Development and Validation of a Deep Learning Model Using Convolutional Neural Networks to Identify Scaphoid Fractures in Radiographs

Alfred P. Yoon, MD; Yi-Lun Lee, MS; Robert L. Kane, BS; Chang-Fu Kuo, MD, PhD; Chihung Lin, PhD; Kevin C. Chung, MD, MS

Abstract

IMPORTANCE Scaphoid fractures are the most common carpal fracture, but as many as 20% are not visible (ie, occult) in the initial injury radiograph; untreated scaphoid fractures can lead to degenerative wrist arthritis and debilitating pain, detrimentally affecting productivity and quality of life. Occult scaphoid fractures are among the primary causes of scaphoid nonunions, secondary to delayed diagnosis.

OBJECTIVE To develop and validate a deep convolutional neural network (DCNN) that can reliably detect both apparent and occult scaphoid fractures from radiographic images.

DESIGN, SETTING, AND PARTICIPANTS This diagnostic study used a radiographic data set compiled for all patients presenting to Chang Gung Memorial Hospital (Taipei, Taiwan) and Michigan Medicine (Ann Arbor) with possible scaphoid fractures between January 2001 and December 2019. This group was randomly split into training, validation, and test data sets. The images were passed through a detection model to crop around the scaphoid and were then used to train a DCNN model based on the EfficientNetB3 architecture to classify apparent and occult scaphoid fractures. Data analysis was conducted from January to October 2020.

EXPOSURES A DCNN trained to discriminate radiographs with normal and fractured scaphoids.

MAIN OUTCOMES AND MEASURES Area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity. Fracture localization was assessed using gradient-weighted class activation mapping.

RESULTS Of the 11,838 included radiographs (4917 [41.5%] with scaphoid fracture; 6921 [58.5%] without scaphoid fracture), 8356 (70.6%) were used for training, 1177 (9.9%) for validation, and 2305 (19.5%) for testing. In the testing test, the first DCNN achieved an overall sensitivity and specificity of 87.1% (95% CI, 84.8%-89.2%) and 92.1% (95% CI, 90.6%-93.5%), respectively, with an AUROC of 0.955 in distinguishing scaphoid fractures from scaphoids without fracture. Gradient-weighted class activation mapping closely corresponded to visible fracture sites. The second DCNN, which examined negative cases from the first DCNN, achieved a sensitivity and specificity of 79.0% and 71.6% with an AUROC of 0.810. This 2-stage DCNN model correctly identified 90% of occult fractures.

CONCLUSIONS AND RELEVANCE In this study, DCNN models were trained to identify scaphoid fractures. This suggests that such models may be able to assist with radiographic detection of occult scaphoid fractures that are not visible to human observers and to reliably detect fractures of other small bones.

Key Points

Question Can deep convolutional neural networks (DCNNs) detect occult scaphoid fractures not visible to human observers?

Findings In this diagnostic study of 11,838 scaphoid radiographs, the DCNN trained to distinguish scaphoid fractures from scaphoids without fracture achieved an overall sensitivity and specificity of 87.1% and 92.1%, respectively, with an area under the receiver operating curve (AUROC) of 0.955; a second DCNN, which examined negative cases from the first DCNN, achieved a sensitivity and specificity of 79.0% and 71.6% with an AUROC of 0.810. This 2-stage DCNN model correctly identified 90% of occult fractures.

Meaning These findings suggest that DCNNs can be trained to reliably detect fractures of small bones, such as scaphoids, and may be able to assist with radiographic detection of occult fractures that are not visible to human observers.

Open Access. This is an open access article distributed under the terms of the CC-BY License.
Introduction

Scaphoid fractures are the most common carpal fracture, with the highest prevalence in active, young adult men. Although the estimated annual incidence is 5 scaphoid fractures per 10,000 people, the actual incidence is likely higher, as physicians fail to detect as many as 20% of scaphoid fractures on the initial radiograph. Failure to detect and treat scaphoid fractures can lead to detrimental consequences for patients. Occult scaphoid fractures, ie, scaphoid fractures that are difficult to detect on radiographs, are increasingly recognized as the etiology responsible for a large proportion of scaphoid nonunions. Patients with scaphoid nonunions are susceptible to further complications, including degenerative wrist arthritis, chronic wrist pain, and carpal collapse. As scaphoid fractures have the greatest prevalence in the younger working population, socioeconomic burden and productivity loss associated with scaphoid fractures are substantial; the yearly estimated economic burden for nonoperative scaphoid fracture management in the United States is approximately $58 million.

