APPLICATION OF ARTIFICIAL INTELLIGENCE TO ASSIST HIP FRACTURE DIAGNOSIS USING PLAIN RADIOGRAPHS

Phichai Udombuathong, Ruthasiri Srisawasdi *, Waravut Kesornsukhon**, San Ratanasanya***

***Department of Orthopedic Surgery, Ananthamahidol Hospital, Lopburi, Thailand**
****Department of Computer and Information Science, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand**

Abstract
Background: Most hip fractures occur among elderly people. They are usually treated in the emergency room where orthopedic surgeons may not be readily available. The problem of delayed diagnosis and treatment results increase risks of further complications and mortality rate. Thus, applying artificial intelligence (AI) can assist physicians having limited experience to rapidly and confidently diagnose hip fractures using radiographs.
Objective: This study aimed to validate AI programs to assist diagnosing of hip fractures on plain radiographs.
Methods: This study employed a retrospective diagnostic study design. From 1 January 2015 to 31 December 2019, compiled ortho pelvis, anterior-posterior (AP) films from the diagnosis of hip fractures at Ananthamahidol Hospital were performed. The performance of the AI program was compared with one orthopedic surgeon who reviewed the same images. The accuracy, sensitivity and specificity of the diagnosis of hip fractures between the orthopedic surgeon and AI program were analyzed.
Results: In total, 217 patients were enrolled in this study. Of these, 56 (28.5%) were male and 161 (74.2%) female. Areas of hip fractures were as follow: intertrochanteric (108, 49.8%), femoral neck (102, 47.0%), subtrochanteric (6, 2.7%) and femoral head (1, 0.5%). The orthopedic surgeon and AI program revealed an accuracy of 93.59% (95%CI 90.8-95.73) vs. 81.24% (95% CI 77.17-84.85), sensitivity of 90.30% (95% CI 85.60-93.90) vs. 89.40% (95%CI 84.50-93.20) and specificity of 97.10% (95%CI 93.60-98.90) vs. 72.5% (95%CI 65.90-78.50), respectively.
Conclusion: Our results showed that the AI model (VGG16) showed a sensitivity of 89.40% vs. 90.30% obtained from the orthopedic surgeon. Thus, improvement in the sensitivity and specificity of AI software is further required. In the future, AI models have the potential as useful tools for emergent screening and evaluation of patients with hip fractures using plain radiographs, especially in the Emergency Department where orthopedic surgeons may not be readily available.

Keywords: Artificial intelligence, Hip fracture

J Southeast Asian Med Res 2022: 6 (1):e0111
https://doi.org/10.55374/jseamed.v6i0.111

Correspondence to:
Udombuathong P, Department of Orthopedic Surgery, Ananthamahidol Hospital, Lopburi, Thailand
E-mail: phichai856@gmail.com

Received: 5 December 2021
Revised: 14 February 2022
Accepted: 24 February 2022

Introduction

The incidence of hip fractures is expected to increase due to rising elderly populations worldwide. In 2000, approximately 1.6 million hip fractures were reported [1] which is expected to increase to 4.5 to 6.3 million cases in 2050 according to the International Osteoporosis Foundation.[2] Hip fractures are a common problem impacting socioeconomic status. The cumulative mortality after one year of hip fractures among patients occurs 20 to 40%.[3, 6, 10] Thus, the mortality rate among patients with hip fractures greatly increases to 28.7% compared with that in the general population.[5] A systematic review of 229,851 patients with hip fractures from 36 countries from 2013 to 2017 revealed that mean overall one year mortality after hip fractures was 22.0%.[6] Risk factors that increased mortality rate included higher American Society of Anesthesiology (ASA) score. Odds ratio of 2.3 for every ASA point added, being male[4], age and type of hip fracture[7] were associated risk factors. Intertrochanteric fracture had a higher mortality rate than that of femoral neck fracture (17.40% vs. 9.83%). One year mortality after hip fractures among patients receiving nonoperative treatment was more than those receiving operative treatment.[8] Seung-Ju Kim, et al. also showed that one year mortality after hip fractures among patients receiving nonoperative treatment was 48.5% while among those receiving surgical treatments was 19.9%.[9]

