From Pixels to Diagnosis: A Systematic Review of Deep Learning in Femoral Fracture Detection and Classification

Swetank Shree*, *Rahul* Bhagawati, *Souptick* Chanda, *Debanga* Raj Neog
Indian Institute of Technology Guwahati, Assam, India

Abstract. Femoral fractures are becoming more common and require fast and accurate diagnosis, which makes them a significant worldwide health concern for older people. It is a major global health problem because it is becoming more common and needs a quick and accurate diagnosis. Traditional X-ray image interpretation risks human error and a lack of consistency, mainly in emergencies. To address these challenges, this review paper explores the development and application of deep learning (DL) techniques, using convolutional neural networks (CNNs) and Vision Transformers (ViTs), for automated femur fracture detection and classification using X-ray and CT imaging. Several models showed excellent diagnostic performance: the Faster R-CNN achieved a multi-class accuracy of 90% with an IoU of 0.87, the ViTs achieved an accuracy of 92% with an AUC of 0.94, and the ResNet50 achieved up to 95% accuracy. Advanced techniques like curriculum learning, attention mechanisms, and data augmentation with GANs have further enhanced the robustness and interpretability of the model. Although these approaches can help radiologists to accurately and quickly recognize fractures, there are limitations in dataset uniformity, transparency, and real-world integration. Clinical adoption requires further study.

Keywords. Femoral fracture, Deep Learning, Convolutional Neural Networks (CNN), Vision Transformer (ViT), X-ray Imaging, Fracture Classification.

*Corresponding author: s.swetank@iitg.ac.in

1 Introduction

The body is supported, movement is allowed, and essential organs: brain, heart, and lungs are protected by the human skeleton which is made up of bones. The skeleton is made up of several types of bones: long, short, flat, irregular, and sesamoid bones. They all have different shapes and functions. Bones are strengthened by bone cells and calcium. This kind of connective tissue is strong [1]. The thinner center, known as bone marrow, is where blood cells are produced. Flat bones protect your vital organs, while femur as a long bone, help you move and hold weight. Bones are hard but flexible structures that can bend when they are forced. But too much force can break them, making them less stable. When an external force is stronger than the bone, it can break it either partially or entirely. This breaks down the structure of the bone, which can cause pain, loss of function, and sometimes bleeding or damage to tissue around it. Age affects bone fractures. For example, children’s bones are more flexible, so they may not break all the way through. Older adults, on the other hand, tend to have weaker bones because they lose bone mass, which makes them more likely to break. Also, the amount of nutrition impacts bone fractures. If bones don’t get enough nutrients, they may become weaker.

Femur fractures are a health problem that affects older people all over the world, along with higher rates of disease and fatalities. As the world’s population grows, this is happening increasingly, It can be difficult to determine what kind of femoral fracture a patient has, and it takes a lot of skill to do it right in an emergency room with a large number of patients. The severity and location of the fracture determine the choice of surgical procedure and implant, suggesting that there must be precise classification. The Arbeitsgemeinschaft für Osteosynthesefragen and Orthopaedic Trauma Association (AO/OTA) system, illustrated in fig. 1 is among the most widely used classification systems. It classifies fractures according to location, structure, and complexity. The manual classification of fractures using radiographs poses many challenges[2]. The procedure is tedious and can show inter- and intra-expert variability among newly trained radiologists. These challenges highlight the importance of a computer-aided diagnostic (CAD) system that can assist clinicians detect femur fractures on radiographs in a timely manner.

