European Journal of Radiology 130 (2020) 109188

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

The effect of deep convolutional neural networks on radiologists' performance in the detection of hip fractures on digital pelvic radiographs

Tsubasa Mawatari[b], Yoshiko Hayashida[a,*], Shigehiko Katsuragawa[b], Yuta Yoshimatsu[a], Toshihiko Hamamura[a], Kenta Anai[a], Midori Ueno[a], Satoru Yamaga[c], Issei Ueda[a], Takashi Terasawa[a], Akitaka Fujisaki[a], Chihiro Chihara[a], Tomoyuki Miyagi[c], Takatoshi Aoki[a], Yukunori Korogi[a]

[a] Department of Radiology, University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishiku, Kitakyushu, Fukuoka, 807-8555, Japan
[b] Department of Radiological Sciences, Graduate school of Health Sciences, Teikyo University, 6-22 Misakimachi, Omuta, Fukuoka, 836-8505, Japan
[c] Department of Medicine and Biosystemic Sciences, Graduate School of Medical Sciences, Kyushu University, Japan

ARTICLE INFO

Keywords:
Femoral Fractures
Radiography
Neural Networks, Computer
Diagnosis, Computer-Assisted
Reference Standards

ABSTRACT

Purpose: The purpose of our study is to develop deep convolutional neural network (DCNN) for detecting hip fractures using CT and MRI as a gold standard, and to evaluate the diagnostic performance of 7 readers with and without DCNN.

Methods: The study population consisted of 327 patients who underwent pelvic CT or MRI and were diagnosed with proximal femoral fractures. All radiographs were manually checked and annotated by radiologists referring to CT and MRI for selecting ROI. At first, a DCNN with the GoogLeNet model was trained by 302 cases. The remaining 25 cases and 25 control subjects were used for the observer performance study and for the testing of DCNN. Seven readers took part in this study. A continuous rating scale was used to record each observer's confidence level. Subsequently, each observer interpreted with the DCNN outputs and rated them again. The area under the curve (AUC) was used to compare the fracture detection.

Results: The average AUC of the 7 readers was 0.832. The AUC of DCNN alone was 0.905. The average AUC of the 7 readers with DCNN outputs was 0.876. The AUC of readers with DCNN output were higher than those without($p < 0.05$). The AUC of the 2 experienced readers with DCNN output exceeded the AUC of DCNN alone.

Conclusion: For detecting the hip fractures on radiographs, DCNN developed using CT and MRI as a gold standard by radiologists improved the diagnostic performance including the experienced readers.

1. Introduction

Hip fractures are one of the most common causes of hospitalization, morbidity, and mortality in the elderly [1], with a lifetime risk of 17.5% for women and 6% for men [2]. The majority of proximal femoral fractures are diagnosed by radiography. However, a minority of patients with suspected proximal femoral fracture and negative radiographs may have a radiographically "occult" fracture. Due to demineralization, overlying soft tissues, non-displacement, or technical factors, fractures may be occult on radiography. The rate of occult hip fractures has been commonly reported as between 3 and 10% of all negative hip/pelvic radiographs obtained from trauma [3]. In the case of "occult" fractures, the patients undergo further imaging, including additional computed tomography (CT), or magnetic resonance imaging (MRI). This increase, however, not only affects diagnostic costs, but also causes delays or missed diagnosis. This can result in undesired patient outcomes including increased mortality rate, length of hospitalization and cost of care [4–6].

Machine learning is an artificial intelligence that develops algorithms to enable computers to learn from existing data without explicit programming. Deep learning is a supervised machine learning method and some "ground truth" exists which is used to train the algorithms [7]. Recently, a machine deep learning technique known as deep convolutional neural networks (DCNNs) - a specific type of deep learning - has been applied to image recognition tasks. DCNNs are well-suited for imaging, therefore, DCNNs have received considerable attention as a

Abbreviations: DCNN, Deep Convolutional Neural Network; ROC, analysis: receiver operating characteristic analysis; AUC, area under ROC curve

* Corresponding author at: Department of Radiology, University of Occupational & Environmental Health, Iseigaoka 1-1, Yahatanishi-ku, Kitakyusyu-shi, 807-8555, Japan.
E-mail address: yhlinda@med.uoeh-u.ac.jp (Y. Hayashida).

https://doi.org/10.1016/j.ejrad.2020.109188
Received 26 April 2020; Received in revised form 12 July 2020; Accepted 18 July 2020
Available online 23 July 2020
0720-048X/

powerful classification tool for medical images.

A few early studies have assessed the fracture detection on radiographs using DCNN [8–10]. In these studies, the reference standard for the training and testing images was based on the assessment of human readers determining which were visible, only within a radiograph. These extraction methodologies could adversely influence the classification accuracy and "occult fracture" being assessed as a "non-fracture case".

The purpose of our study is to develop an automated deep learning system for detecting hip fractures using CT and MRI as a gold standard annotated by radiologists, and to evaluate the diagnostic performance inclusive of the experienced readers in detecting hip fractures on radiographs;

2. Materials and Methods

2.1. Image database

This study was approved by the appropriate ethics review board, which waived the requirement for informed consent.