The definition of delayed surgical treatment of hip fractures varied in several studies. The optimal time for surgical treatment is 48 hours [11-15] starting from admission to the time of operation. From 25 studies, 21% of patients undergoing surgery after 48 hours died within one year, while one year mortality of those undergoing operations within 48 hours was less than 20%. Factors of delayed surgical treatment included patients' medical conditions, i.e., using an anticoagulant drug, having unstable medical conditions etc. The problem of diagnosing hip fractures, especially that of occult hip fracture constituted a number of both false positive and false negative diagnoses found based on radiography findings alone. Additionally, the experience of physicians working in the Emergency Department proved very crucial. Poor sensitivity and specificity of radiographs of the proximal femur and pelvis among patients with pain or suspected trauma around these structures were recorded.[16] Dominquez et al. reported that 4.40% of patients who were suspicious of hip fractures at the Emergency Department received a subsequent diagnosis as having fractures. The incidence of occult hip fracture was 3 to 10%.[18, 19] Magnetic Resonance Imaging (MRI) and Computerized Tomography Scan (CT) are useful tools to help diagnose occult hip fractures. MRI, an investigation of choice to diagnose occult hip fractures, is highly accurate with 100% sensitivity and specificity. MRI can be helpful for patients to receive an operation within the optimal time.[20-22] CT scan [21-23] is also a second line of choice when results obtained from MRI were contraindicated or could not be obtained within 24 hours. However, MRI and CT scans are costly and require more time to prepare patients.

Artificial Intelligence (AI), a useful tool for assisting the diagnosis of hip fractures using plain radiographs can reduce delayed diagnosis resulting in delayed surgical treatment when orthopedic surgeons or radiologists may not be available at the Emergency Department. At present, in general hospitals, all radiographs are recorded to digital files that can be analyzed and interpreted using Picture Archiving Communication Systems (PACS). AI is a field of Computer Science that has a competence of helping diagnosis close to human performances especially deep learning and supervised learning for training AI. The input data provided specific results. After training the AI, it can predict results from all input datasets. AI can prove and continue to predict the results until obtaining the correct results. Urakawa et al.[24] detected fractures from radiographs using Convolutional Neural Networks (CNNs). Visual Geometry Group 16 layer (VGG16) is CNN-selected to detect intertrochanteric hip fractures compared with the performance to detect intertrochanteric hip fractures between VGG16 and orthopedic surgeons. The accuracy of VGG16 and orthopedic surgeons was 95.5% vs. 92.2%, sensitivity 93.3%

vs. 88.3% and specificity 97.4% vs. 96.8%. A similar study conducted by Cheng et al.[25] found that DCNN (Deep Convolutional Neural Network) achieved an accuracy of 91% sensitivity and 98% specificity. Recognition of hip fractures using DCCN could be performed with less than one hour of perceptual training.[26]

Most hip fractures occur among elderly people. They are usually brought to the Emergency Department where orthopedic surgeons may be unavailable at that time. To detect fractures using radiographs or occult hip fractures, the diagnostic problems of hip fractures could result from inexperienced physicians leading to delayed treatment, increasing both complications and mortality rate. Application of AI to assist young and inexperienced physicians to diagnose hip fractures using plain radiographs could help these physicians be confident with rapid diagnosis of hip fractures decreasing adverse events at the process of diagnosis.

Methods

Study population

The study was reviewed and approved by the Institutional Review Board of the Medical Department, Royal Thai Army (approval number S015h/63_EXP). A retrospective diagnostic study design was conducted from 1 January 2015 to 31 December 2019. Patients, undergoing both operative and nonoperative treatments, received a diagnosis of hip fracture and were admitted at Ananthamahidol Hospital.

The sample size was calculated from the prevalence of 17% osteoporosis and the sensitivity of AI to detect hip fractures conducted by Urakawa et al.[24] A total of 524 radiograph images of hip fractures and nonhip fractures were used in this study.

Definition

Hip fracture is a bone fracture from the edge of the femoral head to 5 cm below the lesser trochanter of the femur.

Criteria

The inclusion criteria included patients aged more than 50 years having a diagnosis of osteoporosis or hip fractures from accidents, i.e., slipping, falling or vehicle accidents. The exclusion criteria comprised patients having a diagnosis of bone tumor, periprosthetic hip fractures, nonfractured hip side without instruments, i.e., cement, screw, nail and prosthesis.

