골절 연구 첫 논문, 유서깊은 정형외과 저널

■ TRAUMA

Deep learning for automated hip fracture detection and classification

ACHIEVING SUPERIOR ACCURACY

Z. Zheng,
B. Y. Ryu,
S. E. Kim,
D. S. Song,
S. H. Kim,
J-W. Park,
D. H. Ro

From Seoul National University Hospital, Seoul, South Korea

Aims

The aim of this study was to develop and evaluate a deep learning-based model for classification of hip fractures to enhance diagnostic accuracy.

Methods

A retrospective study used 5,168 hip anteroposterior radiographs, with 4,493 radiographs from two institutes (internal dataset) for training and 675 radiographs from another institute for validation. A convolutional neural network (CNN)-based classification model was trained on four types of hip fractures (Displaced, Valgus-impacted, Stable, and Unstable), using DAMO-YOLO for data processing and augmentation. The model's accuracy, sensitivity, specificity, Intersection over Union (IoU), and Dice coefficient were evaluated. Orthopaedic surgeons' diagnoses served as the reference standard, with comparisons made before and after artificial intelligence assistance.

Results

The accuracy, sensitivity, specificity, IoU, and Dice coefficients of the model for the four fracture categories in the internal dataset were as follows: Displaced (1.0, 0.79, 1.0, 0.70, 0.82), Valgus-impacted (1.0, 0.80, 1.0, 0.70, 0.82), Stable (0.99, 0.95, 0.99, 0.83, 0.89), and Unstable (1.0, 0.98, 0.99, 0.86, 0.92), respectively. For the external validation dataset, the sensitivity and specificity were as follows: Displaced (0.83, 0.94), Valgus-impacted (0.89, 0.90), Stable (0.88, 0.95), and Unstable (0.85, 0.99), respectively. The overall means (Micro AVG and Macro AVG) for the external dataset were Micro AVG (0.83 (SD 0.05), 0.96 (SD 0.01)) and Macro AVG (0.69 (SD 0.02), 0.95 (SD 0.02)), respectively.

Conclusion

Compared to human diagnosis alone, our study demonstrates that the developed model significantly improves the accuracy of detecting and classifying hip fractures. Our model has shown great potential in assisting clinicians with the accurate diagnosis and classification of hip fractures.

Cite this article: ***Bone Joint J*** **2025;107-B(2):213–220.**

Introduction

Hip fractures are common among elderly patients, and pose challenges in terms of diagnosis and treatment.[1] Approximately 250,000 hip fractures occur annually in the USA in people aged 65 years and older.[2] The reported one-year mortality rates for these fractures fall within the range of 12% to 36%, with associated morbidity, mortality, and healthcare costs.[2] Accurate, timely diagnosis is crucial for providing appropriate medical intervention and improving patient outcomes.[3,4]

Traditionally, the diagnosis and classification of hip fractures has relied on clinical evaluation, radiographs, CT,[5] and expert interpretation. The need for precise and efficient healthcare may benefit from advanced diagnostic tools. Latterly, artificial intelligence (AI) has been applied in orthopaedics in various ways.[6,7] Deep learning, a subset of AI, has shown promising results in medical image analysis,[8] offering the potential to enhance accuracy and speed of fracture diagnosis and classification.[9,10]

To date, application of AI in hip fracture classification has primarily focused on distinguishing fracture types, such as femoral neck or trochanteric fractures, as well as on the classification

Correspondence should be sent to D. H. Ro; email: duhyunro@gmail.com

doi:10.1302/0301-620X.107B2. BJJ-2024-0791.R1 $2.00

Bone Joint J
2025;107-B(2):213–220.

