ORIGINAL ARTICLE

Developing A Fast Computer Vision Model for Diagnosing and Classifying Hip Fractures

Muharrem Kanar[1], Ahmet Hamdi Olçar[2], Yusuf Sülek[1], Güngör Alibakan[1], Bilal Gök[3], Raffi Armağan[1]

[1]University of Health Sciences Turkey, Şişli Hamidiye Etfal Training and Research Hospital, Department of Orthopedic and Traumatology, İstanbul, Turkey
[2]University of Health Sciences Turkey, Sancaktepe Şehit Prof. Dr. İlhank Varank Training and Research Hospital, Department of Orthopedic and Traumatology, İstanbul, Turkey
[3]Mardin Training and Research Hospital, Clinic of Orthopedic and Traumatology, Mardin, Turkey

Abstract

Objective: The primary aim of this study was to refine the accuracy and efficiency of hip fracture detection using a new computer vision model based on the YOLOv8 algorithm, thereby addressing the current limitations in diagnostic methodologies and improving dataset constraints.

Methods: We conducted a retrospective study using anterior-posterior (AP) hip radiographs collected from adult patients at University of Health Sciences Turkey, Şişli Hamidiye Etfal and University of Health Sciences Turkey, Sancaktepe Şehit Prof. Dr. İlhank Varank Training and Research Hospital between January 2021 and January 2023. A total of 676 radiographs were analyzed after applying classifications according to the AO/OTA system by orthopedic specialists. The dataset was divided into training, validation, and testing sets, and image augmentations were applied to enhance model training.

Results: The YOLOv8 model achieved a mean Average Precision at 0.5 IOU (mAP50) of 0.877 at the 99th epoch, demonstrating high diagnostic accuracy with a precision rate of 0.891 and recall of 0.797. These metrics indicate the model's effectiveness in accurately detecting and classifying hip fractures.

Conclusion: This study presents a significant enhancement in the use of artificial intelligence for medical imaging, particularly in detecting and classifying hip fractures, thereby demonstrating the potential of AI to augment clinical decision-making. Further studies are recommended to expand the application scope and improve the model's accuracy in various clinical environments.

Keywords: Hip fractures, artificial intelligence, image processing, computer-assisted, deep learning

INTRODUCTION

In emergency services, diagnosis of fractures is important for guiding appropriate treatment to optimize patient outcomes. A missed diagnosis of either overuse or underuse of imaging techniques poses great risks. In addition to overuse, which wastes healthcare resources, these patients will be exposed to radiation for no necessary reason, whereas underuse may result in missed diagnosis. Misdiagnosis may result in delayed or inappropriate treatment, leading to negative impact on recovery time, increased healthcare costs, and, most importantly, potential harm to patients (1,2). These issues become all the more critical in musculoskeletal injuries, where proper imaging forms the cornerstone for effective clinical decision-making. Some of the most significant disadvantages of traditional X-ray radiographs for diagnosis of fractures are image quality, angle, and clarity problems inherent in the emergency situation, which may render the situation indecipherable to human eyes regarding whether a fracture has occurred. Fracture classification is a necessary tool for clinicians because first, there is now a common language with which to describe the type, location, and severity of a fracture. This not only facilitates clinical communications, but also supports research by allowing consistent comparisons

Eur Arch Med Res 2024;40(4):214-220

Kanar et al. AI Model for Hip Fracture Diagnosis

of treatment outcomes (3-5). Artificial intelligence (AI) is being inevitably deployed nearly every field of study, and medicine is not an exception. In the research area of diagnostic medicine, AI has held a huge promise, hence boosting accuracy, especially through computer vision techniques (6,7). However, models developed to this date achieve variable successes when applied to the field of fracture detection. Several studies have explored AI applications for automated medical image analysis, including fracture detection. However, most of these models have limiting factors: either they lack sufficient accuracy, especially for complex fractures, or large, high-quality public datasets that can be used for training more powerful models are not available. The present study aimed to address these gaps by developing a computer-vision-based model capable of diagnosing and classifying hip fractures from radiological images.The present research is based on previous studies, narrowing its focus to problems concerning hip fracture detection, a critical and complex area of orthopedic care. By creating a labeled dataset of radiological images and leveraging the YOLOv8 algorithm, a state-of-the-art object detection method, this study aims to improve both the accuracy and efficiency of fracture diagnosis in clinical settings. Our contributions will provide not only a practical tool for clinicians but also a dataset and methodology that can be further utilized by researchers in the field.