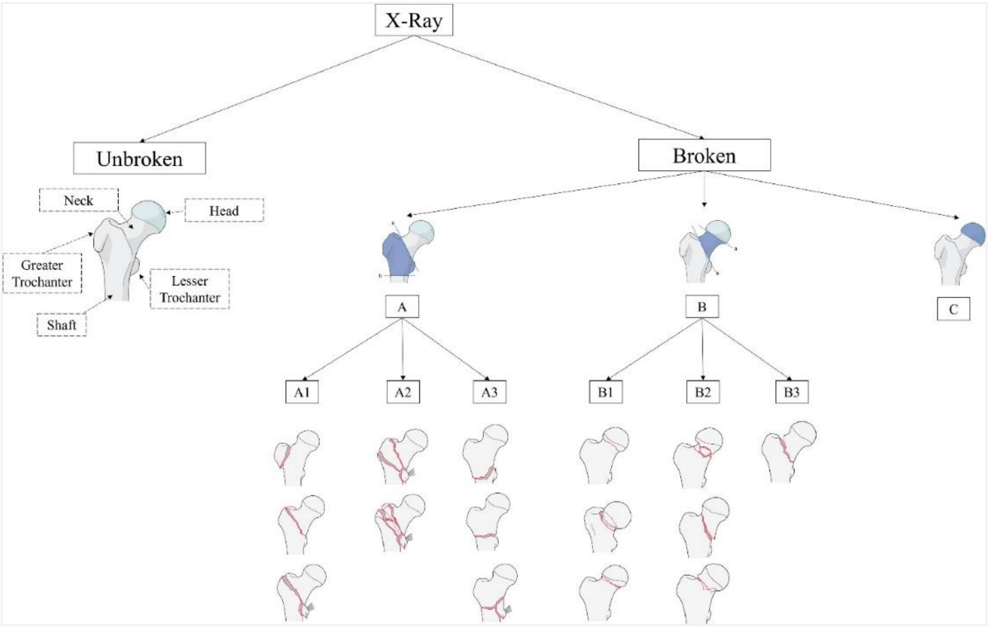

Fig. 1. The above fig. is AO/OTA classification system. These type A, B, and C fractures occur in the trochanteric region, femoral neck, and femoral head. This classification is based on the position and arrangements of the fracture lines [2].

These issues can be handled using Deep Learning (DL) approaches, along with artificial intelligence (AI) as a possible solution. It is a type of machine learning that uses convolutional neural networks (CNNs), which have become very popular. This system recognizes how to detect and classify complex fractures from medical images. This speeds up and improves the accuracy of diagnosis, which can help radiologists make fewer mistakes and improve patient outcomes [3]. The CNN model has a well-defined pipeline that helps improve image quality and minimize noise. These steps are needed to prepare raw X-ray data for accurate analysis. A thresholding technique is used to separate the bone area from the background during segmentation. Following segmentation, feature extraction is carried out to determine the shape and structure of bones, which helps identify defects. Deep learning also improves images by translating and combining them. It helps with tasks like detecting lesions, measuring clinical parameters, planning treatments, and assisting in surgeries. This review shows how deep learning has changed over the past few years for finding and classifying femur fractures, from early proof-of-concept studies to the most recent advances. We systematically analyzed the methodologies, datasets, performance metrics, and clinical significance of the studies. We put together the state-of-the-art in a comparative table. We then look at how these methods affect patients, highlight current gaps, and suggest future directions for research and clinical adoption.

2 Background

The increase in femoral fractures emphasizes the significance of rapid and precise diagnosis to ensure that the patients receive the best care. Radiologists and healthcare professionals rely on medical imaging such as X-rays and computed tomography (CT) scans for traditional diagnosis. This manual approach, however, has several challenges, including differences in interpretation, errors in diagnosis due to fatigue and the lack of poor accuracy for complex fractures. These challenges become even more serious in emergency situations or in busy hospitals where the physicians must make swift and precise decisions. In these situations, mistakes are more prone to occur especially because of the inexperienced medical staff.

Owing to these issues, deep learning (DL) methods, especially convolutional neural networks (CNNs) hold great potential for tasks that require high-level feature extraction and recognition. Earlier applications of CNN involved fracture detection focusing on binary classification tasks where the presence or absence of a fracture was determined. Models have been built to detect fractures at different regions of the body such as femoral neck fractures, intertrochanteric femur fractures, and distal radius fractures [4]. With further improvements, the methodologies have moved to more challenging tasks such as multi-class fracture classification, detection and localization. Several CNN architectures such as RestNet, DenseNet, Faster R-CNN, InceptionV3, RetinaNet and Vision Transformer have significantly improved the accuracy and efficiency of fracture classification. Object detection methods such as Faster Region-based CNN (Faster R-CNN) and You Only Look Once (YOLO), help locate the exact position of the fracture in the image using bounding boxes and simultaneously classify the type of fracture.

Recently, Vision Transformers (ViTs), which use attention-based approaches instead of traditional CNN methods, have become popular. ViTs effectively capture relationships across the entire image, helping to classify complicated fractures accurately. Studies have indicated that ViTs outperform standard CNN models, making diagnoses more accurate and easier for doctors to trust and adopt [5]. Figure 2 demonstrates the steps used for the classification process in ViT.