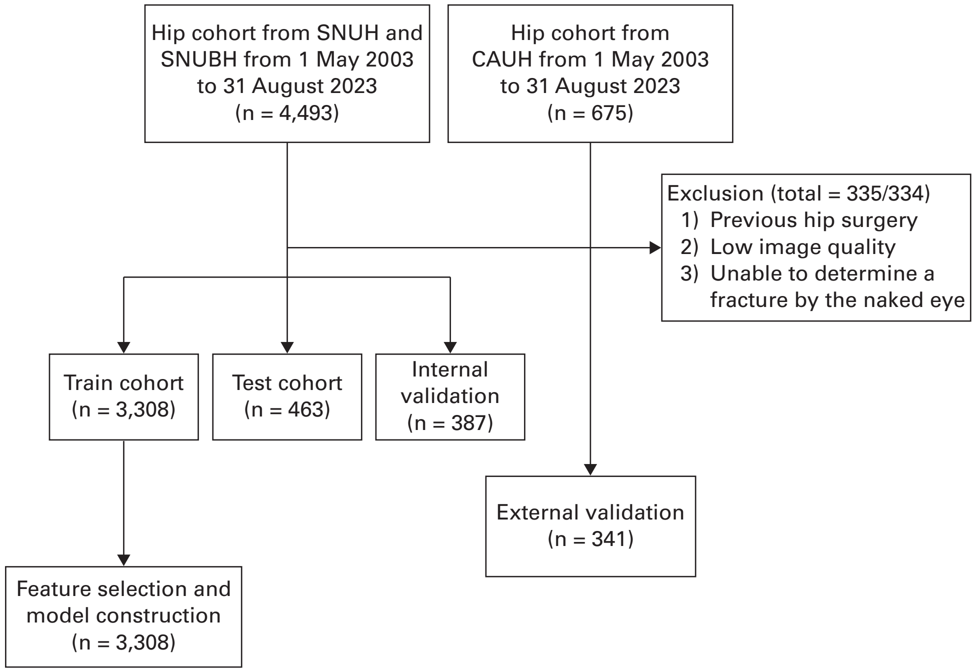

Fig. 1

Patient inclusion flowchart. CAUH, Chung-Ang University Hospital; SNUBH, Seoul National University Bundang Hospital.

Table I. Patient characteristics.

Characteristic	Internal dataset	External dataset
Patients, n	4,493	675
Mean age, yrs (SD)	80.83 (15.86)	77.46 (12.14)
Sex, n (%)		
Male	2,651 (59)	209 (31)
Female	1,842 (41)	466 (69)
Fracture type, n		
Valgus-impacted femur neck fracture	720	173
Displaced femur neck fracture	1,410	91
Stable intertrochanteric fracture	1,044	37
Unstable intertrochanteric fracture	984	40
No visible fracture	335	334

of commonly encountered femoral neck fractures (Garden type).[11–14] Although the potential for AI applications has been highlighted, their usefulness has been constrained by relatively limited datasets. Advancement requires larger and more diverse datasets. This would have particular value in critical settings like emergency departments.

Therefore, the main purpose of this study was to develop a deep learning model that can accurately identify and classify hip fractures, thereby enhancing clinicians' diagnostic accuracy.

Methods

Data collection and processing. This study was approved by the Institutional Review Board of Seoul National University Hospital (IRB no. 1810-004-974) and conducted in accordance with ethical standards. Written informed consent was obtained from all the study participants. The dataset for this study was obtained from three tertiary hospitals in South Korea: Seoul National University Hospital, Seoul National University Bundang Hospital, and Chung-Ang University Hospital. To train the model, a total of 4,493 hip isolated anteroposterior (AP) radiographs were retrospectively collected from 1 May 2003 to 31 August 2023, from two of these hospitals (Seoul National University Hospital and Seoul National University Bundang Hospital). An external validation dataset consisting of 675 radiographs was collected from Chung-Ang University Hospital. A total of 335 images were excluded due to previous hip surgery, poor quality, or a fracture not visible to the naked eye. These images were categorized as "No visible fracture" (Figure 1). Baseline characteristics are presented in Table I.

With reference to the Garden classification for femoral neck fractures[11,15] and the Evans classification[16,17] for intertrochanteric fractures, we classified the labelling into five types: "displaced femur neck fractures", "valgus-impacted femur neck fractures", "stable intertrochanteric fractures", "unstable intertrochanteric fractures", and "no visible fractures". Among them, the images labelled as "No visible fractures" represent that the hip fracture could not be determined as a hip fracture, or the fracture type was not within the scope of this study, and was excluded from the subsequent deep learning (Figure 2).

Data annotation. The images were first converted from Digital Imaging and Communications in Medicine (DICOM) format to Portable Network Graphics (PNG) format. The PNG images were labelled individually for subsequent training using the open-source labelling software LabelImg (Innodata, USA). Two orthopaedic surgeons (ZZ, JWP), each with over

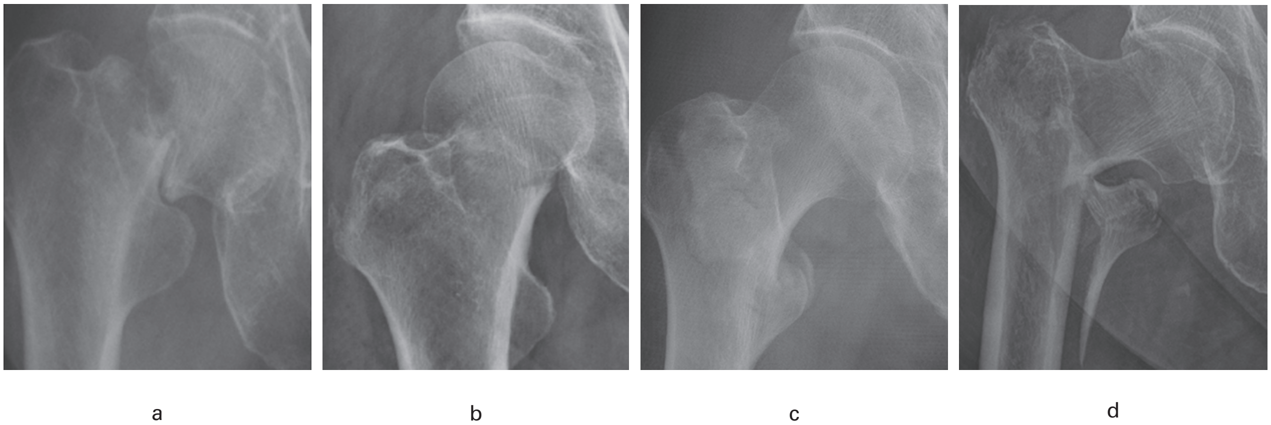

Fig. 2

Four types of fracture. a) Displaced femur neck fracture. b) Valgus-impacted femur neck fracture. c) Stable intertrochanteric fracture. d) Unstable intertrochanteric fracture.