METHODS

Study Design

This retrospective study collected AP hip radiographs obtained from adult patients who presented to the University of Health Sciences Turkey, Şişli Hamidiye Etfal and University of Health Sciences Turkey, Sancaktepe Şehit Prof. Dr. İlhank Varank Training and Research Hospital emergency department between January 2021 and January 2023. Ethical approval for the study was obtained from the University of Health Sciences Turkey, Sancaktepe Şehit Prof. Dr. İlhank Varank Training and Research Hospital Non-Interventional Research Ethics Committee (approval number: 252, date: 13.12.2023). Ethical approval for this study was obtained before data collection and was conducted

in adherence to institutional and national guidelines for the purpose of maintaining patient confidentiality and protection.

Data Sets

The dataset consisted of 748 AP hip radiographs. During data pre-processing, 72 radiographs were excluded due to visual obstruction of the image, either by the presence of a patient's hand or any foreign object, such as keys or coins. A total of 676 hip radiographs were included in the study (Figure 1). Specific exclusions for "visible pollution" were developed in an attempt to make the exclusion process more standardized and include cases with visible extra-hip body parts, objects external to the patient, or where fracture visibility was unclear. The remaining images were then resized to 640x640 pixels using bilinear interpolation to retain the quality of the images during model training. Image quality after resizing was verified to ensure that the main diagnostic features, including bone structures and fracture lines, remained clear. Radiographs were then classified according to the AO/OTA classification system by two orthopedic specialists with 10 and 5 years of experience.

Fracture Labeling

Regions of interest in anterior-posterior (AP) hip radiographs were manually annotated using image annotation software. Fractures were labeled using the AO/OTA classification system, with classes represented as,

1. Normal (no fracture),

2. 31-B (femoral neck fracture),

3. 31-A1 (femoral simple pertrochanteric fracture),

4. 31-A2 (femoral multifragmentary pertrochanteric fracture),

5. 31-A3 (femoral intertrochanteric fracture) (Figure 2).

The orthopedic surgeons who evaluated the images classified all images, and any disagreements were resolved by expert consensus meetings. Ultimately, the same results were obtained in all evaluations. The final number of radiographs analyzed in this dataset was 676 and was divided into 300 normal hips, 110

Figure 1. Distribution of graphs used in the study (Roboflow, Inc., https://roboflow.com/)

Kanar et al. AI Model for Hip Fracture Diagnosis

Eur Arch Med Res 2024;40(4):214-220

femoral neck fractures (31-B), 133 femoral simple pertrochanteric fracture (31-A1), 68 femoral multifragmentary pertrochanteric fracture (31-A2), and 65 femoral intertrochanteric fractures (31-A3). The dataset was divided into three sets: 70% training, 20% validation, and 10% testing.

Augmented Images

Image augmentation increased the training set from 473 to 2365 images by applying transformations such as mirror horizontally (image flipped horizontally to simulate variation in patient positioning), rotation (images were rotated randomly between -15° and +15° to introduce minor variations in position for the radiographs), blur (this involved the use of up to 1.25 pixels of random blurring to simulate motion or low-quality differences in image capturing equipment), noise (random noise, up to 6% of the image, was added to simulate real variability in radiographs), exposure (the exposure values were randomly changed to vary between -16% and +16% to simulate different lighting conditions

Figure 2. Labelling of the graphs used in our study (Roboflow, Inc., https://roboflow.com/)

at the time of radiography). These augmentations were considered to reflect clinically relevant variations to improve the generalizability of the model to different clinical scenarios.