Compiled ortho pelvis and anterior-posterior (AP) films from the diagnosis of hip fractures were performed using 541 images which were divided to 120 images for AI training procedures and 421 images for test procedures. The performance of the AI program was compared with one orthopedic surgeon who reviewed the same images. The accuracy, sensitivity and specificity of the diagnosis of hip fractures between the orthopedic surgeon and AI program were analyzed.

Radiographs using deep learning techniques. the AI model comprised the VGG16, which is a public model and was used in this study. The VGG model was created in 2014 for which the model VGG16 has advantages of a design architecture with conv2D 3x3 pixels, 1 stride same padding and max pooling 2x2 pixels, 2 strides and hyperparameter. The VGG16 has 16 layers with a large network and hyperparameters encompassing about 138 million units.

Procedure

Preprocessing images

The 120 ortho pelvis radiograph images were provided for training procedures while 421 images were used for test procedures. Images were taken from digital film photographs and input to Picture Archiving Communication Systems (PACS) and all image data were recorded in the Digital Imaging and Communications in Medicine (DICOM). The DICOM was changed to JPG and adjusted by histogram equalization to brighten adjustment, reduce noise, and rotate and crop images as shown in **Figure 1.**

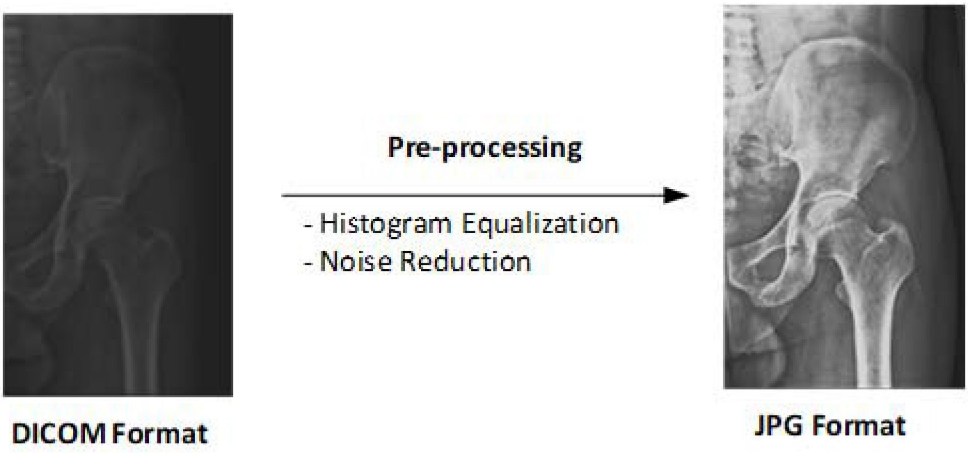

Figure 1. Preprocessing image

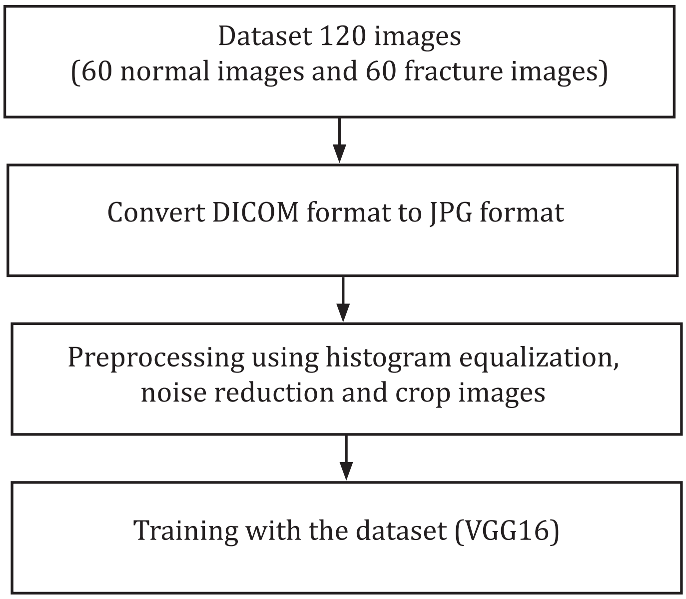

Figure 2. Training procedures

Training Procedures

Images for training using the VGG16 were classified in two groups: nonhip fracture and hip fracture. The size of each image was 200x200 pixels, then the images were input to the Image Data Generator by keras preprocessing that was used in the preprocessing process. The Image Data Generator adjusted the size, rotation and zoom of the images before training procedures. CPU Intel Core i5 Generation 11 and GPU NVIDIA RTX3060 were used for the training procedures involving VGG16 use 50 times as shown in **Figure 2**.