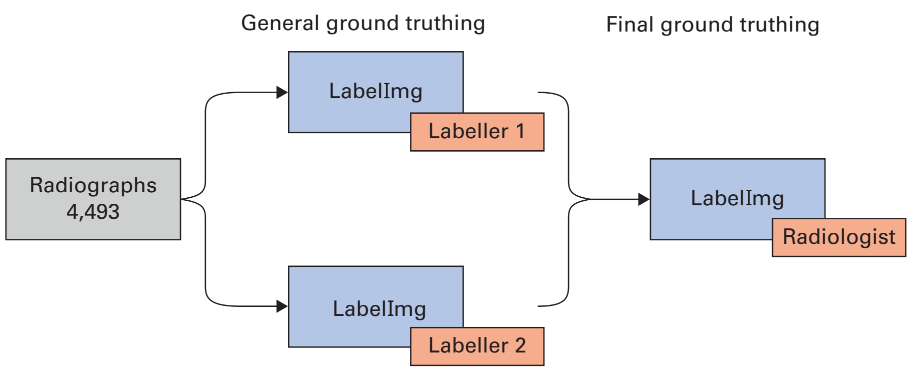

Fig. 3

The process of labelling and creating final ground truth.

five years of experience, along with one radiologist (see Acknowledgements), conducted the labelling process. Labeller 1 labelled the images based on radiographs, reading results, and surgical records, while Labeller 2 double-checked the labels. In case of discrepancies in fracture location or interpretation, both labellers performed a double-check through chart reviews and examining results, and the final confirmation was made by the radiologist. Of the 4,493 images, there were 253 instances where the two labellers disagreed. After a mutual discussion with the radiologist, these cases were categorized as either a specific fracture type or "No visible fracture" (Figure 3).

For the external validation set, a total of 341 images were assessed by three orthopaedic surgeons (A, B, and C). In Phase I, evaluations were conducted without AI assistance. Two weeks later, the Phase II assessments were conducted, again without AI assistance. After another two weeks, in Phase III, the assessments were conducted with AI assistance as described below (Figure 4). At this stage, the AI system performed an initial classification of each radiograph and provided diagnostic suggestions. The surgeons viewed the AI system's analysis results, including the predicted fracture types and possible feature markings (e.g. fracture location and shape). With the assistance of the AI system, the surgeons were able to quickly identify key radiological features and use their professional knowledge to verify and refine the AI's diagnostic results. The results from the three phases were compared to evaluate the AI's role in aiding human interpretation of hip fracture radiographs.

Deep learning model training. YOLO (You Only Look Once) is an AI model widely used in deep learning for object detection due to its real-time performance and multi-object detection capability (Alibaba DAMo Academy, China). It excels in dynamic and complex scenes with many targets. YOLO's efficient feature extraction ensures accurate recognition, and its end-to-end training simplifies model optimization.[18] In this study, DAMO-YOLO-T (Tiny) was employed, and the network was trained using input images resized to a uniform dimension of 1,024 × 1,024 pixels. We configured our training batch size to 64, which allowed for optimal use of our computational resources while maintaining a balance between training speed and memory usage. Training was conducted using four RTX 3,090 graphics processing units (GeForce RTX 3090; NVIDIA Corporation, USA), taking a total of 430 epochs and approximately 17 hours,

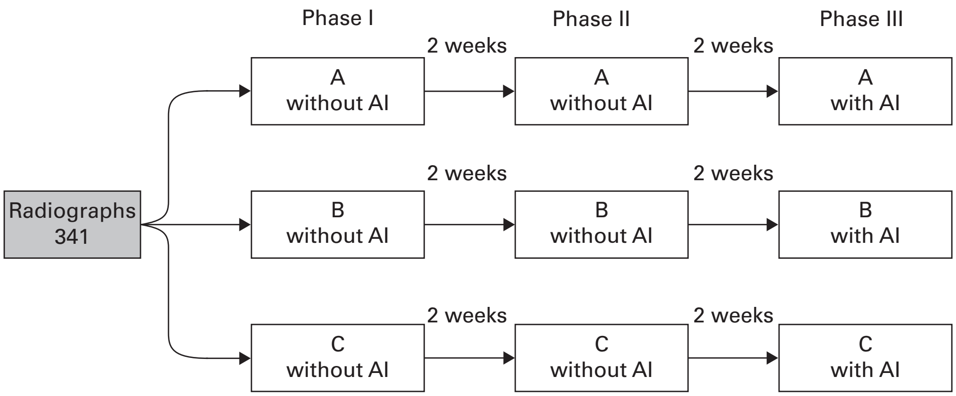

Fig. 4

The validation process of the external dataset. AI, artificial intelligence.