Statistical Analysis

The YOLOv8 algorithm from Ultralytics was used to develop a computer vision model for hip fracture detection. The parameters used during training were Learning rate 0.001, Batch size auto, and epoch number 100 (Figure 3). YOLOv8 is known for its efficient real-time object detection and segmentation capabilities, which make it ideal for medical image analysis tasks. The use of YOLOv8 enabled the localization and classification of fractures. The performance of the model was measured in terms of accuracy, sensitivity, and specificity.

RESULTS

The mean Average Precision at 0.5 IOU (mAP50), which measures model performance based on how well it identifies objects at a threshold of 0.5 IOUs, was 0.877 at the 99th epoch. This score indicates that most of the predicted bounding boxes overlapped with the actual annotations, demonstrating the high degree of accuracy obtained in locating fractures. Our model's precision was 0.891, indicating that most of the fractures identified were positive. We obtain a recall value of 0.797, which shows that the model has a strong true positive rate, which is an essential value in the clinical setting because missing a fracture may lead to severe harm to the patient. The results of the AI model (performance metrics and losses in training) over time are shown in Figure 4, which highlights the evolution of the model across subsequent steps. The first graph shows an increasing accuracy with slight variations in the values: a vague sign indicates that this model has been improving its object recognition and classifier skills over time. More specifically, validation metrics always increase from that point onward, which means that the model generalizes very well and can perform in unseen data. The box loss graph shows

Figure 3. Training arguments of our model (Ultralytics LLC, https://www.ultralytics.com/).

the decrease after training in predicting correct boxes. The results demonstrate that the detection sensitivity has improved and allows for more accurate location of objects. A downward trend is also displayed in the "Class Loss" graph, which indicates better recognition and classification of object classes, as demonstrated by the ability of the proposed method to differentiate different object types more accurately. The final cut in the "Object Loss" graph shows that the model finally detected object presence in images with far fewer errors, leading to fewer false positives and false negatives.

DISCUSSION

The most key finding of this study is the capability of the YOLOv8 model, which can achieve an mAP50 of 0.877, proving that it is highly precise in detecting and locating hip fractures. Considering that the proposed method achieved this level with a low dataset of only 676 images, it also demonstrates how efficiently the model can learn from fewer data without compromising

