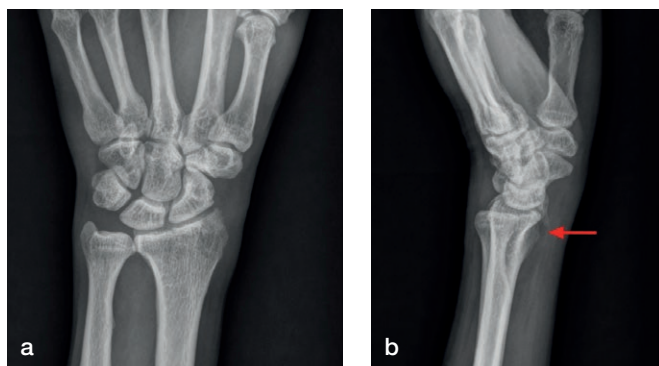


Figure 6. The same wrist with a DRF in the anterior–posterior view radiograph (a) and in the lateral view radiograph (b). The hidden DRF in the anterior–posterior view was apparent in the lateral view (the fracture is shown by the red arrow).

pedic surgeons, who had an accuracy of 92%. All the previous studies mentioned prepared the images that were used in the training datasets and test datasets by manually cropping them into certain matrix sizes before the images were input into the CNNs. Since the images were uniform and had concentrated matrix sizes, the ROIs on the images were recognized faster and more accurately by the deep learning models, thereby remarkably improving the efficiency of the CNNs in the learning and test procedures. We employed and trained the Faster R-CNN model to automatically annotate the ROIs on images as a reliable substitution for manual cropping, which resulted in a low processing time and decreased bias (Urakawa et al. 2019). There is a great potential for the Inception-v4 model to be combined with Faster R-CNN to detect DRFs in clinical practice, where wrist radiographs with both ROIs and irrelevant regions are presented.

We plotted the ROC curve for the test output; at the optimal cut-off point, Inception-v4 showed a sensitivity of 90% (135/150), which is much lower than the specificity (96%). In such conditions, some wrist AP radiographs with DRFs appeared to be misdiagnosed as normal images by the model, resulting in a delay in essential treatment for injured patients. After reviewing the 15 images with fractures, whose predicted values by the Inception-v4 model were less than the threshold (0.64), we found that 5 of them displayed an absence of apparent fracture traits (e.g., fracture lines or fracture fragment displacement). However, such traits were visible on the lateral radiographs corresponding to the AP images, as shown in Figure 6. The ensemble of model analyses using AP and lateral radiographs has the potential to enhance the sensitivity in fracture detection.

In the total enrolled dataset, there was a statistically significant difference between the group of patients with DRFs and the group of patients without DRFs in sex (Table 1). This difference may affect the results of training and testing the CNN models due to difference in anatomical traits in distal radius between the male and the female (Oppermann et al. 2015). But

we cannot declare to what extent the effect of gender difference would be on the results in this study.

There are several limitations in our study. First, the original sample size in our dataset was small. However, we did not increase the original sample size by obtaining new wrist radiographs from other medical centers to maintain uniformity in the image quality. This small sample size might restrict the improvement of the CNN's performance in the training and test procedures. Data augmentation was used to address the sample size issue, since it can reduce over-fitting and improve performance (Wong et al. 2016). Second, the assessment of the diagnostic performance of the deep learning models was based on anterior–posterior wrist radiographs, so the procedure may not represent a practical scenario. Typically, at least 2 wrist radiographs (an anterior–posterior image and a lateral image) are obtained by the reader to review. We will investigate whether the performance of the CNN would improve when the dataset consists of anterior–posterior wrist radiographs and matched lateral radiographs in our next planned project. Finally, we trained Inception-v4 to simply distinguish images with DRFs from normal images. The deep learning algorithm could accurately classify proximal humerus fractures based on Neer's classification on shoulder radiographs (Chung et al. 2018), so as part of our next project we will train the CNN model to classify DRFs based on 1 particular fracture classification system.

In summary, the network exhibited a similar diagnostic capability to that of the orthopedists and a superior performance to that of the radiologists in distinguishing AP wrist radiographs with DRFs from normal radiographs under limited conditions. Further studies are required to determine the feasibility of applying the diagnostic network with the object detection algorithm as an auxiliary in clinical practice under extended conditions.