Test Procedures

The 421 Ortho Pelvis radiograph images were changed from DICOM to JPEG and adjusted using histogram equalization to reduce noise, rotate and crop images, then the preprocessing was repeated using the Image Data Generator I of VGG16. This process took less than 2 minutes to analyze the data of 421 images (**Figure 3**).

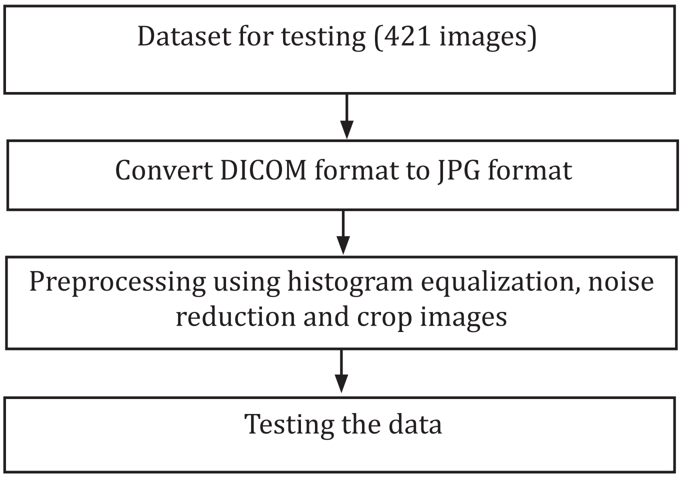

Figure 3. Training procedures

Table 1. Demographic data of enrolled patients

	number	%
Sex		
Male	56	25.8
Female	161	74.2
Age		
Mean± SD	76.89±10.6	
Median (min-max)	80 (50-93)	
Sideding		
Left	116	53.5
Right	101	46.5
Area		
Intertrochanteric fracture	108	49.8
Neck of femur fracture	102	47.0
Subtrochanteric fracture	6	2.7
Femeral head fracture	1	0.5

Table 2. Analysis of interpreted radiographics by AI

Gold Standard	Positive	Negative	Total
Abnormal	194	23	217
Normal	56	148	204
Total	250	171	421
		95%CI	
Sensitivity (%)	89.40	84.50	93.20
Specificity(%)	72.50	65.90	78.50

Table 2. (Continued)

Gold Standard	Positive	Negative	Total
Positive Predictive Value (%)	77.60	71.90	82.60
Negative Predictive Value (%)	86.50	80.50	91.30
Accuracy (%)	81.24	77.17	84.85

Table 3. Analysis of interpreted radiographics by the orthopedic surgeon

Gold standard	Positive	Negative	Total
Abnormal	196	21	217
Normal	6	198	204
Total	202	219	421
		95%CI	
Sensitivity (%)	90.30	85.60	93.90
Specificity(%)	97.10	93.70	98.90
Positive Predictive Value (%)	97.00	93.60	98.90
Negative Predictive Value (%)	90.40	85.70	94.00
Accuracy (%)	93.59	90.80	95.73

Results

In total, 217 patients were enrolled in this study. Of these, 56 (28.5%) were male and 161 (74.2%) female. Areas of hip fractures were as follow: intertrochanteric (108, 49.8%), femoral neck (102, 47.0%), subtrochanteric (6, 2.7%) and femoral head (1, 0.5%) as shown in **Table 1**. The orthopedic surgeon and AI program revealed accuracy of 93.59% (95%CI 90.8-95.73) vs. 81.24% (95% CI 77.17-84.85), sensitivity of 90.30% (95% CI 85.60-93.90) vs. 89.40% (95%CI 84.50-93.20), and specificity of 97.10% (95%CI 93.60-98.90) vs. 72.5% (95%CI 65.90-78.50), respectively (**Tables 2 and 3**).