diagnostic accuracy. This is remarkable compared with findings from similar studies, such as those by Jiménez-Sánchez et al. (8) and Tanzi et al. (8,9) which made use of larger datasets but realized mPA values of 0.87 and 0.81, respectively. Efficiency in this matter from our model is a crucial aspect of a clinical setting where the need for speed and preciseness in diagnosis is paramount toward appropriate patient management. Jiménez-Sánchez et al. (8) used the ResNet-50 and AlexNet architectures to develop deep learning classification and localization models for 1347 images. These models were performed with an mPA value of 87. Tanzi et al. (9) developed a multi-stage architecture using 2453 images. This architecture consists of successive CNNs gradually. These models were performed with an mPA value of 0.81 (9). In another study conducted in 2022, Tanzi et al. (10) obtained an accuracy of 83% in fracture estimation using an architecture consisting of CNNs on 4207 images. This model predicted 29% better than 11 orthopedic surgeons. In this study, we created a model with 676 images using the YOLOv8 system in about 5 hours. This model had an mPA50 value of 0.877.
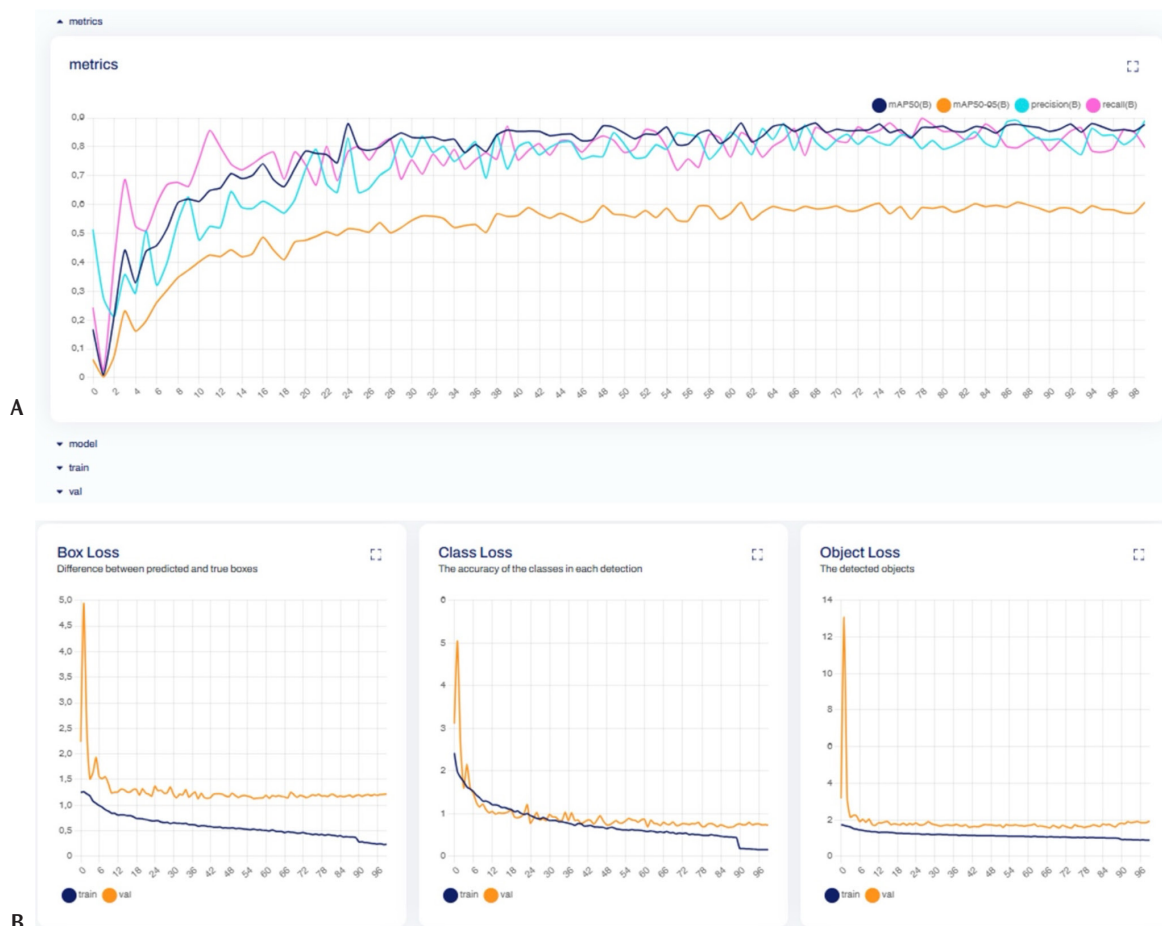

Figure 4. (A, B) Model accuracy measured on validation set, object loss, class loss, box loss graph during training of our model (Object loss measures the algorithm's ability to correctly predict whether there are objects in an imagelabel. Class loss measures the algorithm's ability to correctly guess which class objects in an image belong to. Box loss measures the algorithm's ability to accurately predict the position and shape of objects in an image) (Ultralytics LLC, https://www.ultralytics.com/)