Table II. Performance metrics for hip fracture classification in internal datasets.

Fracture type	Accuracy	Sensitivity	Specificity	IoU	Dice
Displaced femoral neck fractures	1.0	0.79	1.0	0.70	0.82
Valgus-impacted femoral neck fractures	1.0	0.80	1.0	0.70	0.82
Stable intertrochanteric fractures	0.99	0.95	0.99	0.83	0.89
Unstable intertrochanteric fractures	1.0	0.98	0.99	0.86	0.92

IoU, Intersection over Union.

ensuring sufficient exposure to the training data for robust learning. We employed a momentum of 0.9 to accelerate convergence in the right direction and mitigate any oscillations during updates. Weight decay was set at 5e-4 to regularize the network and prevent overfitting by penalizing larger weights. A critical component of our training regimen was the implementation of a warmup period spanning 5 epochs, where the learning rate was gradually increased from a lower starting point to 0.05. This strategy helps to stabilize the training process in the initial phases, preventing any drastic updates that could derail the learning trajectory. Furthermore, we adjusted the base learning rate per image to 0.00015 (calculated as 0.01 divided by the batch size of 64). This per-image learning rate scaling ensures that the overall learning rate remains consistent regardless of batch size variations, promoting stable and consistent model training across different settings.

A deep neural network was designed and tuned based on an 80% training set, a 10% test set, and a 10% validation set. Ground truth labels of hip fracture classification were applied as follows: 1) 720 valgus-impacted femur neck fractures, 2) 1,410 displaced femur neck fractures, 3) 1,044 stable intertrochanteric fractures, 4) 984 unstable intertrochanteric fractures, and 5) 335 no visible fractures.

Evaluation metrics. Accuracy, sensitivity, specificity, Intersection over Union (IoU), and Dice coefficient were used to evaluate the model's performance for internal and external test datasets. Accuracy was defined as (TP (true positive) + TN (true negative))/(TP + TN + FP (false positive) + FN (false negative)). Sensitivity was defined as TP/(TP + FN) and specificity was defined as TN/(TN + FP). IoU was defined as (area of overlap intersection)/(area of union). It was determined by overlap intersection and union between two boxes: ground truth and prediction. Dice and IoU are highly effective in evaluating the overlap between predicted and actual segmented areas, especially in the highly imbalanced datasets common in medical imaging. Dice and IoU values greater than 0.7 and close to 1 indicate high precision and accuracy of the segmentation model.[19,20]

Results

Internal validation and testing. For the internal dataset, the diagnostic performance of the model is shown in Table II. The model exhibited a high level of accuracy for all fracture types, with accuracy values reaching 1.0 for displaced, valgus-impacted, and unstable intertrochanteric fractures, and 0.99 for stable intertrochanteric fractures. Sensitivity values ranged from 0.79 to 0.98, indicating the model's ability correctly to identify true positives. Specificity for displaced and valgus-impacted femoral neck fractures was 1.0, while the other two types were 0.99. The IoU and Dice coefficients, which measure the overlap and similarity between the predicted and actual fracture regions, also showed robust performance, with IoU values ranging from 0.70 to 0.86 and Dice coefficients from 0.82 to 0.92.

External validation. For the external dataset, the diagnostic performance of the model is shown in Table III. The mean

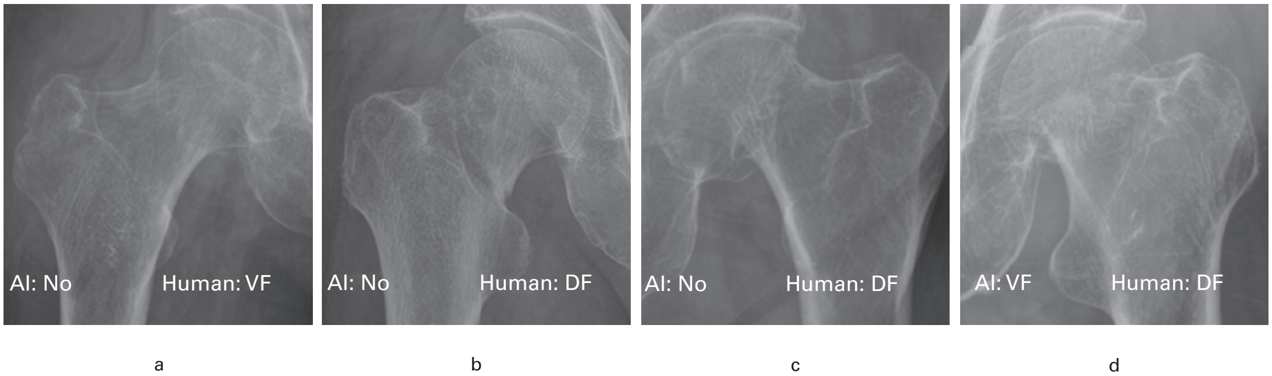

Fig. 5

The representative examples of differing diagnoses by trained artificial intelligence (AI) and humans. DF, displaced femur neck fractures; No, no identified fractures; VF, valgus-impacted femur neck fractures.