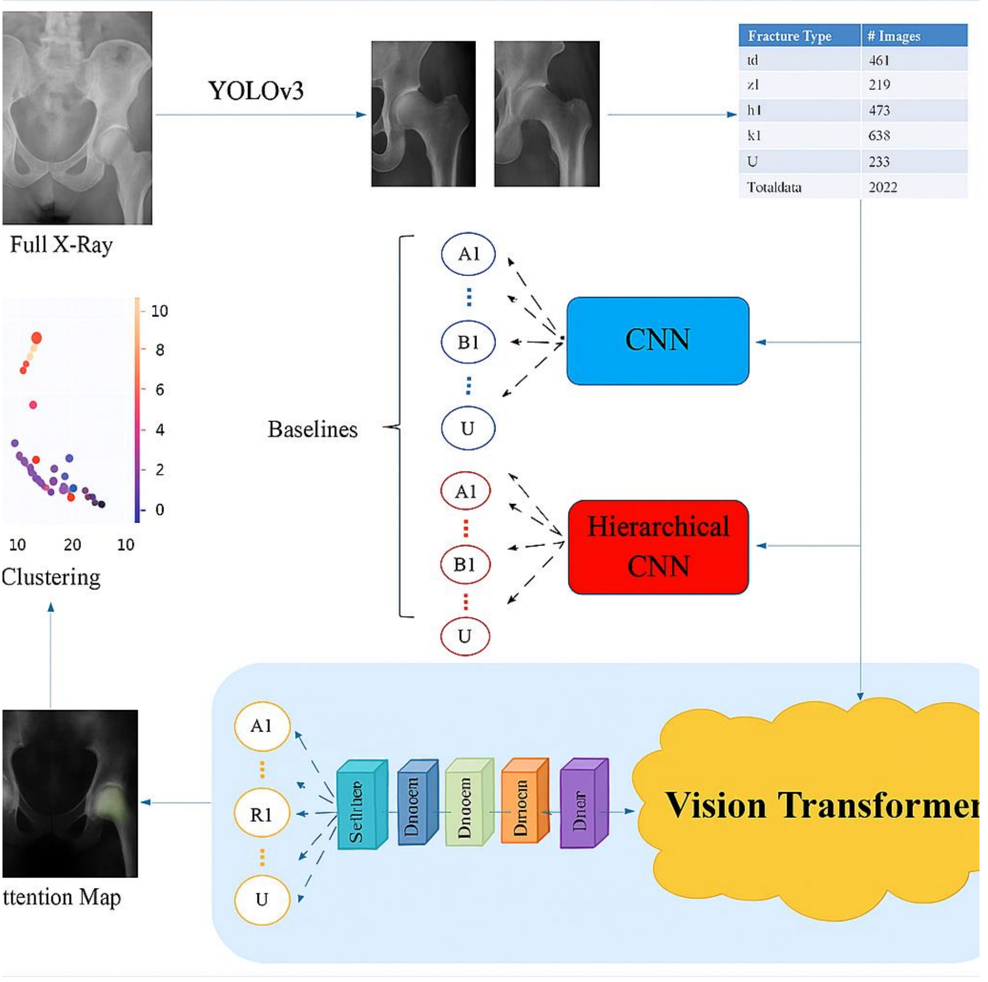

Fig. 2. The YOLOv3 network was used to process images by cropping the region corresponding to the left and right femurs. A CNN and a hierarchical CNN were then applied to the dataset, and the resulting models were used as baselines comparisons. Next attention maps were analyzed using a modified ViT encoder for 7-class classification. Finally, clustering techniques were used to evaluate ViT's feature extraction capability of the ViT model [5].

Limited data is the major concern in medical imaging. To overcome these issues, data augmentation methods like Generative Adversarial Networks (GANs) and Digitally Reconstructed Radiographs (DRRs) were used by researchers. These techniques assist in generating synthetic medical images from existing data, thereby enhancing the accuracy and reliability of models. Methods like curriculum learning (teaching models using progressively more challenging examples) and genetic algorithms (automatically optimizing CNN parameters) also help improve model performance [6]. With these developments, several challenges remain, including inconsistent labelling standards, lack of data availability, and concerns regarding the transparency of automated decision-making. Solving these challenges requires bigger and consistent datasets, better visualization techniques to help understand model decisions, and easier integration of advanced tools into clinical practice.

3 Methods

Many studies have researched multiple deep learning methods to address the problem of femur fractures. The approaches can be categorized into three parts: image classification, object detection, and segmentation tasks. CNN architectures such as AlexNet, VGG, and ResNet were used with pre-trained large datasets, then fine-tuned for fracture detection. More recent studies introduced two-step methods: first, localizing the hip/femur region, then classifying fracture type, as well as architectures like attention mechanisms and Transformers to improve focus on fracture features [7]. Deep learning models have been improved using advanced techniques like spatial transformer networks for region alignment, generative adversarial networks (GANs) for synthetic data augmentation, and Bayesian optimization for hyperparameter tuning.