When a scaphoid fracture is suspected, patients are managed with cast immobilization for approximately 2 weeks with interval follow-up radiograph. Although this empirical approach facilitates the management of potential occult scaphoid fractures, it leads to unnecessary immobilization for 80% of patients who never had a fracture. During these weeks, patients accumulate indirect costs, such as taking time off work and traveling long distances for clinic visits. Several cost-effectiveness analyses suggest that magnetic resonance imaging (MRI) for a suspected occult scaphoid fracture may be more cost-effective than empirical immobilization and interval radiographs. Nevertheless, MRI remains among the most expensive imaging modalities, and its use must be determined judiciously. An inexpensive diagnostic test that is sensitive and specific for occult scaphoid fractures may improve patient outcomes, yield cost savings from obviating the need for advanced imaging, and prevent unnecessary immobilization.

Recent advances in the field of image analysis have demonstrated that computer models can assist and even outperform humans in detecting features of radiographs. This takes place through a process called deep learning, whereby computers can learn features and data patterns not readily visible to the human eye. A class of deep learning neural networks commonly applied to image recognition is an algorithm called a deep convolutional neural network (DCNN). In the medical field, DCNNs have been applied to identify diabetic retinopathy, accurately discriminate osteoporosis from nonosteoporosis, enhance detection of pulmonary tuberculosis, classify skin cancers, and identify fractures. Given that physicians are unable to detect 1 in 5 scaphoid fractures on radiographs, a DCNN that assists physicians with identifying scaphoid fractures would improve patient outcomes. To our knowledge, no study has tried to construct a DCNN that can identify radiographically occult fractures in any bone. In this study, we aimed to create a DCNN that could reliably detect both apparent and occult scaphoid fractures using plain radiographs. We hypothesized the DCNN would detect occult scaphoid fractures with greater than 70% sensitivity and specificity. Given that unaided identification of occult scaphoid fractures on plain radiographs is extremely difficult or nearly impossible, a 70% detection accuracy would have clear clinical consequences.

Methods

Data Set
Hand radiographs from Chang Gung Memorial Hospital (CGMH) and Michigan Medicine (MM) between January 2001 and December 2019 were collected in Digital Imaging and Communications in Medicine (DICOM) format. Posteroanterior or scaphoid view hand radiographs querying for scaphoid fractures in adults older than 18 years were included. A group of senior musculoskeletal radiologists provided final image interpretations. Radiographs with ambiguous or conflicting reports were reviewed by a hand surgeon (K.C.C.), and final diagnoses were made based on the surgeon's...
interpretation, subsequent imaging, and follow-up clinic examination notes. Any confirmatory imaging with repeated radiographs, computed tomography (CT) scans, and MRI were also obtained and reviewed to ascertain the veracity of the initial radiology reading and confirm the ground truths of occult fractures. Subsequently, the occult fracture data set was populated by compiling the injury radiograph (first radiograph) taken at these patients’ initial hospital visits after traumatic injury. Radiographs were deidentified before transmission to the investigators for model development. This study was considered exempt from regulatory review by the University of Michigan institutional review board and CGMH. It was considered exempt from informed consent because it was secondary research, for which consent is not required. This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline for prediction model development.

In total, there were 4183 DICOM images from 2176 patients treated at MM and 13 339 DICOM images from 5553 patients treated at CGMH. The following images were excluded: (1) oblique or lateral views; (2) images with poor quality (ie, poor image detail, contrast, or inappropriate film darkness); (3) fractures older than 4 weeks; (4) images with unclear ground truth without confirmatory images or contradictory radiologic diagnoses; (5) images of chronic hand conditions with bony changes around the scaphoid; (6) images of patients with psoriatic arthritis or rheumatoid arthritis; (7) images with external immobilization (casts, splints, external fixations); and (8) images with hardware (screws, plates, wires, pins). All radiographic qualities were confirmed by 2 team members, 1 hand surgeon (A.P.Y.) and 1 machine learning engineer (Y.L.). After exclusion, 8329 images from 3777 patients treated at CGMH and 3509 images from 1943 patients treated at MM remained for inclusion. Of those, 2305 images from 1137 patients treated at both institutions were retained for the evaluation of the overall pipeline. The remaining images were used in the training and evaluation of the detection and classification models. **Figure 1** depicts the image selection and data allocation process.