Table III. Performance metrics for hip fracture classification in external datasets.

Fracture type	A-Se	A-Sp	B-Se	B-Sp	C-Se	C-Sp	AVG-Se	AVG-Sp
Displaced femoral neck fractures	0.88	0.90	0.75	0.95	0.86	0.96	0.83	0.94
Valgus-impacted femoral neck fractures	0.80	0.95	0.93	0.85	0.94	0.91	0.89	0.90
Stable intertrochanteric fractures	0.84	0.94	0.89	0.94	0.92	0.96	0.88	0.95
Unstable intertrochanteric fractures	0.86	0.98	0.80	0.99	0.89	0.99	0.85	0.99
Micro AVG	0.83	0.96	0.78	0.95	0.87	0.97	0.83	0.96
Macro AVG	0.68	0.95	0.68	0.94	0.72	0.97	0.69	0.95

A, B, and C represent three different doctors.
AVG, average; Se, sensitivity; Sp, specificity.

Table IV. Studies evaluating the use of deep learning in hip fracture detection and/or classification.

Model	Classification	Algorithm	Internal database total/ train/test/valid	External database	Accuracy	Specificity	Sensitivity
Our model	VF/DF/SF/UF	DAMO-YOLO	AP 4493/3566/494/433	AP 675	1.00 VF	1.00 VF	0.80 VF
					1.00 DF	1.00 DF	0.79 DF
					0.99 SF	0.99 SF	0.95 SF
					1.00 UF	0.99 UF	0.98 UF
Mutasa et al[13]	Garden I/II/III/IV/NF	GAN	AP 1063/300/105/-	-	0.92 NF	0.93 NF	0.91 NF
					0.80 G I/II	0.93 G I/II	0.54 G I/II
					0.86 G III/IV	0.83 G III/IV	0.91 G III/IV
Yamada et al[12]	FNF/INT/NF	Xception	AP 1703/1553/150/ Lateral 1220/1070/150/	-	0.98	-	-

AP, anteroposterior; DF, displaced femur neck fractures; FNF, femoral neck fractures; G, Garden; GAN, generative adversarial network; INT, intertrochanteric fractures; NF, no fractures; SF, stable intertrochanteric fractures; UF, unstable intertrochanteric fractures; VF, valgus-impacted femur neck fractures.

Micro AVG sensitivity across different fracture types for the three clinicians is 0.83 (SD 0.05) and the specificity is 0.96 (SD 0.01), while the Macro AVG sensitivity is 0.69 (SD 0.02) and specificity is 0.95 (SD 0.02). This means that although the Macro AVG sensitivity is slightly lower, indicating potential challenges in classifying certain specific types of fractures, the general performance of the AI system on all samples in the external validation set is excellent.

Diagnostic performance of the model compared with orthopaedic surgeons. Figure 5 illustrates typical examples of differing diagnoses by AI (trained model) and humans. In the first three images (a, b, and c), each clinician has their corresponding diagnosis, while the AI failed to identify them. However, in the last image (d), there was a discrepancy between the clinicians and the AI in the diagnosis of valgus-impacted femoral neck fractures and displaced femoral neck fractures.

Discussion

In this study, we successfully developed a model that accurately detects and screens for a variety of hip fracture types using datasets based on isolated AP radiographs from three hospitals in South Korea. Both the results of the internal validation set (Table II) and the external validation set (Table III) demonstrate the excellent performance of the developed model, indicating good prospects for clinical application. Figure 4 highlights several typical examples of differing diagnoses by AI and clinicians. From the perspective of AP radiographs, due to the femoral external rotation or overlapping appearance of the fracture site, AI could not make a correct assessment, as in cases B and C. AI may miss some obvious fractures. Clinicians typically assess cortical continuity.[21] It seems occasionally that the cortex included in the box may be insufficient, with misalignment of the proximal and distal cortices, or the failure to connect the positions of the proximal and distal fragments. Therefore, we have considered widening the boundaries of the boxes during labelling, providing more information about the proximal and distal fragments. This may improve accuracy. However, this change would risk including unnecessary regions for learning.

Valgus-impacted femoral neck fractures describe a fracture where the proximal bone impacts the distal bone in a valgus direction, so there is inherently only slight displacement.[22] Generally these fractures are relatively stable, classified as Garden type I, which is an undisplaced, incomplete fracture.[23,24] However, even this fracture can ultimately become unstable.[25] From a clinical perspective, it should be considered a displaced femur neck fracture. Like case D in Figure 5, although AI may detect a valgus-impacted femoral neck fracture, in our interpretation it belongs to a displaced femoral neck fracture. In clinical practice, the definition between valgus-impacted femur neck fractures and displaced femur neck fractures is highly ambiguous, posing a challenge even for professionals. This is why the sensitivity of our model performance to valgus-impacted femoral neck fractures and displaced femoral neck fractures is relatively low. Indeed, Table IV shows that Mutasa et al's[13] study also reported low sensitivity in Garden type I and II fractures, indicating that they encountered similar issues.