Kanar et al. AI Model for Hip Fracture Diagnosis

Eur Arch Med Res 2024;40(4):214-220

Despite using fewer images, our model produced outcomes comparable to those of other models in the literature. The training performance of the proposed model is shown in Figure 4. As can be seen, the proposed model follows the learning curve steadily. Figure 5 shows some of the test images that illustrate how our model detects and classifies proximal femur fractures. In solving AI problems, artificial neural networks model the connections between biological counterparts using weights between nodes. A positive weight reflects exciting relations, whereas inhibition links are represented by negative values. The sum of the products obtained from the weighted inputs determines the overall model output. Such common architectures are CNN and UNET (11). This research study employed Ultralytics YOLOv8, the most advanced form of the real-time object detection and image segmentation model. YOLOv8 is an engine behind a range of cutting-edge deep learning and computer vision advances that allow it to realize very fast speeds coupled with high accuracy. In this regard, our model was more efficiently developed with a higher predictive value. Ultralytics' YOLOv8 is the latest version of a well-known real-time object detection and image segmentation model.

Built on top of the latest developments in deep learning and computer vision, YOLOv8 performs incomparably with respect to the features of speed and accuracy. Whereas in some studies, a lot of data input is required for the classification and reporting of proximal femur fracture, in our study, from the results, it will be very obvious that the machine learned in a very short period of time with less data (12). For the model in our studies, we used 657 hip X-ray images, and we noticed that CNN training generally requires thousands of X-rays and some programing expertise. YOLOv8 has indeed eased our lives as medical professionals in terms of model building and has been an immense help in the spread of AI in medicine. This development underlines not only the efficiency of YOLOv8 for handling medical imaging tasks with limited datasets and demonstrates that AI technologies are becoming increasingly accessible and applicable in health care; thus, they could be a game-changer in diagnostic and therapeutic practices. Many researchers identify and classify fracture performance by comparing various computer vision models developed to those of doctors in recognizing and classifying fractures (13). These state-of-the-art studies present computer vision models with higher accuracy rates than those of
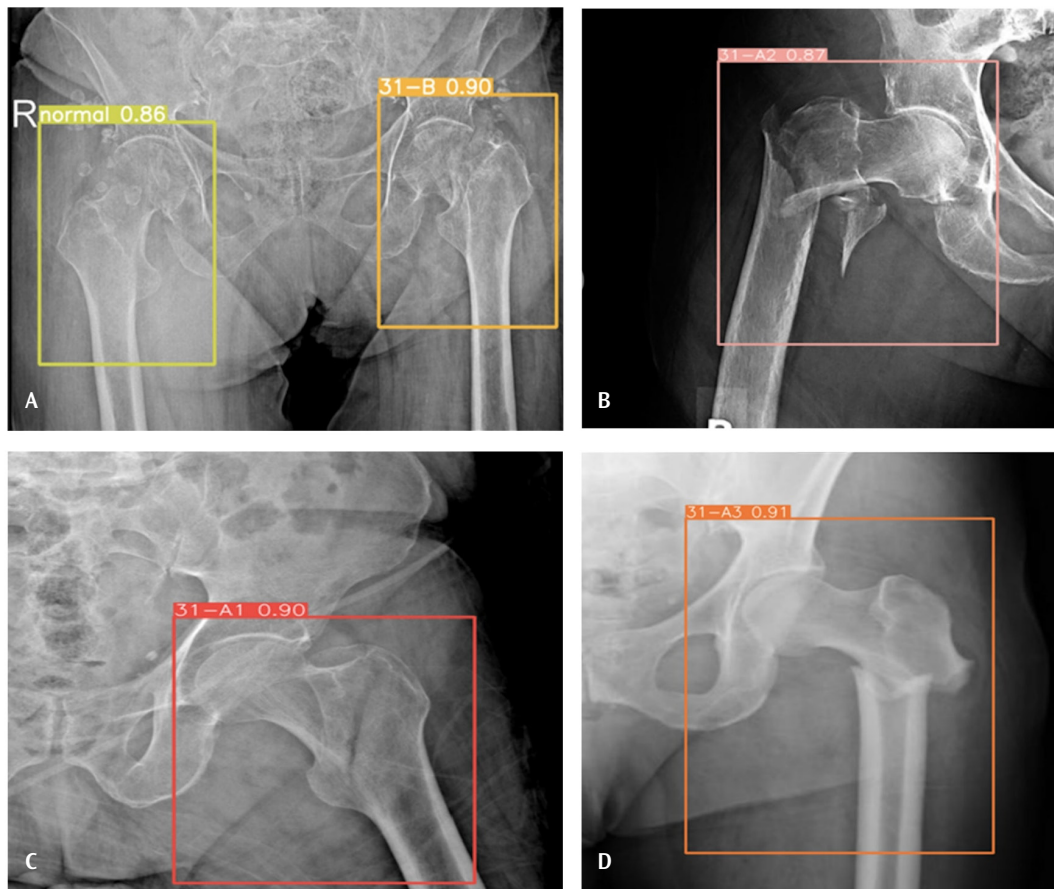