**Detection Model and Image Cropping**

Only scaphoid images were needed for DCNN development; therefore, a scaphoid detection model based on Cascade R-CNN (Region-based Convolutional Neural Network)\(^\text{17}\) was trained and used to isolate the scaphoid in a bounding box in hand radiographs. A total of 2851 images from 1572 patients treated at MM and 6682 images from 3011 patients treated at CGMH were used for training. The images were standardized by image depth and were converted from grayscale to red-green-blue (RGB) color. To generalize the model, random flip, scale, random brightness, rotation (<15°), flip, resizing, and standardization were applied. Because other data augmentation techniques did not improve model performance, they were not implemented. The images were subsequently cropped around the bounding box to isolate the scaphoid.

**Classification Model**

From the ImageNet Large Scale Visual Recognition Competition,\(^\text{18}\) several high-performing DCNNs were identified, such as AlexNet,\(^\text{19}\) Inception v3,\(^\text{20}\) and VGG-16.\(^\text{21}\) A recent study\(^\text{22}\) showed that another DCNN, EfficientNet, achieved higher accuracy than Inception v3 (95.6% vs 94.4%) with half as many parameters. For these reasons, we used the EfficientNetB3 DCNN architecture to train our model.

We developed 2 classification models. The first apparent fracture model was trained on images of radiographically apparent fractures. Because occult fractures are not readily observable by human experts and are confirmed using secondary imaging modalities, such as MRI, we developed a second DCNN that was based on the first DCNN and further adjusted using occult fracture images. This 2-stage design aimed to maximize the detection of occult scaphoid fractures.
Apparent Fracture Model
The apparent fracture model was trained to detect radiographically evident fracture. A total of 3991 scaphoid fracture radiographs from 1911 patients (2435 images [61.0%] from CGMH and 1556 images [39.0%] from MM) and 5542 normal scaphoid radiographs from 2672 patients (4247 images [76.6%] from CGMH and 1295 images [23.4%] from MM) were used for training and validation of the apparent fracture model. Because the number of CGMH images was approximately 2.5-fold larger than the number of MM images, we oversampled MM images to balance the sampling weight. The model was trained based on MM pretrained weights using the AdamW optimizer\textsuperscript{23} with learning rate set as $2 \times 10^{-5}$; weight decay, $1 \times 10^{-6}$; and batch size, 16. The learning rate was reduced if the validation loss did not improve for 6 epochs. The training process concluded when the model performance did not improve after 15 epochs. The model outputted predicted scores for presence or absence of fracture and a gradient-weighted class activation mapping (Grad-CAM) image\textsuperscript{24} to visualize the possibility of a pixel to represent fracture (Figure 2). After overlaying with the original image, high probability areas were highlighted as the most likely fracture sites.

Occult Fracture Model
To facilitate recognition of occult fractures, the model was further trained with 49 occult fracture images from 27 patients treated at CGMH and 90 occult fracture images from 51 patients treated at MM with 417 control images from 351 patients treated at CGMH and 139 control images from 118 patients treated at MM. The ratio of case to control was 1:4. Of these, 565 images (81.3%) from 451 patients were used in training and 130 images (18.7%) from 96 patients were used for validation. Control images for the testing set were randomly picked from a larger pool of normal radiographs to
better mimic real-life scenarios. The previously described image augmentation techniques were also applied to the occult model. To mitigate the small ratio of case (occult fracture) to control images, occult fracture images were oversampled once. Based on the pretrained weights of the apparent fracture model, the occult model was trained with AdamW optimizer with learning rate set as $1 \times 10^{-5}$; weight decay, $1 \times 10^{-6}$; and batch size, 16. The learning rate, training policy, and output types were identical to the apparent fracture model.