Intracapsular fractures and extracapsular fractures occur with similar frequency but have different clinical features, treatment methods, and prognoses.[26,27] Despite grey areas, such as in basicervical fractures where both treatment methods could be used, intracapsular fractures are typically treated with arthroplasty or multiple cannulated screws, while intertrochanteric fractures are mostly treated with cephalomedullary nailing or sliding hip screws.[28,29] The specific treatment strategies will depend on many factors including age, bone mineral density, underlying comorbidities, but most importantly fracture classification.[30–32] For Garden type I or II femoral neck fractures, the subsequent risk of post-traumatic osteonecrosis and nonunion is relatively small, making osteosynthesis a more favourable surgical option.[33] In contrast, for Garden type III or IV displaced femoral neck fractures, with a high risk of osteonecrosis and nonunion, arthroplasty is favoured over osteosynthesis. For intertrochanteric fractures, the Evans classification for the evaluation of the fracture stability is crucial.[16] Currently, the use of intramedullary nails, mostly cephalomedullary nails, has increased, substituting the previously widely used sliding hip screws,[34] especially when fracture stability is a concern.[32] The distinctive radiological features of each fracture type introduced above were classified in the current study using the newly developed model. For more accurate diagnosis, advanced imaging modalities including CT or MRI scans are often used in the clinical setting. However, these are not always readily available. With the rigorous algorithm developed on the large set of accurately classified radiographs in the current study, the correct diagnosis and classification of hip fractures could be made from plain radiographs.

During our labelling process, we considered whether to classify fractures as valgus-impacted femoral neck fractures for internal fixation surgery or as displaced femoral neck fractures for arthroplasty. However, due to the lack of clear quantitative criteria, overcoming this purely by adding cases would be unhelpful. Clinically, internal fixation is more urgent to preclude possible late displacement, as delaying arthroplasty would not significantly alter the surgical plan.[35] Accordingly, there is reason to believe that valgus-impacted femoral neck fractures should be identified more sensitively. In other words, if the distinction between valgus-impacted and displaced femoral neck fractures is highly ambiguous in the AI's predictive results, it may be more helpful clinically to prioritize diagnosing valgus-impacted femur neck fractures.

Implementation of deep learning techniques for the detailed classification of hip fractures may facilitate personalized treatment plans, aiming for early mobilization to reduce postoperative complications and improve long-term mortality rates.[36] Timely and accurate diagnosis is crucial, as delayed treatment increases morbidity and mortality.[37] For inexperienced personnel, identifying the four types of fractures in this study can be difficult. Our research model serves as an initial diagnostic tool to support clinicians.

This study has some limitations. First, it only utilized AP radiographs for training. Future research should include lateral radiographs to avoid missing certain fractures. Second, fractures occurring in relation to abnormal morphology are rare, potentially leading to model overfitting if used for training. In the future, they could be categorized separately with a specific model being developed. Third, for valgus-impacted femoral neck fractures and displaced femoral neck fractures, ambiguous label definitions may hinder AI learning and reduce robustness. Therefore, establishing clear and refined standards, such as prioritizing valgus-impacted femoral neck fractures when their confidence value exceeds a certain threshold, is necessary for future algorithm implementation.

In conclusion, we developed a model using data from three institutions that accurately detected various hip fracture types and accurately localized the region. The currently developed model is highly consistent. Our findings support the integration of deep learning systems into clinical settings to assist diagnosing hip fractures.

Take home message

- This deep learning model, based on clinical surgical treatment methods, significantly improves the accuracy of diagnosing and classifying hip fractures.
- In clinical practice, precise fracture classification enables personalized treatment plans aimed at early mobilization, reducing postoperative complications and improving long-term mortality rates.

References

1. **Ansari H**, **Jaglal S**, **Cheung AM**, **Jain R**, **Weldon J**, **Kurdyak P**. Osteoporosis management in adults with schizophrenia following index hip fracture event: a 10-year population-based retrospective cohort study, Ontario, Canada. *Osteoporos Int*. 2024;35(7):1289–1298.
2. **Richmond J**, **Aharonoff GB**, **Zuckerman JD**, **Koval KJ**. Mortality risk after hip fracture. 2003. *J Orthop Trauma*. 2003;17(8 Suppl):S2–5.