Figure 5. (A-D) Some images in which we tested our model's introduction of proximal femur fractures (Ultralytics LLC, https://www.ultralytics.com/)

Eur Arch Med Res 2024;40(4):214-220

Kanar et al. AI Model for Hip Fracture Diagnosis

physicians. In our work, there is no direct comparison between doctors and our model. No matter how good the identification and classification of a fracture are using computer vision models, the doctor always has the discretion on treatment planning. For this reason, computerized vision methods used in fracture diagnosis should be regarded as a means to speed up the diagnosis process for physicians and as a tool to contribute to the training of residents. Because we thought that differences in proximal femur radiographs could reduce the reliability of our model, we excluded such radiographs from our study. We have excluded of some radiographs based on predefined strict criteria aimed at maintaining high image quality for effective model training. In this way, we attempted to prevent difficulties that may arise in the diagnosis and classification of fractures in radiographs that were not taken in the appropriate position. Our study only used hip radiographs from adult patients, so we were not able to evaluate the performance of our model in pediatric patients. In addition, since our model did not include radiographs of patients with additional pathologies, such as coxarthrosis, bone cysts, pelvis, and acetabular fractures, we could not test our ability to recognize and classify proximal femur fractures in such cases.

Study Limitation

Epidemiological research indicates a progressive escalation in the incidence of PFFs (proximal femoral fractures) with advancing age, commencing at 40 years and accelerating markedly beyond the age of 75 (14). Consequently, this study exclusively utilized adult radiographic images. The model developed here has not been evaluated for application to pediatric populations. The quality and quantity of a dataset is very important for fracture detection and classification in computer vision. The fact that the images in the dataset are labeled correctly and are sufficiently clear increases the success of the model. Therefore, cleaner and high-quality images are required for each class to develop a better model. When we examine the studies in the literature, we find that the desired performance can be achieved using an appropriate artificial neural network architecture; however, the models and datasets used are not shared. This makes it difficult for academic publications to be verifiable and reproducible. In addition, the publication of the model used by the authors will contribute to scientific research. Therefore, in our study, we tried to overcome this problem by presenting both our model and the labeled data set in the appendix of our publication.

CONCLUSION

This study demonstrated the power of the YOLOv8 model in the detection and classification of proximal femoral fractures, with an mAP50 of 0.877 with high precision of 0.891, and recall of 0.797. These findings indicate that AI diagnostics can be highly accurate and reliable and thus have great potential to improve clinical decision-making. The results also demonstrate the efficiency of the model on a smaller but well-annotated dataset, thus reducing computational demands with the intent of making advanced diagnostics more accessible to resource-constrained medical facilities. This is encouraging further studies to extend this model to a wide range of different patient demographics and fracture types, hence broadening the clinical utility of the tool to help further advances in medical imaging.