The study population consisted of three-hundred and sixteen patients who were diagnosed with proximal femoral fractures by CT and MRI between April 2004 and April 2018. The age-matched 25 patients, diagnosed with no fractures at the same period, were recruited as control cases. Radiography and CT were performed in all 341 cases and MRI in 42 cases. One patient had fractures bilaterally and a further 11 patients underwent radiography twice to confirm the first radiograph assessment showing no hip fracture. In total, 328 hip fractures in 327 radiographs and 25 control subjects were used in this study. All radiographs were manually checked and annotated retrospectively by the three radiologists (Y.H, T.T, and M.U) referring to CT and MRI for ROI selection. At first, a DCNN with the GoogLeNet model was trained by 302 cases (Fig. 1). The remaining 25 cases and 25 control subjects were used in the observer performance study, additionally for DCNN testing. The distribution of cases included in training and testing is shown in Table 1. Visualization of the fracture was evaluated using a five-point scale by 2 radiologists (Y.H, and T.T) with mutual agreement referring to CT and MRI. A five-point scale was described as below;

- Score 1: Apparent fracture on radiograph
- Score 2: Possible fracture on radiograph
- Score 3: Obscure fracture can be detected by radiograph
- Score 4: Fracture diagnosed by CT images
- Score 5: Fracture diagnosed only by MRI

2.2. Selection of ROI

We selected ROIs on the radiographs as input images of DCNN as shown in Fig. 2. Square shape ROIs were manually placed bilaterally for training and unilaterally for testing. The matrix size of the ROI varied between 256 x 256 and 300 x 300, depending on the area of the hip joint region including the femoral head and trochanter. For the training database, 303 ROIs with fractures and 247 ROIs without fracture were selected from 302 radiographs. Fifty-four ROIs were eliminated due to the artificial joints or metal components. For testing, a total of 50 ROIs including 25 ROIs with fractures and 25 ROIs without fractures were selected from 50 testing database. All ROI images were resized using a bicubic interpolation method to the matrix size of 224 x224 which is required for our DCNN.

2.3. Data augmentation

We performed the data augmentation by image rotation with the allowance of plus and minus one degree, and horizontal flipping. Finally, the number of ROI images for the training increased from 550 to 3300. Fig. 3 showed receiver operating characteristic (ROC) curves of 4 different data augmentation strategies. The area under the ROC curve (AUC) without data augmentation was 0.875. The AUC of the combination of horizontal flipping and rotation was the highest (0.905). As a result, we performed the data augmentation by both image rotation plus and minus one degree and horizontal flipping for all training sets.

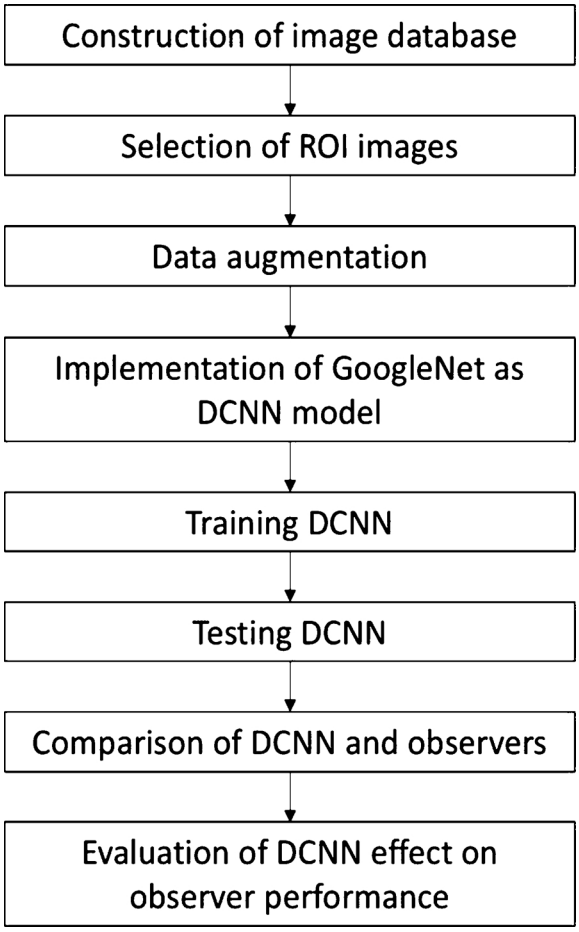

Fig. 1. shows the overall flowchart explaining the process for synthesizing, training, testing and evaluating the DCNN in this study. At first, two image databases for training and testing our DCNN were constructed after selecting the ROI of proximal femur. We employed GoogLeNet as a DCNN model. For training our DCNN model, we augmented ROI images of the training database by image rotation and horizontal flipping. After we trained our DCNN by using training datasets, the detection accuracy of DCNN for hip fractures was compared with observers using testing datasets. In addition, the effect of DCNN on observers' performance with DCNN output was also evaluated.