3.1 Datasets and Annotations

Annotated datasets of high quality are essential. The standard of dataset creation and annotation is important for the training, and clinical significance of deep learning models in femoral fracture detection. The studies that were looked at used different imaging methods (X-ray/CT/MRI), labelling methods, and classification criteria to identify the types of fractures and where they took place [1]. Common annotations consist of binary labels (fracture vs. normal) and occasionally multi-class labels following the AO/ATO classification system. It categorises fractures based on their location and complexity, or the Garden classification for femoral neck fractures. A dataset consisting of 350 axial CT images of patients with intertrochanteric femoral fracture was collected from a single medical institution. Two experienced orthopaedic surgeons manually plotted the fracture boundaries and labelled each image according to the AO/OTA classification system (Types A1, A2, and A3). We reached a consensus on how to fix the differences in labelling. We made sure that dataset was evenly split between all types of fractures to avoid class imbalance during training. In a more comprehensive effort, about 7,000 high-resolution X-ray images were collected from several clinical centres. Two certified radiologists agreed to classify each image as either “fractured” or “non-fractured” [10]. Potential fracture locations were highlighted with the help of a semi-automated method; however final validation was done manually. To improve the quality` training, cases that were unclear or overlapped were discarded, and the dataset was divided up so that both classes were equally represented. As shown in Figure 3, cross validation is used in smaller datasets, and patient-wise data splits are performed for training, validation, and testing to prevent leaks.

https://doi.org/10.1051/bioconf/202520401018

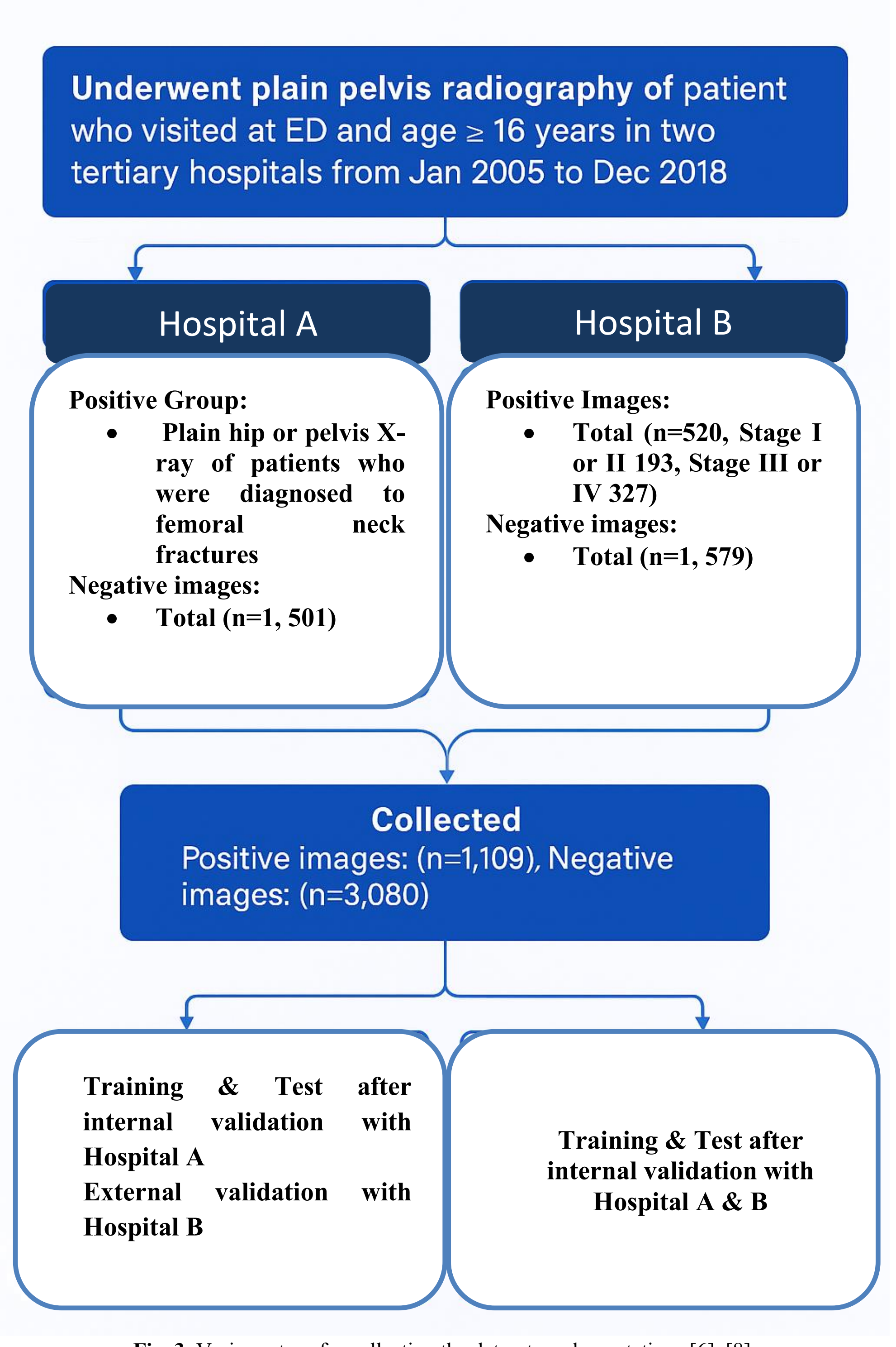

Fig. 3. Various steps for collecting the datasets and annotations [6], [8].