**Statistical Analysis**

The DCNN’s fracture detection accuracy was evaluated by the area under the receiver operating characteristic curve (AUROC), sensitivity, and specificity. Optimal cutoff values for sensitivity and specificity were assessed with the Youden index, which is the point on the ROC that maximizes both sensitivity and specificity. Fracture localization accuracy was estimated as a secondary end point with Grad-CAM. Confidence intervals for sensitivity and specificity are exact Clopper-Pearson confidence intervals.25,26

The analysis was carried out on a DGX-1 server with ubuntu version 18.04 (Canonical) operating system using Python version 3.7 (Python Software Foundation). Four NVIDIA V100 were used to train the model. The training and inference of the detection model were completed under mmddetection version 1.0.0 and Pytorch version 1.4 framework. The bony landmarks were manually annotated using the labelme package,27 which was modified to accept original DICOM images. For training of the classification model, the tensorflow version 2.2 framework was used. Automatic mixed precision was used in the training process. Pydicom version 1.4.2,28 tensorflow, and opencv version 4.1.0 were used for image processing.

**Results**

Of the 11,838 included radiographs (4917 [41.5%] with scaphoid fracture; 6921 [58.5%] without scaphoid fracture), 8356 (70.6%) were used for training, 1177 (9.9%) for validation, and 2305 (19.5%) for testing. The complete pipeline of our model was as follows. First, the detection model detected and cropped the scaphoid; then, the apparent fracture model classified apparent fractures, after which the occult fracture model identified occult fractures for images predicted as normal by the apparent fracture model. Instead of evaluating the 3 models separately, 2305 images that were not used in the training process were used to test the full pipeline. The test data set included 1379
control images (59.8%), 904 apparently fractured scaphoid images (39.2%), and 22 occult fracture images (1.0%). All test images except for 1 from the MM data set passed the scaphoid detection test. The remaining 2304 images were used for testing the pipeline (eFigure in the Supplement).

**Apparent Fracture Model Performance**

The apparent fracture model test data set included 1379 normal, 904 apparently fractured, and 22 occultly fractured scaphoid images (eTable 1 in the Supplement). The apparent fracture model correctly predicted 1271 of 1379 true-normal images (92.2%), 795 of 903 images (88.0%) of apparent fractures, and 11 of 22 images (50.0%) of occult fractures (eTable 2 in the Supplement). The model achieved an AUROC of 0.955 (Figure 3) with a sensitivity of 87.1% (95% CI, 84.8%-89.2%) and specificity of 92.1% (95% CI, 90.6%-93.5%) when tested with a set of normal and apparent scaphoid fractures. The positive predictive value (PPV) was 88.2% (95% CI, 86.1%-90.0%), and the negative predictive value (NPV) was 91.4% (95% CI, 90.0%-92.7%), with a 40.0% fracture prevalence (Table). The localization of the fracture lines was also demonstrated by Grad-CAM images (Figure 4A).

**Occult Fracture Model Performance**

We tested the occult fracture model using 1390 images that were predicted as normal (119 false-negative and 1271 true-negative) by the apparent fracture model; therefore, this test data set included 1271 normal images (91.4%), 108 apparently fractured images (7.8%), and 11 occultly fractured scaphoid images (0.8%) (eTable 1 in the Supplement). Among these, the occult fracture model correctly predicted 910 of 1271 true-normal images (71.6%), 85 of 108 images (78.7%) of apparent fractures, and 9 of 11 images (81.8%) of occult fractures (eTable 2 in the Supplement). The occult fracture model achieved an AUROC of 0.810 (Figure 3) with a sensitivity of 79.0% (95% CI, 70.6%-86.0%) and specificity of 71.6% (95% CI, 69.0%-74.1%). The PPV was 20.6% (95% CI, 18.7%-22.8%)