2.4. Transfer learning and testing with DCNN

We used GoogLeNet (Inception V1) [11] as a DCNN model on MATLAB 2018b (Mathworks Inc., Natick, MA). Conceptual images were shown in Fig. 4. An output layer and a fully connected layer of GoogLeNet were modified according to our task. We used 3300 ROI images in the training database and the entire GoogLeNet was re-trained. After building the models, we examined the accuracy of the trained models. Fifty ROI images were used to evaluate DCNN detection accuracy. DCNN returns the percentage of probability for each category by the bar graph as shown in Fig. 4. The categories for which the output value was higher was observed as the diagnosis made by the model.

Table 1
Summary information for the training and test image data sets.

Parameter	Training	Test data sets		P value	
			Control		vs control
	291	25	25		
No. of men	69	6	7	0.73	0.75
No. of women	222	19	18		
Pelvic radiography	302	25	25		
No. of men	73	6	7	0.78	0.75
No. of women	229	19	18		
Age (y)	81 ± 11	84 ± 9	82 ± 6	0.10	0.36
Men	77 ± 12	80 ± 8	82 ± 3	0.53	0.65
Women	82 ± 10	86 ± 9	82 ± 6	0.12	0.25
Hip joints					
Fractures	303	25	0		
No. of right	166	15	0	0.61	
No. of left	137	10	0		
Non fracture	247	0	25		
No. of right	101	0	9	0.63	
No. of left	146	0	16		
Fracture site					
femoral neck	151	18	-		
trochanteric	150	7	-		
subtrochanteric	2	0	-		
Score					
1	125	4	-	0.01	
2	61	6	-	0.64	
3	44	7	-	0.07	
4	59	4	-	0.67	
5	14	4	-	0.02	

Note. –Data are the mean ± standard deviation or numbers of images.
Score 1: Apparent fracture on radiograph.
Score 2: Possible fracture on radiograph.
Score 3: Obscure fracture can be detected by radiograph.
Score 4: Fracture diagnosed by CT images.
Score 5: Fracture diagnosed only by MRI.

2.5. Observer performance

An independent test method was used for evaluation. One radiologist (Y.H) lined up the 25 fracture cases and 25 healthy cases in random order. These 50 cases were also used for evaluation of DCNN detection accuracy in testing phase. The seven readers took part in the observer performance study with no prior knowledge of the CT or MRI results. Seven readers included 3 radiologists with 9, 13 and 24 years of experience respectively, an orthopedist with 22 years of experience, a radiology specialist trainee with 4 years of experience, a general physician with 4 years of experience, and one senior resident. Firstly, readers evaluated the radiographs alone without DCNN outputs. Each reader interpreted with the information of the affected side independently. A continuous rating with a 0-100 scale was used to represent each observer's confidence level. Subsequently, we conducted a second observer performance study with DCNN output. The time interval between the first and the second interpretation was more than one month.

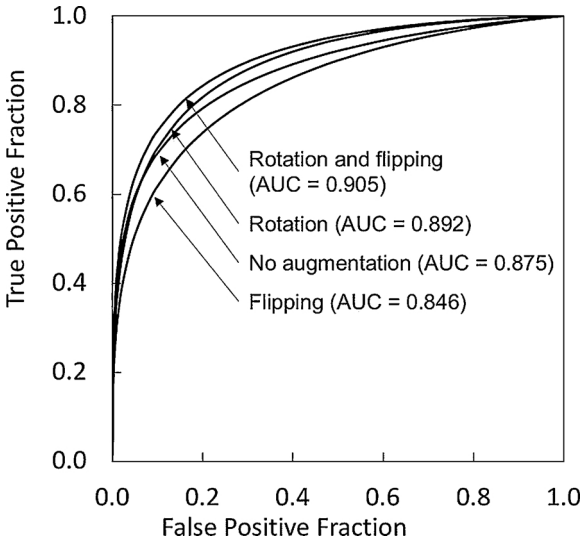

Fig. 3. Receiver operating characteristic (ROC) curves of 4 different data augmentation strategies in DCNN models: 1 horizontal flipping, 2 rotation with plus and minus one degree, 3 both rotation and flipping and, 4 no augmentation. The AUC of DCNN models with both rotation and flipping augmentation were the highest (AUC = 0.905). As a result, we performed the data augmentation by both image rotation with plus and minus one degree, and horizontal flipping for all training sets in this study.