Ethics

Ethics Committee Approval: Ethical approval for the study was obtained from the University of Health Sciences Turkely, Sancaktepe Şehit Prof. Dr. İlhank Varank Training and Research Hospital Non-Interventional Research Ethics Committee (approval number: 252, date: 13.12.2023).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.K., A.H.O., Y.S., G.A., B.G., R.A., Concept: M.K., A.H.O., Y.S., G.A., B.G., R.A., Design: M.K., A.H.O., Y.S., B.G., R.A., Data Collection or Processing: M.K., A.H.O., Y.S., G.A., B.G., R.A., Analysis or Interpretation: M.K., A.H.O., B.G., R.A., Literature Search: M.K., A.H.O., Y.S., G.A., B.G., R.A., Writing: M.K., A.H.O., Y.S., G.A., B.G., R.A.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. 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;29:343-8.

2. French SD, Green S, Buchbinder R, Barnes H. Interventions for improving the appropriate use of imaging in people with musculoskeletal conditions. Cochrane Database Syst Rev. 2010;2010:CD006094.

3. Bachmann LM, Haberzeth S, Steurer J, ter Riet G. The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review. Ann Intern Med. 2004;140:121-4.

Kanar et al. AI Model for Hip Fracture Diagnosis

Eur Arch Med Res 2024;40(4):214-220

4. Beckenkamp PR, Lin CC, Macaskill P, Michaleff ZA, Maher CG, Moseley AM. Diagnostic accuracy of the Ottawa Ankle and Midfoot Rules: a systematic review with meta-analysis. Br J Sports Med. 2017;51:504-10.

5. Marsh JL, Slongo TF, Agel J, Broderick JS, Creevey W, DeCoster TA, et al. Fracture and dislocation classification compendium - 2007: Orthopaedic Trauma Association classification, database and outcomes committee. J Orthop Trauma. 2007;21:S1-133.

6. Singh A, Randive S, Breggia A, Ahmad B, Christman R, Amal S. Enhancing prostate cancer diagnosis with a novel artificial intelligence-based web application: Synergizing deep learning models, multimodal data, and insights from usability study with pathologists. Cancers (Basel). 2023;15:5659.

7. Agrawal D, Poonamallee L, Joshi S, Bahel V. Automated intracranial hemorrhage detection in traumatic brain injury using 3D CNN. J Neurosci Rural Pract. 2023;14:615-21.

8. Jiménez-Sánchez A, Kazi A, Albarqouni S, Kirchhoff C, Biberthaler P, Navab N, et al. Precise proximal femur fracture classification for interactive training and surgical planning. Int J Comput Assist Radiol Surg. 2020;15:847-57.

9. Tanzi L, Vezzetti E, Moreno R, Aprato A, Audisio A, Massè A. Hierarchical fracture classification of proximal femur X-Ray images using a multistage Deep Learning approach. Eur J Radiol. 2020;133:109373.

10. Tanzi L, Audisio A, Cirrincione G, Aprato A, Vezzetti E. Vision transformer for femur fracture classification. Injury. 2022;53:2625-34

11. Xu F, Xiong Y, Ye G, Liang Y, Guo W, Deng Q, et al. Deep learning-based artificial intelligence model for classification of vertebral compression fractures: A multicenter diagnostic study. Front Endocrinol (Lausanne). 2023;14:1025749.

12. Paalvast O, Nauta M, Koelle M, Geerdink J, Vijlbrief O, Hegeman JH, et al. Radiology report generation for proximal femur fractures using deep classification and language generation models. Artif Intell Med. 2022;128:102281.

13. Zeren MT, Arslankaya S, Altuntaş Y, Cam N, Kırelli Y, Özdemir MH. Doctors versus YOLO: comparison between YOLO algorithm, orthopedic and traumatology resident doctors and general practitioners on detection of proximal femoral fractures on X-ray Images with multi methods. International Journal on Artificial Intelligence Tools. 2024;33:2350056.

14. Hashimoto K, Shinyashiki Y, Ohtani K, Kakinoki R, Akagi M. How proximal femur fracture patients aged 65 and older fare in survival and cause of death 5+ years after surgery: A long-term follow-up. Medicine (Baltimore). 2023;102:e33863.