3.2 Image Classification

Predicting a class label for an entire input image, such as “fracture” or “not fracture”, or specific fracture subtypes, is the main goal of image classification models. Transfer learning is used in many research, in which a CNN is fine-tuned on a medical image dataset after first being pre-trained on a large-scale medical image dataset such as ImageNet. inceptionV3, Resnet, VGG, DenseNet, and EfficientNet are some examples of common architectures. Tanzi et al. did a study that used a multi-stage, hierarchial approach with a cascade of InceptionV3 networks that were specially designed for the AO/OTA classification system [2]. This system separates fractures by their location and shape. Another study looked at periprosthetic femur fractures (PFFs) and compared different models. It found that Resnet50 worked best for both binary and multi-class classifications. Bae et al. made a model based on ResNet18 with a convolutional Block Attention Module (CBAM) that can find femoral neck fractures, whether they are displaced or not.

To solve the common problem of limited and unbalanced medical datasets, many people use data augmentation methods. Classic methods use random variations like flipping, rotating, scaling, and changing the brightness and contrast [9]. Generative Adversarial Networks (GANs) and Digitally Reconstructed Radiographs (DRRs) are two more advanced methods for making synthetic images. These methods add more variety and size to the training data. Mutasa et al. conducted a study showing that the inclusion of these advanced techniques substantially enhanced the efficacy of their model in detecting and classifying femoral neck fractures.

3.3 Object Detection

Object detection models go beyond classification by identifying the location of a fracture within an image using bounding boxes and its type. Faster R-CNN with a Feature Pyramid Network (FPN) backbone is a popular choice for this task. Qi et al. used a Faster R-CNN model with a ResNet-50 and FPN backbone to locate and classify nine types of femoral shaft fractures. Alzaid et al. evaluated Faster R-CNN and RetinaNet for detecting and classifying periprosthetic fractures. Potter et al. created a model based on the VarifocalNet FPN for finding and locating proximal femur fractures from start to finish. This model worked better than several benchmark FPNs and even a transformer model [10].

3.4 Segmentation

Segmentation models are used to accurately define the exact boundaries of a fractured bones, which is a crucial step for 3D reconstruction and surgical planning. Vicory et al. used a 3D U-Net model on CT volumes to segment fractured femurs, including small fragments, which is important for creating accurate patient-specific 3D models. The authors further studied the impact of different loss functions on model performance and discovered that combining Dice and Binary Cross-Entropy (BCE) loss worked effectively.

3.5 Advanced Training and Interpretability Methods

Some studies have focused on optimizing the training process and making the models' predictions more transparent using the methods of curriculum learning and explainable AI (XAI). Jiménez-Sánchez et al. introduced a curriculum learning (CL) approach to improve classification performance when dealing with limited, imbalanced, and noisy datasets [11] This method involves strategically ordering, sampling, or weighting training data from "easy" to "hard" examples, using prior clinical knowledge or a self-paced uncertainty score. Several works incorporated visualization techniques to interpret model behaviour. Gradient Class Activation Maps (Grad-CAM) and Eigen-CAM generate heatmaps highlighting the specific regions of an image that the model focused on to make its prediction. They help build trust and allows clinicians to verify that the model is learning relevant features.

3.6 Performance Metrics

The performance of these models is evaluated using a range of metrics, including accuracy, precision, sensitivity (recall), specificity, F1-score, and area under the ROC curve (AUC). For detection tasks, mean average precision (mAP) at a certain IoU threshold is used to measure joint localization-classification performance [4].

In predictive modelling, there are four key metrics to evaluate the performance of a classifier. The number of occurrences that were correctly predicted to be positive is known as true positives, or TP. True negatives (TN) refer to instances that were correctly identified as not having the condition. False positives (FP) are cases where models incorrectly predicted as positive. Lastly, false negatives (FN) are instances that are incorrectly classified as negative. These measures are important for evaluating model’s accuracy and precision, when the risks of misinterpretation are high.

1) Accuracy: It shows how many predictions the model got correct out of all the predictions it made. The accuracy ranges from 0 to 1, indicating how well the model differentiates between fractured and non-fractured cases.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (1)$$

2) Precision: It represents the percentage of true positives that were accurately categorized as positive. It indicates the model’s efficacy in accurately identifying fracture cases while excluding false positives.

$$Precision = \frac{TP}{(TP+FP)} \qquad (2)$$

3) Sensitivity (Recall): Sensitivity is also known as Recall. It measures the model's capability to detect actual positive cases. It is the ratio of correctly identified positives to the total actual positives.

$$Sensitivity = \frac{TP}{(TP+FN)} \qquad (3)$$

4) Specificity: The specificity demonstrates the proportion of unbroken bones correctly recognized as non-fractured. The specificity values range from 0 to 1 to show the probability of having no fractures, conditioned on truly being non-fractured.

$$Specificity = \frac{TN}{(TN+FP)} \qquad (4)$$

5) Area under the curve (AUC): The AUC is the area under the ROC curve. The AUC value can analyze full prediction scores without setting a particular threshold. The AUC value equals the area under the ROC curve to the x-axis, which exploits sensitivity as the ordinate and (1-specificity) as the abscissa.