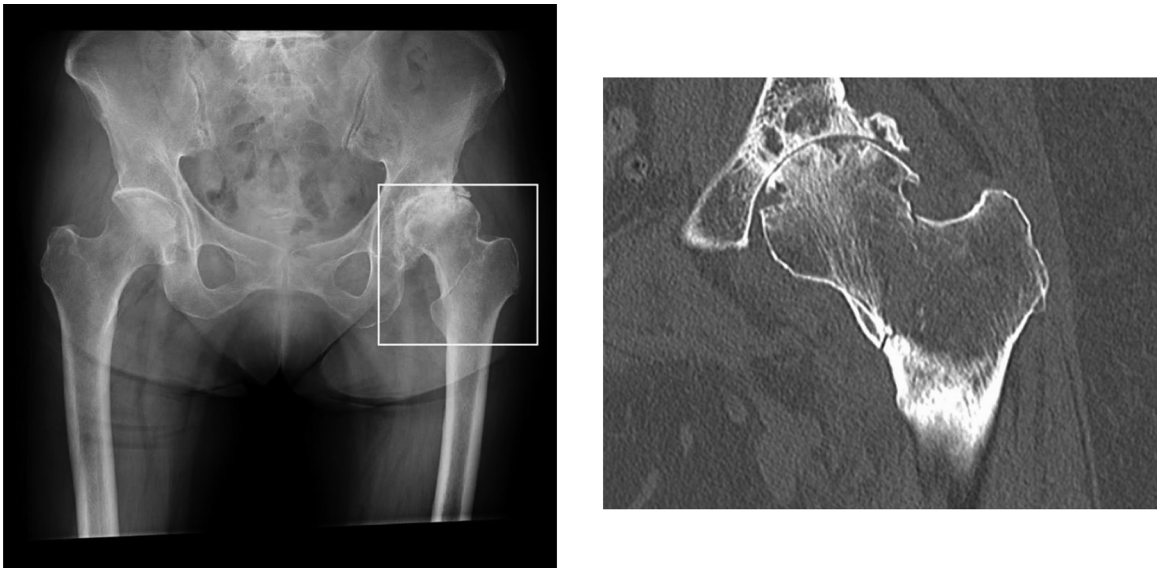

Fig. 2. a) Pelvic radiograph and b) Coronal multiplanar reconstruction (MPR) of the pelvic CT image in an 83-year-old with left femoral neck fracture.
For the input of DCNN, a square shape ROI was manually drawn at the proximal femur as shown on a). In this test case, 4 out of 7 readers and DCNN assessed this as a "fracture".

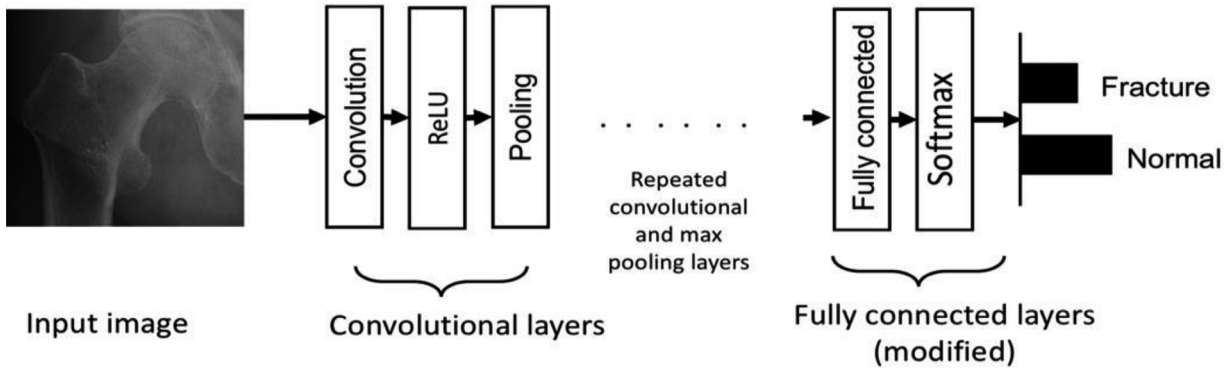

Fig. 4. Conceptual images of GoogLeNet used in this study. The square shape ROIs of the proximal femur were used as input images. Image features of input images were automatically extracted in convolution layers with the activation function of rectifier, ReLU and pooling layers. The extracted features were finally processed in fully connected layers. An output layer and a fully connected layer of GoogLeNet were modified in accordance with our task. Softmax function calculates the percentage of probability for each category (fracture and normal). The category with the higher output value is observed as the diagnosis made by the DCNN.

2.6. Data analysis

We evaluated the detection accuracy on hip fractures by means of ROC analysis. The statistical significance of the differences between average AUC values for readers with and without DCNN outputs was evaluated by Student's t-test. We also evaluated the DCNN detection accuracy itself by the ROC analysis. In addition, we evaluated the beneficial and detrimental effects of DCNN outputs following the methods of previous studies [12,13].

3. Results

The sensitivity of DCNN alone for the detection of hip fracture was 88% (score 1-3:100%, score 4: 75%, score 5: 50%) and the specificity of DCNN alone was 72%, whereas, the average sensitivity of readers for the detection of hip fracture was 78.3% (score 1-3: 94%, score 4: 53.6%, score 5: 37.5%) and the average specificity of readers was 72.6%.