**Table. Model Performance Detecting Apparent and Occult Scaphoid Fractures**

| Measure                  | Apparent fracture model (95% CI), % | Occult fracture model (95% CI), %a | Overall pipeline (95% CI), % |
|--------------------------|------------------------------------|------------------------------------|-------------------------------|
| Sensitivity              | 87.1 (84.8-89.2)                   | 79.0 (70.6-86.0)                   | 97.2 (95.9-98.2)              |
| Specificity              | 92.1 (90.6-93.5)                   | 71.6 (69.0-74.1)                   | 66.0 (63.4-68.5)              |
| Positive predictive value| 88.2 (86.1-90.0)                   | 20.6 (18.7-22.8)                   | 65.7 (64.1-67.4)              |
| Negative predictive value| 91.4 (90.0-92.7)                   | 97.3 (96.3-98.1)                   | 97.2 (96.0-98.1)              |
| AUROC                    | 0.96                               | 0.81                               | NAa                           |

Abbreviations: AUROC, area under the receiver operating characteristic curve; NA, not applicable.

a The occult fracture model performance included both apparent and occult fractures.

b Unable to calculate AUROC.
18.7%-22.8%), and the NPV was 97.3% (95% CI, 96.3%-98.1%) for a disease prevalence of 8.6% (Table). The model accurately localized the fracture site in Grad-CAM images when an occult fracture was detected. The predicted fracture location on the Grad-CAM images coincided with the fractures identified on follow-up CT scans and MRIs (Figure 4B).

Overall Pipeline Performance
The overall pipeline model yielded 900 true-positive, 469 false-positive, 26 false-negative, and 910 true-negative results. It achieved a sensitivity of 97.2% (95% CI, 95.9%-98.2%) and specificity of 66.0% (95% CI, 63.4%-68.5%). The PPV was 65.7% (95% CI, 64.1%-67.4%), and the NPV was 97.2% (95% CI, 96.0%-98.1%) for a disease prevalence of 40.2% (Table). Of the 22 occult images included in the test data set, the overall model correctly identified 20 occult fracture images and misclassified 2 as normal: a 90.9% accuracy rate of detecting occult scaphoid fractures.

Discussion
We developed a DCNN that differentiated between normal and fractured scaphoids with high specificity and sensitivity. With an AUROC of 0.955, the results of the apparent fracture model suggest that deep learning methods can detect small bone fractures, such as scaphoid fractures. Our occult fracture model also had a high accuracy, with an AUROC of 0.810, although this result must be interpreted cautiously because it originates from a test data set that included both apparent and occult fracture images. Nevertheless, given that the overall model correctly identified 20 of 22 occult fractures (90.9%), this suggests that neural networks may be able to detect fractures not visible to human observers. In addition, the model correctly localized the occult fracture sites, as seen in the Grad-CAM images (Figure 4). Although further clinical testing of the model is warranted, we propose that DCNNs may have the capacity to detect occult fractures and may outperform human observers in detecting them.

Our model pipeline consisted of 2 separate DCNNs, the apparent and occult fracture models. We intentionally implemented this 2-step process because the true prevalence of scaphoid fractures
in patients presenting with acute wrist injury is reported to be between 2% and 54%,\textsuperscript{29-31} whereas occult scaphoid fracture prevalence is likely to be even lower. Because the pretest probability of occult fractures is lower, passing the image through the apparent fracture model to detect any apparent scaphoid fractures will concentrate the pretest probability for the occult fracture model, increasing its diagnostic performance. In the clinical setting, this 2-step process could not only increase the probability of detecting a true occult fracture but also increase the model's sensitivity to rule out scaphoid fractures, especially in cohorts with low prevalence of occult scaphoid fractures, precluding the need for advanced imaging.