6) Mean Average Precision (mAP): mAP aggregates the average precision across all object classes into a single performance score. Let C be the total number of classes, and let AP_1, AP_2, ..., APC be the average precisions for each class.

$$mAP = \frac{1}{C}\sum_{i=1}^{c} AP_i \qquad (5)$$

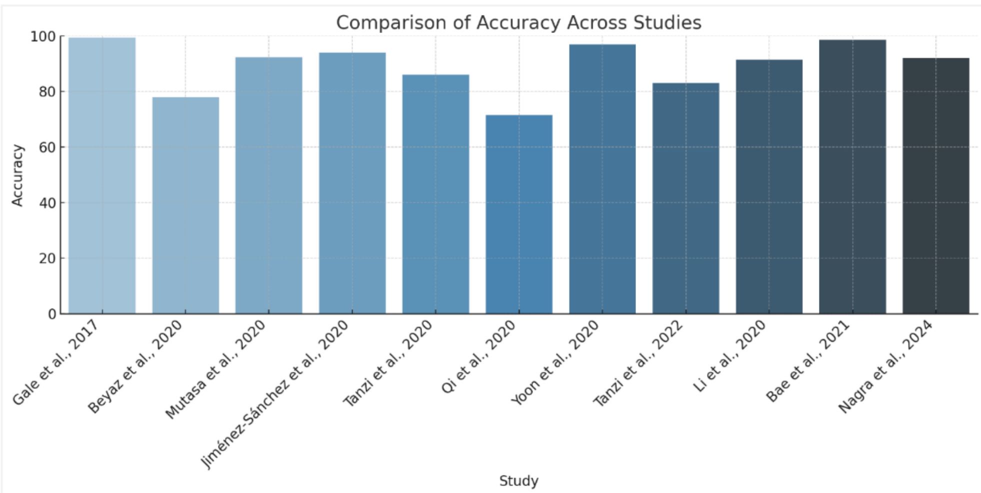

Fig. 4. Performance metrics, year-wise comparison.

4 Results and Discussion

Proximal femur fracture (PFF) detection and classification using radiographic images has been substantially enhanced with recent advances in deep learning. Obtaining a 94% F1-score and up to 95% accuracy in a binary classification task, ResNet50 showed outstanding performance. An excellent average accuracy of 90% was maintained when it was broadened to multi-class classification involving AO/OTA fracture types (A, B, and C). The faster R-CNN outperformed RetinaNet in localization tasks, achieving higher mean Intersection-over-Union (IoU) and average precision, resulting in better fracture region annotations [12].

Vision Transformers (ViT) were effective alternatives that improved the performance of CNNs. With an accuracy of 92% and an AUC of 0.94, ViT models outperformed conventional convolutional architectures. Grad-CAM-generated heatmaps showed that ViT models could effectively focus on clinically significant fracture locations, even in obscure or complex instances, boosting diagnostic reliability.

Prior methods that used synthetic augmentation techniques, such as GANs and digitally reconstructed radiographs, for better model generalization were expanded upon by these advances. The accuracy of fracture detection and subtype identification was substantially improved by systems incorporating the localization and classification phases [13]. Hierarchical methods further improved multi-stage classification, which aligned with AO classification. The study also showed that artificial intelligence might significantly boost physicians' accuracy in diagnosis. Deep learning models have shown great potential in detecting, localizing, and classifying femur fractures, especially those that use architectures like ResNet50, Faster R-CNN, and ViT. Combined with expert guidance, their integration into clinical workflows improves diagnostic precision and speeds up decision-making in critical situations. These tools are progressively noticed as replacements and collaborative aids that support radiologists and orthopaedic surgeons in managing fracture cases more effectively and accurately.