Discussion

The AI model (VGG16) used in this study revealed an accuracy of 81.24% which was less than that reported by Takaaki et al. (95.5%).[24] However, the accuracy of the orthopedic surgeon was similar between the two studies (93.59% vs. 92.22%). The difference resulted from the quality of images such as shooting distances, brightness, clearance, noise and size of images. In this study, among 217 images of hip fractures, 21 (9.7%) were occult hip fractures of which the incidence of the occult hip fracture was 3 to10 %[18, 19]: this could confound the accuracy of AI. The accuracy of AI used in this study could not detect the occult hip fracture to the same degree as those of the orthopedic surgeon. Preprocessing of images is an important procedure resulting in the accuracy of AI. In addition, the number of images, which were used for training AI procedures could have affected the accuracy of AI. In this study, 120 images (60 images of hip and 60 images of nonhip fractures) were used for the step of training procedures. Other related studies [24, 25] could provide more images for this step than our study. Takaaki et al. [24] used 2,678 images for training procedures of AI images (1,408 images of hip and 1,270 images of nonhip fractures) while Cheng et al. [25] used 3,605 images for training procedures of AI images (1,975 images of hip and 1,630 images of nonhip fractures). Additionally, the area of hip fractures used in the study of Takaaki et al. was specific at the intertrochanteric area of the femur while those in the study of Cheng et al. was specific at the neck and intertrochanteric area of the femur. Compared with our study, the areas of hip fractures starting from the head to the subtrochanteric area of the femur were used covering larger areas than those of related studies. Different areas had varieties of bone architectures and artifacts that reduced the accuracy of AI. Thus, one advantage of our study was being able to detect overall areas of the hip fractures.

The competence of AI for detecting hip fractures using plain radiographs was reliable as a screening tool to diagnose hip fracture because the sensitivity did not significantly differ from the performance of the orthopedic surgeon, 89.40% (95%CI= 84.50-93.20) vs. 90.30% (95% CI =85.60-93.90), respectively. Further studies to develop the competence of AI are required using more images for the AI training step as well as improving qualities of images in the pre-processing image. In the future, multihospital assessment of hip fractures would also be required to validate the developed AI model.

Conclusion

Our results showed that the AI model (VGG16) showed a sensitivity of 89.40% vs. 90.30% obtained from the orthopedic surgeon. However, improvement in the accuracy of AI software, both sensitivity and specificity, is further required. In the future, AI models have the potential to be a useful tool for emergent screening and evaluation of patients with hip fractures using plain radiographs, especially in the Emergency Department where orthopedic surgeons may not be readily available.

References

1. Johnell O, Kanis JA. An estimate of worldwide prevalence and disability associated with osteoporotic fractures. Osteoporos Int 2006; 17: 1726-33.
2. Gullberg B, Johnell O, Kanis JA. Worldwide projections for hip fracture. Osteoporos Int 1977; 7: 407-13.
3. Guzon-lllecas O, Fernandez EP, Villarias NC, Donate FJQ, Pena M, Alonso-Blas C, et al. Mortality after osteoporotic hip fracture: incidence, trends and associated factors. J. Orthop Surg Res 2019; 14: 1-9.
4. CL Daugaard, HL Jorgensen, T Riis, JB Lauritzen, BR Duss, SVD Mark. Is mortality after hip fracture associated with surgical delay or admission during weekend and public holidays? Acta Orthop 2012; 83: 609-13.
5. Ricci G, Longaray MP, Goncalves RZ, Neto ASU, Manente M, Barbosa LBH. Evaluation of the mortality rate one year after hip fracture and factors relating to diminished survival among elderly people. Rev Bras Orthop 2012; 47: 304-09.
6. Downey C, Kelly M, Quinland JF. Changing trends in the mortality rate at 1-year post hip fracture-a systematic review. World J Orthop 2019; 10: 166-75.
7. Cui Z, Feng H, Meng X, Zhuang S, Liu Z, Ye K, et al. Age-specific 1-year mortality rate after hip fracture based on populations in mainland China between the year 2000 and 2018. Arch Osteoporos 2019; 14: 1-10.
8. Chow SKH, Qin JH, Wong RMY, Yuen WF, Ngai WK, Tang N, et al. One year mortality in displaced intracapsular hip fracture and associated risk: a report of Chinese-based fragility fracture registry. J Orthop Surg Res 2018; 13: 1-6
9. Seung-Ju Kim, Hyundai-Soo Park, Dong-Woo Lee. Outcome of nonoperative treatment for hip fractures in elderly patients: A systematic review of recent literature. J Orthop Surg 2020; 28: 1-8
10. Moran CG, Wenn RT, Taylor AM, Sikand M. Early mortality after hip fracture: is delay before surgery important? JBJS Am 2005; 87: 483-89
11. Shiga T, Wajima Z, Ohe Y. Is operative delay associated with increased mortality of hip fracture patients? Systemic review, meta-analysts and meta-regression. Can J Anesth 2008; 55: 146-54.
12. Klestil T, Roder C, Stotter C, Winkler B, Nehrer S, Lutz, M, et al. Impact of timing of surgery in elderly hip fracture patients: a systematic review and meta-analysis. Sci Rep 2018; 8: 1-15