Table 2 shows the readers' detection accuracy (AUC) of hip fractures with and without DCNN outputs. The AUCs of the 7 readers were 0.920,0.886, 0.842 0.839, 0.827, 0.810, and 0.698, respectively. The AUC of DCNN alone was 0.905. The AUCs of the 7 readers with DCNN outputs were 0.934,0.928, 0.896 0.866, 0.862, 0.841, and 0.800 respectively. The AUC of both experienced and less-experienced readers with DCNN output were higher than those without ($P < 0.05$). The AUC of the 2 experienced readers (reader No.2 and 3) with the DCNN output exceeded the AUC of DCNN alone.

The overall performance of the DCNN and the readers is illustrated by the average ROC curves in Fig. 5. The average AUC value of all readers had significantly increased from 0.832 to 0.876 after they referred the DCNN outputs ($P < 0.05$). However, the average AUC of the 7 readers with the DCNN outputs was lower than that of DCNN alone (AUC = 0.905)

Fig. 6 shows the beneficial and detrimental effects of the DCNN outputs for the individual readers. In all 50 test cases, the average number of cases affected positively and adversely by the DCNN output were 7.0 and 4.1, respectively ($P = 0.05$). In the 25 normal cases, the average number of cases affected positively and adversely were 5.4 and 2.1, respectively ($P = 0.07$). In the 25 fracture cases, the average number of cases affected positively and adversely were 1.6 and 2.0, respectively ($P = 0.34$). The effect of the DCNN output was significantly more beneficial than detrimental in all cases. When considering the average number of cases affected positively and adversely, there was no statistically significant difference between the normal cases and the cases with fractures.

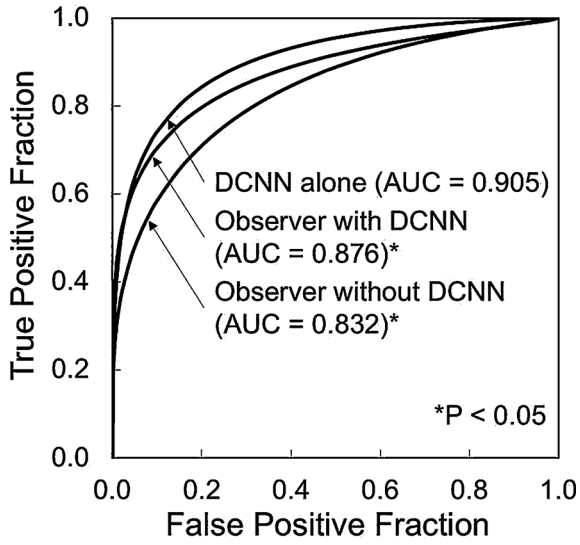

Fig. 5. The overall performance of the DCNN and the readers is illustrated by the average ROC curves in Fig. 5. The average AUC value of all the readers had significantly increased from 0.832 to 0.876 after they referred DCNN outputs ($P < 0.05$). However, the average AUC of the 7 readers with DCNN outputs was lower than that of DCNN alone (AUC = 0.905).

Table 2
The readers' detection accuracy of hip fractures with and without DCNN outputs.

Observer	No.1	No.2	No.3	No.4	No.5	No.6	No.7	Average	SD
Exp.years	24	22	13	9	4	4	2	11.1	8.254
AUC without DCNN	0.886	0.920*	0.839	0.842	0.827	0.810	0.698	0.832	0.065
AUC with DCNN	0.896	0.934*	0.928*	0.862	0.841	0.866	0.800	0.876	0.044

Note. –Data are AUC, SD or numbers of years.
AUC = the area under ROC curve, SD = standard deviation.
* Exceeded the AUC of DCNN alone (= 0.905).