3. **Sato Y**, **Takegami Y**, **Asamoto T**, **et al**. Artificial intelligence improves the accuracy of residents in the diagnosis of hip fractures: a multicenter study. *BMC Musculoskelet Disord*. 2021;22(1):407.
4. **Whitehouse MR**, **Berstock JR**, **Kelly MB**, **et al**. Higher 30-day mortality associated with the use of intramedullary nails compared with sliding hip screws for the treatment of trochanteric hip fractures: a prospective national registry study. *Bone Joint J*. 2019;101-B(1):83–91.
5. **Woldeyesus TA**, **Gjertsen J-E**, **Dalen I**, **et al**. Preoperative CT improves the assessment of stability in trochanteric hip fractures. *Bone Jt Open*. 2024;5(6): 524–531.
6. **Kim SE**, **Lee JH**, **Choi BS**, **Han H-S**, **Lee MC**, **Ro DH**. Performance of ChatGPT on solving orthopedic board-style questions: a comparative analysis of ChatGPT 3.5 and ChatGPT 4. *Clin Orthop Surg*. 2024;16(4):669–673.
7. **Lee DW**, **Han H-S**, **Lee MC**, **Ro DH**. Prediction of postoperative gait speed change after bilateral primary total knee arthroplasty in female patients using a machine learning algorithm. *Orthop Traumatol Surg Res*. 2024;103842.
8. **Kim SE**, **Nam JW**, **Kim JI**, **Kim J-K**, **Ro DH**. Enhanced deep learning model enables accurate alignment measurement across diverse institutional imaging protocols. *Knee Surg Relat Res*. 2024;36(1):4.
9. **Cha Y**, **Kim J-T**, **Park C-H**, **Kim J-W**, **Lee SY**, **Yoo J-I**. Artificial intelligence and machine learning on diagnosis and classification of hip fracture: systematic review. *J Orthop Surg Res*. 2022;17(1):520.
10. **Langerhuizen DWG**, **Janssen SJ**, **Mallee WH**, **et al**. What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? A systematic review. *Clin Orthop Relat Res*. 2019;477(11):2482–2491.
11. **Garden RS**. Low-angle fixation in fractures of the femoral neck. *J Bone Joint Surg Br*. 1961;43-B(4):647–663.
12. **Yamada Y**, **Maki S**, **Kishida S**, **et al**. Automated classification of hip fractures using deep convolutional neural networks with orthopedic surgeon-level accuracy: ensemble decision-making with antero-posterior and lateral radiographs. *Acta Orthop*. 2020;91(6):699–704.
13. **Mutasa S**, **Varada S**, **Goel A**, **Wong TT**, **Rasiej MJ**. Advanced deep learning techniques applied to automated femoral neck fracture detection and classification. *J Digit Imaging*. 2020;33(5):1209–1217.
14. **Murphy EA**, **Ehrhardt B**, **Gregson CL**, **et al**. Machine learning outperforms clinical experts in classification of hip fractures. *Sci Rep*. 2022;12(1):2058.
15. **Collin PG**, **D'Antoni AV**, **Loukas M**, **Oskouian RJ**, **Tubbs RS**. Hip fractures in the elderly-: a clinical anatomy review. *Clin Anat*. 2017;30(1):89–97.
16. **Evans EM**. The treatment of trochanteric fractures of the femur. *J Bone Joint Surg Br*. 1949;31-B(2):190–203.
17. **Andersen E**, **Jørgensen LG**, **Hededam LT**. Evans' classification of trochanteric fractures: an assessment of the interobserver and intraobserver reliability. *Injury*. 1990;21(6):377–378.
18. **Redmon J**, **Divvala S**, **Girshick R**, **Farhadi A**. You only look once: unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016, Las Vegas, Nevada, USA
19. **Müller D**, **Soto-Rey I**, **Kramer F**. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes*. 2022;15(1):210.
20. **Nobel SMN**, **Hossain MA**, **Kabir MM**, **Mridha MF**, **Alfarhood S**, **Safran M**. SegX-Net: a novel image segmentation approach for contrail detection using deep learning. *PLoS One*. 2024;19(3):e0298160.
21. **Lu Y**, **Uppal HS**. Hip fractures: relevant anatomy, classification, and biomechanics of fracture and fixation. *Geriatr Orthop Surg Rehabil*. 2019;10:2151459319859139.
22. **Smrke D**, **Makovec G**, **Weis M**, **Biscević M**. Impacted femoral neck fracture, surgery or not? *Med Arh*. 2006;60(6 Suppl 1):5–8.
23. **Wang Y**, **Ma J**, **Bai H**, **et al**. A three-dimensional measurement study of fracture displacement in Garden I femoral neck fracture: a retrospective study. *BMC Musculoskelet Disord*. 2023;24(1):623.
24. **Cai Z**, **Zhang Z**, **Ren L**, **Piao C**, **Xiang L**. Does Garden type I incomplete femoral neck fracture really exist in older adults? To evaluate the stability and consistency of Garden classification. *BMC Surg*. 2022;22(1):276.
25. **Zhang H**, **Xu X**, **Wu S**, **Liu Y**, **Mei J**. A finite element analysis study based on valgus impacted femoral neck fracture under diverse stances. *Comput Methods Biomech Biomed Engin*. 2022;25(5):475–486.
26. **Bhandari M**, **Swiontkowski M**. Management of acute hip fracture. *N Engl J Med*. 2017;377(21):2053–2062.
27. **Arshi A**, **Su L**, **Lee C**, **Sassoon AA**, **Zeegen EN**, **Stavrakis AI**. Comparison of complication profiles for femoral neck, intertrochanteric, and subtrochanteric geriatric hip fractures. *Arch Orthop Trauma Surg*. 2023;143(1):49–54.
28. **Bhandari M**, **Devereaux PJ**, **Swiontkowski MF**, **et al**. Internal fixation compared with arthroplasty for displaced fractures of the femoral neck. A meta-analysis. *J Bone Joint Surg Am*. 2003;85-A(9):1673–1681.
29. **Socci AR**, **Casemyr NE**, **Leslie MP**, **Baumgaertner MR**. Implant options for the treatment of intertrochanteric fractures of the hip: rationale, evidence, and recommendations. *Bone Joint J*. 2017;99-B(1):128–133.
30. **Guyen O**. Hemiarthroplasty or total hip arthroplasty in recent femoral neck fractures? *Orthop Traumatol Surg Res*. 2019;105(1S):S95–S101.
31. **Parker MJ**, **Khan RJK**, **Crawford J**, **Pryor GA**. Hemiarthroplasty versus internal fixation for displaced intracapsular hip fractures in the elderly. A randomised trial of 455 patients. *J Bone Joint Surg Br*. 2002;84-B(8):1150–1155.
32. **Horwitz DS**, **Tawari A**, **Suk M**. Nail length in the management of intertrochanteric fracture of the femur. *J Am Acad Orthop Surg*. 2016;24(6):e50–8.
33. **Laursen JO**. Treatment of intracapsular fractures of the femoral neck in Denmark: trends in indications over the past decade. *Acta Orthop Belg*. 1999;65(4):478–484.
34. **Anglen JO**, **Weinstein JN**, **American Board of Orthopaedic Surgery Research Committee**. Nail or plate fixation of intertrochanteric hip fractures: changing pattern of practice. A review of the American Board of Orthopaedic Surgery database. *J Bone Joint Surg Am*. 2008;90-A(4):700–707.
35. **Vidakovic H**, **Kieser D**, **Hooper G**, **Frampton C**, **Wyatt M**. Valgus-impacted subcapital neck of femur fractures: a systematic review, meta-analysis with cost analysis of fixation in-situ versus nonoperative management. *Hip Int*. 2024;34(2):260–269.
36. **Maheshwari K**, **Planchard J**, **You J**, **et al**. Early surgery confers 1-year mortality benefit in hip-fracture patients. *J Orthop Trauma*. 2018;32(3):105–110.
37. **Ryan DJ**, **Yoshihara H**, **Yoneoka D**, **Egol KA**, **Zuckerman JD**. Delay in hip fracture surgery: an analysis of patient-specific and hospital-specific risk factors. *J Orthop Trauma*. 2015;29(8):343–348.