13. Sobolev B, Guy P, Sheehan KJ, Kuramoto L, Sutherland JM, Levy AR, et al. Mortality effects of timing alternative for hip fracture surgery. CMAJ 2018; 190: 923-32.
14. Matisson L, Bojan A, Enocson A. Epidemiology, treatment and mortality of trochanteric and subtrochanteric hip fracture:data from the Swedish fracture register. BMC Muscoskel Disord 2018; 19: 1-8.
15. Sa-Ngasoonsong, Kulachote N, Sirisreetreerax N, Chanplakorn P, Laohajaroensombat S, Seeraprom J, et al. Outcome of early hip fracture surgery in elderly patients in Ramathipbodi Hospital. Rama Med J 2013;36: 3-12.
16. MW Kirby, C Spritzer. Radiographic detection of hip and pelvic fractures in the Emergency Department. AJR Am J Roentgenol 2010; 194: 1054-60.
17. Dominquez S, Lui P, Roberts C, Mendell M, Richman PB. Prevalence of traumatic hip and pelvic fractures in patients with suspected hip fracture and negative initial standard radiographs. Acad Emerg Med 2005; 12: 366-70.
18. Je S, Kim H, Ryu S, Cho S, Oh S, Kang T, et al. The consequence of delayed diagnosis of occult hip fracture. J Trauma Inj 2015; 28: 91-97.
19. Sodozai Z, Davies R, Warner J. The sensitivity of CT scan in diagnosis occult femoral neck fractures. Injury Int J 2016; 47: 2769-71.
20. Yun BJ, Hunink M, Prabhakar AM, Heng M, Lui SW, Qudsi R, et al. Diagnosis imaging stratigies for occult hip fractures. Acad Emerg Med 2016; 23: 1161-69.
21. Raval P, Mayne AIW, PM Yeap, Oliver TB, Jirawala A, Sripada S. Outcome of MRI detected occult neck of femur fracture: Do they represent a less severe injury with improved outcomes? Hip Pelvis 2019; 31: 18-22.
22. Foex BA, Russell A. CT versus MRI for occult hip fracture. Emerg Med J 2018; 35: 645-47.
23. Trenton T. Kellock, Bharti Khurana, Jacob C. Mandell. Diagnostic performance of CT for occult proximal femoral fractures. A systematic review and meta-analysis. AJR Am J Roentgenol 2019; 213: 1324-30.
24. Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fracture with orthopedist level accuracy using a deep convolutional neural network. Skeletal Radiol 2019; 48: 239-44.
25. CT Cheng, TY Ho, TY Lee, CC Chang, CC Chou, CC Chen, et al. Application of a deep learning algorithm for detecting and visualization of hip fracture on plain radiographs. Eur Radiol 2019; 29: 5469-77.
26. Adams M, Chen W, Holcdorf D, McCusker MW, Howe PDL, Gaillard F. Deep learning versus perceptual training for the detection of neck of femur fracture. J Med Imaging Radiat Oncol 2019; 63: 27-32.