Neural networks have previously been applied to other musculoskeletal conditions. Several CNNs correctly classified distal radius fractures with an AUROC greater than 0.95,\textsuperscript{32,33} whereas others have reliably detected osteoporosis from hand radiographs.\textsuperscript{12} Although these illustrate the potential role of DCNNs in diagnostic radiology, a clinically meaningful application of this technology may not be possible for certain diagnoses. For example, distal radius fractures are relatively straightforward to detect on radiographs without computer vision. The standard of care for osteoporosis diagnosis is a dual-energy x-ray absorptiometry (DEXA) scan, which is comparable in cost to a hand radiograph ($125) and takes 10 minutes to administer.\textsuperscript{34} The reported sensitivity of a DEXA scan for osteoporosis is 98%,\textsuperscript{35} suggesting that a DCNN capable of detecting osteoporosis in hand radiographs with similar sensitivity may not be cost-effective compared with a DEXA scan when considering the added cost of the software. Similarly, Lindsey et al\textsuperscript{32} reported that the mean sensitivity and specificity of distal radius fracture detection by physicians are 80.8% and 91.5%, respectively. These estimates are likely even greater for hand surgeons and musculoskeletal radiologists, who regularly diagnose distal radius fractures. Scaphoid fractures, unlike some other musculoskeletal conditions, represent a suitable clinical dilemma for DCNNs because as many as 20% of these fractures are not readily visible to physicians on initial radiographs. This poses a unique challenge that can be overcome with computer vision.

Artificial intelligence (AI) is expected to resolve several practical challenges that radiologists confront daily. For example, approximately 40% of all inpatient imaging examinations are designated as requiring immediate attention.\textsuperscript{36} Such high-volume, urgent radiology interpretations can lead to observer fatigue that may diminish diagnostic accuracy, particularly in conditions, such as scaphoid fractures, for which radiographic diagnosis is already elusive. Preanalysis of radiographs with DCNNs could decrease observer fatigue and reduce missed fractures. DCNNs can make a prediction in seconds; therefore, clinical integration of these models should be effortless. Furthermore, AI can recognize complex image patterns and mathematical motifs that are not discernable to human eyes, facilitating detection of occult fractures.\textsuperscript{36} By contrast, trained radiologists assess new images based on knowledge of patterns learned from prior experience, which are vulnerable to human subjectivity.\textsuperscript{37} If DCNNs can assist physicians in reliably diagnosing occult fractures and elucidating obscure findings in other imaging modalities, immeasurable benefits to both patients and health care delivery could be achieved.

This DCNN benefited from a training data set composed of radiographs from 2 centers on 2 different continents, increasing both the diversity and power of the data set. Because the scaphoid was isolated from the hand radiographs and subsequently processed using image processing techniques, model performance was optimized while minimizing overfitting.\textsuperscript{38} Lastly, reliable ground truths were confirmed using conclusive radiology reports by 2 radiologists and/or follow-up imaging.

**Limitations**

This study has limitations. While our long-term aim is to create a tool that enhances the standard of care for scaphoid fracture and brings tangible benefits to these patients, further refinement and validation with multiple prospective data sets are needed before this tool can be integrated into clinical workflow. Although the DCNN was trained with a sizeable data set from 2 academic institutions, incorporating radiographs from additional populations will increase the accuracy and
generalizability of the DCNN. Because of their multilayered neural networks and abstractions, an inherent limitation of DCNNs is that the mechanistic steps of how the model reached its conclusions cannot be discerned.39 This is an important consideration because we must ensure that the model’s inferences are from the fracture site and not from irrelevant parts of the radiograph. However, the DCNN’s accuracy was demonstrated in multiple test data sets and Grad-CAM images, instilling confidence in the model’s conclusions. Indication bias is another potential limitation of our study, given that radiographs included in the training set belonged to patients with a high likelihood of scaphoid fracture. In addition, the high sensitivity and specificity of the occult model may be partially attributable to the selection of control images, which were selected based on low fracture probability predicted by the apparent fracture model. Because the apparent fracture model was highly confident that these images were normal, this may have facilitated detection of subtle features, such as occult fractures. In addition, it must be noted that the test data set for the occult fracture model included images of both apparently and occultly fractured scaphoids; therefore, the performance results for the occult fracture model are not exclusive representations of the model’s capacity to detect occult fractures. However, of the 22 occult fracture images in the test data set, 20 (90.9%) were correctly identified as having a fracture by the overall model, indicating that detection of occult fractures by DCNNs is possible. On the contrary, the relatively high false-positive rate, as evidenced by the low PPV, will likely improve with a larger training data set. But the current model’s high NPV suggests that occult scaphoid fractures are not missed, which is preferable for a clinical test, especially in low-resource settings where advanced imaging techniques are not readily available. Furthermore, lateral view radiographs were excluded to maintain a lower number of parameters, especially because most scaphoid fractures are visible on posteroanterior or scaphoid views.