Deep learning models for femur fracture detection have rapidly progressed [14]. By 2020, several models achieved sensitivities and specificities between 85 to 95%, which means they can detect many fractures with a low false-positive rate. For instance, Mutasa et al.'s system detected hip fractures with 91% sensitivity at 93% specificity, and Tanzi et al.'s 2020 multi-stage CNN achieved 86% accuracy on multi-class classification tasks [8]. These algorithms also have a high negative predictive value (NPV). Mutasa's two-class classification model achieved an NPV of 0.86 at approximately 92% accuracy, which indicates confidence in ruling out fractures. A high NPV is crucial for an evaluation tool to safely say an X-ray has no fracture. However, even with 90% AUCs, one must consider that an algorithm may still miss certain fracture types. Therefore, clinicians cannot blindly trust an AI result; these systems are best utilized as a second opinion. The information about the papers, including the published year, methods, dataset type and size, key performance, and clinical impact, is summarized in Table 1 and Figure 4.

Table 1: Comparison of deep learning studies on femur fracture detection/classification by model, dataset, performance, and clinical impact. (Acc = Accuracy; AUC = area under ROC curve; IoU = intersection-over-union; IT = intertrochanteric)

Sl.No.	*Study (Year)*	*Deep Learning Method*	*Data (Images & Type)*	*Key Performance*	*Clinical Impact*
(1)	Beyaz et al., 2020	5-layer CNN + Genetic Algorithm	234 pelvic X-rays (augmented to 2106)	Sensitivity: 83%, Specificity: 73%, Accuracy: 78–79%, Kappa: 0.55	Custom CNN showed potential on small datasets
(2)	Mutasa et al., 2020	U-Net + CNN; GAN & DRR augmentation	1,063 hip X-rays (augmented to 9063)	Binary Accuracy 92.3%, AUC 0.92; Three-class Accuracy 86%, AUC 0.96	Proposed for ER triage; flagged subtle fractures
(3)	Jiménez-Sánchez et al., 2020	ResNet-50 + AlexNet	1,347 hip X-rays	F1 87% (3-class), AUC 0.95; F1 94% (binary), AUC 0.98	AO classification; 100% correct region localization
(4)	Tanzi et al., 2020	Multi-stage CNN (InceptionV3)	2,453 proximal femur X-rays	3-class Accuracy 86%, 5-class Accuracy 81%, AUC 0.95	Clinician + AI improved accuracy by 14%
(5)	Qi et al., 2020	Faster R-CNN + FPN (ResNet-50)	2,333 femur X-rays	mAP 68.8%, Accuracy 71.5%, 80% in common classes	Outperformed junior surgeons in shaft fracture detection
(6)	Yoon et al., 2020	Faster R-CNN + Bayesian optimization	3,343 CT images	Binary Accuracy 97%, Multi-class Accuracy 90%, IoU 0.87	Automated fracture stability classification

(7)	Tanzi et al., 2022	Vision Transformer (ViT)	4,207 proximal femur X-rays	Accuracy 83%, Precision 0.77, Recall 0.76, F1 0.77	Self-attention improved localization
(8)	Li et al., 2020	Faster R-CNN with bounding box regression	350 axial CT images	Accuracy 91.42%, Sensitivity 91.55%, Precision 90.18%	Classified AO A1- A3 types effectively
(9)	Bae et al., 2021	ResNet-18 with CBAM attention	4,189 pelvic/hip X-rays	Internal AUC 0.999, Accuracy 98.6%; External AUC 0.977, Accuracy 97.1%	Validated across multiple hospitals
(10)	Nagra et al., 2024	Vision Transformer (ViT) with Grad-CAM	~7,000 high-resolution radiographs	Accuracy 92%, AUC 0.94	Improved interpretability with attention maps

Table 2: Ethical and Regulatory Considerations in Clinical Deployment

Aspect	*Description*
1. Data Privacy and Patient Consent	Ensure anonymization, secure storage, and informed consent in compliance with HIPAA, GDPR, and local data protection laws.
2. Bias, Fairness, and Generalizability	Using diverse datasets and external validation to avoid demographic or institutional bias also ensures equitable model performance.
3. Interpretability and Clinical Accountability	Use of explainable AI (e.g.- Grad-CAM) to improve transparency. Clinicians must retain final diagnostic responsibility.
4. Regulatory Approval and Standards Compliance	Get certification from regulatory agencies (FDA, EMA, CDSCO) and follow SaMD guidelines for safety post-deployment.
5. Ethical Use and Monitoring	Conduct regular audits, bias checks, and ethical reviews to maintain fairness, transparency, and safety post-deployment.

5 Conclusion

As artificial intelligence becomes more and more an essential component of clinical process, deep learning techniques have an important area of research in medical imaging. This review gives a comprehensive overview of deep learning methodologies implemented in X-ray imaging for bone fracture analysis. Researchers have looked at several machine learning models on public and institutional datasets, achieving excellent performance on multiple diagnostic tasks. This review adds to what we already know by putting together the latest advancements in the field, given us new ideas about how research is changing, and addressing how quickly AI technologies are moving forward. It highlights four main diagnostic tasks that are important for bone imaging. It describes a broad computational methodology for handling these tasks and identifies possible directions for future research on fracture detection and classification. Though deep learning models have shown diagnostic accuracy comparable to that of clinical professionals, it is still challenging to implement them in real-world clinical practice as people are worried about how transparent and trustworthy they are. To promote higher clinical adoption and utility of AI-driven diagnostic systems, future research is expected to improve model interpretability, integrate multimodal clinical data, provide data-driven therapy recommendations, and improved visualization techniques.