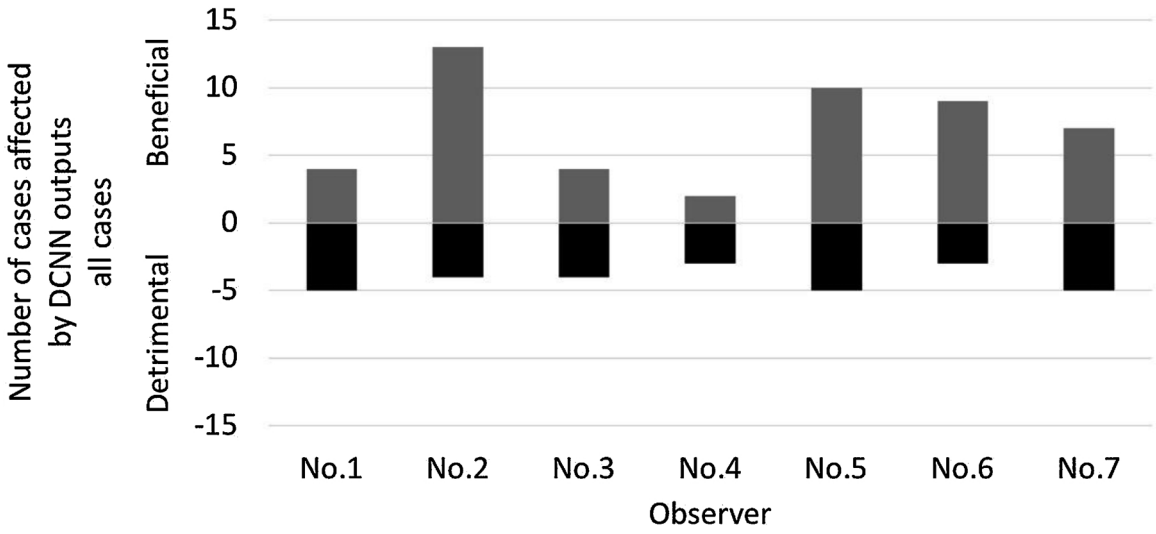

Fig. 6. The number of cases affected by DCNN outputs for the individual observers was shown on Fig. 6. The year of experience of each observer was described in Table 2. We assumed that a clinically relevant change in the confidence rating occurred when the difference was greater than 30 units on the confidence level. In all 50 test cases, the average number of cases affected positively and adversely by the DCNN output were 7.0 and 4.1, respectively (P = 0.05). The effect of the DCNN output was significantly more beneficial than detrimental in all cases.

4. Discussion

We developed a deep learning system for detecting hip fractures on radiographs using CT or MRI as a gold standard annotated by radiologists retrospectively. We also evaluated the diagnostic performance of readers with and without DCNN output. The DCNN output improved the diagnostic performance inclusive of the experienced readers.

Nowadays, a number of deep learning to radiological images of various modalities has been developed that can aid with detection [14–16], diagnosis [17–19] staging [20] and subclassification [21,22] of conditions.

Artificial intelligence and the automation of bony fracture detection on radiographs have been discussed [8–10,23–27]. Olczak et al [8] achieved an accuracy of 83% for fracture detection of the hand, wrist, and ankle radiographs. Yee Liang Thian et al. [23] reported that the ability of DCNNs not only in detection but also localization of radius and ulna fractures. A DCNN method for predicting hip fractures has been also described in a few reports [24–26]. Urakawa et al. [24] achieved an accuracy of 95.5% for fracture detection of the hip radiographs. Justin D.K et al. [27] reported that accuracy of their model was 93.7%, and Cheng et al. [26] achieved an AUC of 0.98 for identifying hip fractures. The validity of the visualization mapping of images for location of the hip fractures had also been investigated.

However, some previous studies used radiologists' reports as the reference standard [8,9], applying automated annotation by using language extraction software [8]. Other studies used the reference standard annotated by specialists retrospectively with radiograph [23] without using CT or MRI as a gold standard for all teaching and testing data [10,24,26,27]. We believe that such inclusion criteria cause the omission of false negative cases from the onset of the study. Several reports assess the human readers' performance comparing DCNN. However, the human readers' performance with or without DCNN output have not been assessed in most previous studies [8–10,24–26].

Our AUC of DCNN alone is 0.905, reaching the expert readers level (AUC of reader 2 is 0.920). Our performance for fracture detection was slightly lower than that of prior published work [9,10,23–27] on orthopedic radiographs. In the previous studies [9,10,23,27], the reference standard for training and testing images was based on the assessment of human readers determining what was visible by radiograph. In our study, all our training data was manually checked and annotated by radiologists referring to CT and MRI retrospectively. This shows we provided higher-level reference standards not only for readers but for DCNN and it may cause relatively low AUC values compared to previous studies. Although more time consuming, it provides more accurate data labeling. Our DCNN outputs could significantly improve readers' performance. Interestingly, the average AUC values of readers with DCNN outputs was lower than the AUC values of the DCNN alone. Furthermore, AUC values of experts with DCNN outputs were higher than that of the DCNN alone. This indicates that the readers could not fully agree with DCNN outputs. One might consider that our DCNN output shows only the probabilities of fracture without locating the fracture line. DCNN outputs are "black-box" for the readers. Therefore, it is challenging for them to depend on this without comprehending the quality of DCNNs. A valid diagnosis should not be made by radiograph alone. It also meant that, the model could never outperform the human. With our methodology, DCNN may have the potential to identify additional abstract features that have not been apparent to the human reader. In future studies considerably larger data sets pooled from multiple centers (the fracture case only detected by MRI not evident in the radiograph) are needed to confirm the potential of DCNN.

We used GoogLeNet (a.k.a. Inception V1) from Google,the winner of the ILSVRC 2014 competition[11] as a DCNN model. GoogLeNet had been pre-trained on ImageNet which included a large number of non-medical images [28]. Such an approach may significantly lower the development costs and increase practical use. There are some prior publications that GoogLeNet applied to radiographs, such as chest radiographs [29], mammography [30] and pelvic radiographs [31]. Adams, M. et al. [31] reported that the GoogLeNet DCNN outperfromed the AlexNet DCNN to detecting femoral neck fracture on radiograph. We believe that GoogLeNet is more efficient at extracting relevant features from the training data when compared with older CNN architectures such as AlexNet or Inceptions networks used in the previous reports [9,23].