Conclusions

In this study, we developed a DCNN to identify apparent scaphoid fractures on radiographs. It achieved high sensitivity and specificity, suggesting that DCNNs can be trained to reliably detect fractures in small bones. In addition, this study found that the DCNN could detect occult fractures that are not readily visible to physicians. This enhanced diagnostic capacity can help to solve medical problems with high monetary or quality-of-life costs and improve fracture care.
Statistical analysis: Lee, Kuo, Lin.

Obtained funding: Yoon, Kuo, Chung.

Administrative, technical, or material support: All authors.

Supervision: Yoon, Kuo, Chung.

Conflict of Interest Disclosures: Dr Chung reported receiving funding from the National Institutes of Health; receiving book royalties from Wolters Kluwer and Elsevier; and serving as a consultant for Axogen and Integra outside the submitted work. No other disclosures were reported.

Funding/Support: This research was funded by a joint Chang Gung Memorial Hospital–University of Michigan Medical Center grant (CORPG3K0201, awarded to Drs Lin and Chung) and the National Endowment for Plastic Surgery Grant from the Plastic Surgery Foundation (694527, awarded to Dr Yoon).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: The authors thank the Maintenance Project of the Center for Artificial Intelligence in Medicine (funded by grants CLRPG3H0012 and CIRPG3H0012 to Dr Kuo) at Chang Gung Memorial Hospital for statistical assistance and support.

REFERENCES

1. Rhemrev SJ, Ootes D, Beeres FJ, Meylaerts SA, Schipper IB. Current methods of diagnosis and treatment of scaphoid fractures. *Int J Emerg Med*. 2011;4(1):4. doi:10.1186/1865-1380-4-4

2. Kawamura K, Chung KC. Treatment of scaphoid fractures and nonunions. *J Hand Surg Am*. 2008;33(6):988-997. doi:10.1016/j.jhsa.2008.04.026

3. Shetty S, Sidharthan S, Jacob J, Ramesh B. ‘Clinical scaphoid fracture’: is it time to abolish this phrase? *Ann R Coll Surg Engl*. 2011;93(2):146-148. doi:10.1308/147870811X560886

4. Waeckerle JF. A prospective study identifying the sensitivity of radiographic findings and the efficacy of clinical findings in carpal navicular fractures. *Ann Emerg Med*. 1987;16(7):733-737. doi:10.1016/S0196-0644(87)80563-2

5. Reigstad O, Grimsgaard C, Thorkildsen R, Reigstad A, Røkkum M. Scaphoid non-unions, where do they come from? the epidemiology and initial presentation of 270 scaphoid non-unions. *Hand Surg*. 2012;17(3):331-335. doi:10.1142/S0218810412500268

6. Van Tassel DC, Owens BD, Wolf JM. Incidence estimates and demographics of scaphoid fracture in the U.S. population. *J Hand Surg Am*. 2010;35(8):1242-1245. doi:10.1016/j.jhsa.2010.05.017

7. Fusetti C, Garavaglia G, Papaloizos M, Wasserfallen J, Büchler U, Nagy L. Direct and indirect costs in the conservative management of undisplaced scaphoid fractures. *Eur J Orthop Surg Traumatology*. 2003;13(4):241-244. doi:10.1007/s00590-003-0101-6

8. Brydie A, Raby N. Early MRI in the management of clinical scaphoid fracture. *Br J Radiol*. 2003;76(905):296-300. doi:10.1259/bjr/19790905

9. Dorsay TA, Major NM, Helms CA. Cost-effectiveness of immediate MR imaging versus traditional follow-up for revealing radiographically occult scaphoid fractures. *AJR Am J Roentgenol*. 2001;177(6):1257-1263. doi:10.2214/ajr.177.6.17771257

10. Karl JW, Swart E, Strauch RJ. Diagnosis of occult scaphoid fractures: a cost-effectiveness analysis. *J Bone Joint Surg Am*. 2015;97(22):1860-1868. doi:10.2106/JBJS.O.00099

11. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist’s guide. *Radiology*. 2019;290(3):590-606. doi:10.1148/radiol.2018180547

12. Teale N, Teitel J, Morris MR, Sani N, Mitten D, Hammert WC. Convolutional neural network for second metacarpal radiographic osteoporosis screening. *J Hand Surg Am*. 2020;45(3):175-181. doi:10.1016/j.jhsa.2019.11.019

13. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410. doi:10.1001/jama.2016.17216

14. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. 2017;284(2):574-582. doi:10.1148/radiol.2017162326

15. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. doi:10.1038/nature21056

16. Olczak J, Fahlgberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop*. 2017;88(6):581-586. doi:10.1080/17453674.2017.1344459
17. Cai Z, Vasconcelos N. Cascade R-CNN: delving into high quality object detection. arXiv. Preprint published December 3, 2017. Accessed April 1, 2021. https://arxiv.org/abs/1712.00726

18. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis. 2015;115(3):211-252. doi:10.1007/s11263-015-0816-y

19. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90. doi:10.1145/3065386

20. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. arXiv. Preprint published December 11, 2015. Accessed April 1, 2021. https://arxiv.org/abs/1512.00567

21. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. Preprint published April 10, 2015. Accessed March 29, 2021. https://arxiv.org/abs/1409.1556

22. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. arXiv. Preprint published December 11, 2015. Accessed April 1, 2021. https://arxiv.org/abs/1512.00567

23. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. Preprint published April 10, 2015. Accessed March 29, 2021. https://arxiv.org/abs/1409.1556

24. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE; 2017:618-626. doi:10.1109/ICCV.2017.74

25. MedCalc. Diagnostic test evaluation calculator. Accessed July 31, 2020. https://www.medcalc.org/calc/diagnostic_test.php

26. Puza B, O’Neill T. Generalised Clopper–Pearson confidence intervals for the binomial proportion. J Stat Computation Simulation. 2006;76(6):489-508. doi:10.1080/10629360500107527

27. Wada K. Labelme: image polygonal annotation with Python. Accessed March 29, 2021. https://github.com/wkentaro/labelme

28. Mason D. Pydicom: an open source DICOM library. Accessed March 29, 2021. https://github.com/pydicom/pydicom

29. Grover R. Assessment of scaphoid injuries and the detection of fractures. J Hand Surg Br. 1996;21(3):341-343. doi:10.1016/S0266-7681(05)80197-4

30. Eyler Y, Sever M, Turkut A, et al. The evaluation of the sensitivity and specificity of wrist examination findings for predicting fractures. Am J Emerg Med. 2018;36(3):425-429. doi:10.1016/j.ajem.2017.08.050

31. Duckworth AD, Bujise GA, Moran M, et al. Predictors of fracture following suspected injury to the scaphoid. J Bone Joint Surg Br. 2012;94(7):961-968. doi:10.1302/0301-620X.94B7.28704

32. Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci USA. 2018;115(45):11591-11596. doi:10.1073/pnas.1806905115

33. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol. 2018;73(5):439-445. doi:10.1016/j.crad.2017.11.015

34. Choosing Wisely. Bone-density tests. Accessed March 29, 2021. https://www.choosingwisely.org/patient-resources/bone-density-tests/

35. Cramer CM, Nasraoui O, Wiklund P. CAD system for the detection of fractures in scaphoid with deep convolutional neural network. IEEE Trans Med Imaging. 2017;36(5):1156-1165. doi:10.1109/TMI.2016.2624122

36. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500-510. doi:10.1038/s41568-018-0016-5

37. Wong SC, Gatt A, Stamatescu V, McDonnell MD. Understanding data augmentation for classification: when to warp? In: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). IEEE; 2016:1-6. doi:10.1109/DICTA.2016.7797091

38. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol. 2018;73(5):439-445. doi:10.1016/j.crad.2017.11.015

39. Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H. Understanding neural networks through deep visualization. arXiv. Preprint published June 22, 2015. Accessed March 29, 2021. https://arxiv.org/abs/1506.06579

SUPPLEMENT.

eTable 1. Composition of Test Data Sets
eTable 2. Confusion Matrices
eFigure. Flow of Testing