Author information:
Z. Zheng, MD, Orthopaedic Surgeon
S. E. Kim, MD, Orthopaedic Surgeon
Department of Orthopedic Surgery, Seoul National University Hospital, Seoul, South Korea; Department of Orthopedic Surgery, Seoul National University College of Medicine, Seoul, South Korea.

B.Y. Ryu, MD, Orthopaedic Surgeon, Department of Orthopedic Surgery, Seoul National University Hospital, Seoul, South Korea.

D. S. Song, B.Eng, AI Engineer, CONNECTEVE, Seoul, South Korea.

S. H. Kim, MD, Orthopaedic Surgeon, Department of Orthopedic Surgery, Chung-Ang University Hospital, Chung-Ang University College of Medicine, Seoul, South Korea.

J-W. Park, MD, Orthopaedic Surgeon, Department of Orthopedic Surgery, Seoul National University College of Medicine, Seoul, South Korea; Department of Orthopedic Surgery, Seoul National University Bundang Hospital, Seongnam, South Korea.

D. H. Ro, MD, Orthopaedic Surgeon, Department of Orthopedic Surgery, Seoul National University Hospital, Seoul, South Korea; Department of Orthopedic Surgery, Seoul National University College of Medicine, Seoul, South Korea; CONNECTEVE, Seoul, South Korea; Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, South Korea.

Author contributions:
Z. Zheng: Data curation, Software, Writing – original draft, Writing – review & editing.
B.Y. Ryu: Data curation.
S. E. Kim: Supervision.
D. S. Song: Formal analysis, Software, Validation, Visualization.
S. H. Kim: Resources.
J-W. Park: Conceptualization, Project administration, Resources, Supervision, Writing – review & editing.
D. H. Ro: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

J-W. Park and D. H. Ro are joint senior authors.

Funding statement:
The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this article: this work was supported by the Technology Development Program (RS-2023-00258466)

funded by the Ministry of SMEs and Startups (MSS, South Korea), as received by D. H. Ro.

ICMJE COI statement:
D. H. Ro reports that this work was supported by the Technology Development Program (RS-2023-00258466) funded by the Ministry of SMEs and Startups (MSS, South Korea).

Data sharing:
The data that support the findings for this study are available to other researchers from the corresponding author upon reasonable request.

Acknowledgements:
We thank Jong Jin Go for the assistance in radiological confirmation of fracture classification.

Ethical review statement:
This study was approved by the Institutional Review Board of Seoul National University Hospital (IRB No. 1810-004-974). All investigations were conducted in conformity with ethical principles of research. Informed consent was exempted due to the retrospective nature of the study.

This article was primary edited by G. Scott.