6 Acknowledgement

We sincerely acknowledge the STARS project (Project ID: STARS-2/2023-0552) for providing the funding and resources we needed to do this research.

References

1. A. Jiménez-Sánchez *et al.*, "Precise proximal femur fracture classification for interactive training and surgical planning," *Int J Comput Assist Radiol Surg*, vol. 15, no. 5, pp. 847–857, May 2020, doi: 10.1007/s11548-020-02150-x.

2. G. Marullo *et al.*, "Classification of AO/OTA 31A/B femur fractures in X-ray images using YOLOv8 and advanced data augmentation techniques," *Bone Rep*, vol. 22, Sep. 2024, doi: 10.1016/j.bonr.2024.101801.

3. L. Tanzi, E. Vezzetti, R. Moreno, A. Aprato, A. Audisio, and A. Massè, "Hierarchical fracture classification of proximal femur X-Ray images using a multistage Deep Learning approach," *Eur J Radiol*, vol. 133, Dec. 2020, doi: 10.1016/j.ejrad.2020.109373.

4. Y. Qi *et al.*, "Ground truth annotated femoral X-ray image dataset and object detection based method for fracture types classification," *IEEE Access*, vol. 8, pp. 189436–189444, 2020, doi: 10.1109/ACCESS.2020.3029039.

5. L. Tanzi, A. Audisio, G. Cirrincione, A. Aprato, and E. Vezzetti, "Vision Transformer for femur fracture classification," *Injury*, vol. 53, no. 7, pp. 2625–2634, Jul. 2022, doi: 10.1016/j.injury.2022.04.013.

6. P. Rajpurkar *et al.*, "MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs," May 2018, [Online]. Available: http://arxiv.org/abs/1712.06957

7. J. Bae *et al.*, "External Validation of Deep Learning Algorithm for Detecting and Visualizing Femoral Neck Fracture Including Displaced and Non-displaced Fracture on Plain X-ray," *J Digit Imaging*, vol. 34, no. 5, pp. 1099–1109, Oct. 2021, doi: 10.1007/s10278-021-00499-2.

8. S. J. Yoon, T. H. Kim, S. Bin Joo, and S. E. Oh, "Automatic multi-class intertrochanteric femur fracture detection from ct images based on ao/ota classification using faster r-cnn-bo method," *J Appl Biomed*, vol. 18, no. 4, pp. 97–105, 2020, doi: 10.32725/jab.2020.013.

9. S. Mutasa, S. Varada, A. Goel, T. T. Wong, and M. J. Rasiej, "Advanced Deep Learning Techniques Applied to Automated Femoral Neck Fracture Detection and Classification," *J Digit Imaging*, vol. 33, no. 5, pp. 1209–1217, Oct. 2020, doi: 10.1007/s10278-020-00364-8.

10. A. Alzaid, A. Wignall, S. Dogramadzi, H. Pandit, and S. Q. Xie, "Automatic detection and classification of peri-prosthetic femur fracture," *Int J Comput Assist Radiol Surg*, vol. 17, no. 4, pp. 649–660, Apr. 2022, doi: 10.1007/s11548-021-02552-5.

11. A. Jiménez-Sánchez *et al.*, "Curriculum learning for improved femur fracture classification: Scheduling data with prior knowledge and uncertainty," *Med Image Anal*, vol. 75, Jan. 2022, doi: 10.1016/j.media.2021.102273.

12. W. Gale, L. Oakden-Rayner, G. Carneiro, A. P. Bradley, and L. J. Palmer, "Detecting hip fractures with radiologist-level performance using deep neural networks," *Computer Vision and Pattern Recognition*, no. 06504, Nov. 2017, [Online]. Available: http://arxiv.org/abs/1711.06504

13. J. Bae *et al.*, "External Validation of Deep Learning Algorithm for Detecting and Visualizing Femoral Neck Fracture Including Displaced and Non-displaced Fracture on Plain X-ray," *J Digit Imaging*, vol. 34, no. 5, pp. 1099–1109, Oct. 2021, doi: 10.1007/s10278-021-00499-2.

14. A. Berrajaa, M. Merras, and I. Berrajaa, "Advanced CNN based on genetic algorithm to automated femoral neck fracture classification," *Signal Image Video Process*, vol. 18, no. 6–7, pp. 5229–5238, Aug. 2024, doi: 10.1007/s11760-024-03228-8.