Supplementary data

The inclusion and exclusion criteria for this study and the Appendix are available as supplementary data in the online version of this article, <http://dx.doi.org/10.1080/17453674.2019.1600125>

Conception and design of this study: KG, YL, and DX. Collection and preparation of the datasets: KG, DX, YL, YS, TZ, KH, and KZ. Training and testing of models: YL, DX, MB, YS, TZ, and KH. Analysis and interpretation of the results and data: KG, YL, LP, and WW.

Acta thanks Martin Gerdin Wörnberg, Max Gordon and Takaaki Urakawa for help with peer review of this study.

Chung S W, Han S S, Lee J W, Oh K S, Kim N R, Yoon J P, Kim J Y, Moon S H, Kwon J, Lee H J, Noh Y M, Kim Y. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop* 2018; 89(4): 468-73.

- Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, Watanabe R, Okiyama N, Ohara K, Fujimoto M. Deep learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumor diagnosis. *Br J Dermatol* 2019; 180(2): 373-81.
- He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. 2015 IEEE International Conference on Computer Vision; 2015.
- Hua K L, Hsu C H, Hidayati S C, Cheng W H, Chen Y J. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Onco Targets Ther* 2015; 8: 2015-22.
- Kim D H, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* 2018; 73(5): 439-45.
- Kumar S N, Lenin Fred A, Sebastin Varghese P. Compression of CT images using contextual vector quantization with simulated annealing for telemedicine application. *J Med Syst* 2018; 42(11): 218.
- Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; 284(2): 574-82.
- Li Y, Shen L. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors (Basel)* 2018; 18(2).
- 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* 2018; 26(3): 83-93.
- Mitra A, Banerjee P S, Roy S, Roy S, Setua S K. The region of interest localization for glaucoma analysis from retinal fundus image using deep learning. *Comput Methods Programs Biomed* 2018; 165: 25-35.
- Nishio M, Sugiyama O, Yakami M, Ueno S, Kubo T, Kuroda T, Togashi K. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLoS One* 2018; 13(7): e0200721.
- Olczak J, Fahlberg N, Maki A, Razavian A S, Jilert A, Stark A, Skoldenberg O, Gordon M. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop* 2017; 88(6): 581-86.
- Oppermann J, Bredow J, Beyer F, Neiss W F, Spies C K, Eysel P, Dargel J, Wacker M. Distal radius: anatomical morphometric gender characteristics. Do anatomical preshaped plates pay attention on it? *Arch Orthop Trauma Surg* 2015; 135 (1): 133–39.
- Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2017; 39(6): 1137-49.
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg A C, Li F F. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015; 115(3): 211-52.
- Szegedy C, Ioffe S, Vanhoucke V, Alemi A A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*; 2017.
- Ting D S W, Cheung C Y, Lim G, Tan G S W, Quang N D, Gan A, Hamzah H, Garcia-Franco R, San Yeo I Y, Lee S Y, Wong E Y M, Sabanayagam C, Baskaran M, Ibrahim F, Tan N C, Finkelstein E A, Lamoureux E L, Wong I Y, Bressler N M, Sivaprasad S, Varma R, Jonas J B, He M G, Cheng C Y, Cheung G C M, Aung T, Hsu W, Lee M L, Wong T Y. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017; 318(22): 2211-23.
- 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(2): 239-44.
- Waever D, Madsen M L, Roling J H D, Borris L C, Henriksen M, Nagel L L, Thorninger R. Distal radius fractures are difficult to classify. *Injury* 2018; 49(Suppl .1): S29-S32.
- Wang Z, Meng Q, Wang S, Li Z, Bai Y, Wang D. Deep learning-based endoscopic image recognition for detection of early gastric cancer: a Chinese perspective. *Gastrointest Endosc* 2018; 88(1): 198-99.
- Wong S C, Gatt A, Stamatescu V, McDonnell M D. Understanding data augmentation for classification: when to warp? *International Conference on Digital Image Computing: Techniques and Applications*; 2016.
- Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. *Exp Dermatol* 2018; 27(11): 1261-7.