There are several limitations to this study. Firstly, we only included proximal femoral fractures without evaluating all potential fractures on a pelvic radiograph, such as iliac, sacral or distal femoral fractures. Secondly, the number of the cases only detected by MRI not evident in the radiograph is small. Thirdly, our DCNN model returns only the probability of the hip fractures and not the location of the fractures. Fourthly, ROI images with the matrix size of only 224 × 224 is permitted as input of the DCNN, ROI images with free matrix sizes cannot

be used. The finally, in the selection of the 25 fracture cases for accuracy testing, we randomly selected with the aim of including higher score cases. It may cause the selection bias.

DCNN developed using the higher-level reference standards has the potential to increase the efficiency of diagnosis and expand access to "expert level" medical image interpretation. For the creation of a practical DCNN, radiologists must be involved in giving attention to the performance of training data, and this methodology provides more accurate data labeling. Future studies with larger numbers are needed to confirm the potential of DCNN.

Funding

The authors state that this work has not received any funding.

Guarantor

The scientific guarantor of this publication is Prof. Yukunori Korogi.

Declaration of Competing Interest

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and Biometry

One of the authors has significant statistical expertise.

Informed Consent

Written informed consent was waived by the Institutional Review Board.

Ethical Approval

Institutional Review Board approval was obtained.

Study subjects or cohorts overlap

N/A

Methodology

Methodology:

- Retrospective
- diagnostic study
- performed at one institution

CRediT authorship contribution statement

Tsubasa Mawatari: Writing - original draft, Data curation, Software. **Yoshiko Hayashida:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Shigehiko Katsuragawa:** Writing - original draft, Software, Methodology, Formal analysis. **Yuta Yoshimatsu:** Investigation. **Toshihiko Hamamura:** Investigation. **Kenta Anai:** Investigation. **Midori Ueno:** Investigation. **Satoru Yamaga:** Investigation. **Issei Ueda:** Software. **Takashi Terasawa:** Investigation, Validation. **Akitaka Fujisaki:** Investigation. **Chihiro Chihara:** Investigation. **Tomoyuki Miyagi:** Investigation. **Takatoshi Aoki:** Supervision, Validation. **Yukunori Korogi:** Supervision.

Acknowledgements

2.1 N/A

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2020.109188.

References

[1] O. Johnell, J.A. Kanis, An estimate of the worldwide prevalence and disability associated with osteoporotic fractures, Osteoporos Int 17 (12) (2006) 1726–1733.
[2] P. Kannus, J. Parkkari, H. Sievanen, A. Heinonen, I. Vuori, M. Jarvinen, Epidemiology of hip fractures, Bone 18 (1 Suppl) (1996) 57S–63S.
[3] P.F. Rizzo, E.S. Gould, J.P. Lyden, S.E. Asnis, Diagnosis of occult fractures about the hip. Magnetic resonance imaging compared with bone-scanning, J Bone Joint Surg Am 75 (3) (1993) 395–401.
[4] T. Shiga, Z. Wajima, Y. Ohe, Is operative delay associated with increased mortality of hip fracture patients? Systematic review, meta-analysis, and meta-regression, Can J Anaesth 55 (3) (2008) 146–154.
[5] N. Simunovic, P.J. Devereaux, M. Bhandari, Surgery for hip fractures: Does surgical delay affect outcomes? Indian J Orthop 45 (1) (2011) 27–32.
[6] S. Shabat, E. Heller, G. Mann, R. Gepstein, B. Fredman, M. Nyska, Economic consequences of operative delay for hip fractures in a non-profit institution, Orthopedics 26 (12) (2003) 1197–1199 discussion 1199.
[7] G. Zaharchuk, E. Gong, M. Wintermark, D. Rubin, C.P. Langlotz, Deep Learning in Neuroradiology, AJNR Am J Neuroradiol 39 (10) (2018) 1776–1784.
[8] J. Olczak, N. Fahlberg, A. Maki, A.S. Razavian, A. Jilert, A. Stark, O. Skoldenberg, M. Gordon, Artificial intelligence for analyzing orthopedic trauma radiographs, Acta Orthop 88 (6) (2017) 581–586.
[9] D.H. Kim, T. MacKinnon, Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks, Clin Radiol 73 (5) (2018) 439–445.
[10] S.W. Chung, S.S. Han, J.W. Lee, K.S. Oh, N.R. Kim, J.P. Yoon, J.Y. Kim, S.H. Moon, J. Kwon, H.J. Lee, Y.M. Noh, Y. Kim, Automated detection and classification of the proximal humerus fracture by using deep learning algorithm, Acta Orthop 89 (4) (2018) 468–473.
[11] L.W. Szegedy C, Y. Jia, P. Sermanet, P. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, Computer Vision and Pattern Recognition (CVPR) (2015) 1–9.
[12] Y. Matsuki, K. Nakamura, H. Watanabe, T. Aoki, H. Nakata, S. Katsuragawa, K. Doi, Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT: evaluation with receiver operating characteristic analysis, AJR Am J Roentgenol 178 (3) (2002) 657–663.
[13] S. Tsukuda, A. Heshiki, S. Katsuragawa, Q. Li, H. MacMahon, K. Doi, Detection of lung nodules on digital chest radiographs: potential usefulness of a new contralateral subtraction technique, Radiology 223 (1) (2002) 199–203.
[14] M.C. Taylor AG, J.A. Mongan, Automated detection of clinically-significant pneumothorax on frontal chest X-rays using deep convolutional neural networks, PLoS Med. 15 (11) (2018) e1002697.
[15] J.R. Zech, M.A. Badgeley, M. Liu, A.B. Costa, J.J. Titano, E.K. Oermann, Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study, PLoS Med 15 (11) (2018) e1002683.
[16] T. Nakao, S. Hanaoka, Y. Nomura, I. Sato, M. Nemoto, S. Miki, E. Maeda, T. Yoshikawa, N. Hayashi, O. Abe, Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography, J Magn Reson Imaging 47 (4) (2018) 948–953.
[17] P. Rajpurkar, J. Irvin, R.L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C.P. Langlotz, B.N. Patel, K.W. Yeom, K. Shpanskaya, F.G. Blankenberg, J. Seekins, T.J. Amrhein, D.A. Mong, S.S. Halabi, E.J. Zucker, A.Y. Ng, M.P. Lungren, Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists, PLoS Med 15 (11) (2018) e1002686.
[18] N. Bien, P. Rajpurkar, R.L. Ball, J. Irvin, A. Park, E. Jones, M. Bereket, B.N. Patel, K.W. Yeom, K. Shpanskaya, S. Halabi, E. Zucker, G. Fanton, D.F. Amanatullah, C.F. Beaulieu, G.M. Riley, R.J. Stewart, F.G. Blankenberg, D.B. Larson, R.H. Jones, C.P. Langlotz, A.Y. Ng, M.P. Lungren, Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet, PLoS Med 15 (11) (2018) e1002699.
[19] K. Yasaka, H. Akai, O. Abe, S. Kiryu, Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study, Radiology 286 (3) (2018) 887–896.
[20] K. Yasaka, H. Akai, A. Kunimatsu, O. Abe, S. Kiryu, Liver Fibrosis: Deep Convolutional Neural Network for Staging by Using Gadoxetic Acid-enhanced Hepatobiliary Phase MR Images, Radiology 287 (1) (2018) 146–155.
[21] P. Chang, J. Grinband, B.D. Weinberg, M. Bardis, M. Khy, G. Cadena, M.Y. Su, S. Cha, C.G. Filippi, D. Bota, P. Baldi, L.M. Poisson, R. Jain, D. Chow, Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas, AJNR Am J Neuroradiol 39 (7) (2018) 1201–1207.
[22] A. Hosny, C. Parmar, T.P. Coroller, P. Grossmann, R. Zeleznik, A. Kumar, J. Bussink, R.J. Gillies, R.H. Mak, H. Aerts, Deep learning for lung cancer prognostication: A

retrospective multi-cohort radiomics study, PLoS Med 15 (11) (2018) e1002711.
[23] Y.L. Thian, Y. Li, P. Jagmohan, D. Sia, V.E.Y. Chan, R.T. Tan, Convolutional Neural Networks for Automated Fracture Detection and Localization on Wrist Radiographs, Radiology: Artificial Intelligence 1 (1) (2019) e180001.
[24] T. Urakawa, Y. Tanaka, S. Goto, H. Matsuzawa, K. Watanabe, N. Endo, Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network, Skeletal Radiol 48 (2) (2019) 239–244.
[25] O.-R.L. Gale W, G. Carneiro, et al., Detecting hip fractures with radiologist-level performance using deep neural networks, arXiv:1711.06504 (2017).
[26] C.T. Cheng, T.Y. Ho, T.Y. Lee, C.C. Chang, C.C. Chou, C.C. Chen, I.F. Chung, C.H. Liao, Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs, Eur Radiol 29 (10) (2019) 5469–5477.
[27] J.D. Krogue, K.V. Cheng, K.M. Hwang, P. Toogood, E.G. Meinberg, E.J. Geiger, M. Zaid, K.C. McGill, R. Patel, J.H. Sohn, A. Wright, B.F. Darger, K.A. Padrez, E. Ozhinsky, S. Majumdar, V. Pedoia, Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning, Radiology: Artificial Intelligence 2 (2) (2020).
[28] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, IEEE Trans Med Imaging 35 (5) (2016) 1285–1298.
[29] P. Lakhani, B. Sundaram, Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks, Radiology 284 (2) (2017) 574–582.
[30] D. Lévy, Arzav Jain, Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks arXiv preprint arXiv:1612.00542. (2016).
[31] M. Adams, W. Chen, D. Holcdorf, M.W. McCusker, P.D. Howe, F. Gaillard, Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures, J Med Imaging Radiat Oncol 63 (1) (2019) 27–32.