



The International Journal for Engineering and
Modern Science

Journal homepage: <https://modernacademy-journal.synceg.net>



Article

Long Bone Fracture Detection Using Detection Transformers

Soaad Nasser Eldin Ali Kadry * 

Department of Computer Engineering & Information Technology, Modern Academy for Engineering and Technology, Elmokattam, Cairo 11439, Egypt; soaadnassereldin@gmail.com

ARTICLE INFO.

Article history:

Received 12 November 2025

Revised 30 November 2025

Accepted 5 January 2026

Available online 11 January 2026

Keywords:

Radiology
Bone Fracture
X-rays
Deep learning
Image detection
Transformers Model

ABSTRACT

Accurate diagnosis of long bone fractures is crucial to prevent complications such as delayed healing or permanent deformities. In medical imaging, machine learning (ML), particularly deep learning (DL) models, have emerged as powerful tools for enhancing diagnostic accuracy, reducing human error, and streamlining workflows through automation. This study investigates the Detection Transformer (DETR), a transformer-based DL model, for automated fracture analysis in X-ray images. A dataset of 3,000 long bone radiographs, manually labeled and annotated by an expert, was preprocessed and augmented to support robust model training and evaluation. DETR achieved a mean Average Precision (mAP) of 80.9%, demonstrating strong sensitivity and reliable localization of fracture regions. Its end-to-end architecture and ability to capture global contextual dependencies make it particularly effective in identifying subtle or complex fracture patterns. These results highlight DETR's potential as a clinically relevant tool, with future improvements anticipated through larger datasets, advanced augmentation strategies, and architectural refinements.



© 2026 Modern Academy Ltd. All rights reserved

1. Introduction

Medical imaging has become an indispensable component of modern healthcare, enabling clinicians to visualize internal structures and diagnose a wide range of conditions with

* Corresponding author

E-mail address: soaadnassereldin@gmail.com

greater precision and confidence. Among the available modalities, X-rays remain the most widely used due to their affordability, accessibility, and ability to capture fine skeletal details in a relatively short time [1]. Although MRI and ultrasound continue to attract research interest, their higher cost and limited availability constrain widespread clinical use. Moreover, developing diagnostic models that generalize across modalities and anatomical regions remains particularly difficult, as this requires large, well-annotated datasets for each imaging type and bone category—resources that are rarely available in practice. For this reason, the present study focuses on X-ray imaging, the most practical and clinically relevant modality for bone fracture classification and detection.

The human skeleton consists of several categories of bones, including long, short, flat, and irregular bones, each with distinct structural and functional roles. Among these, long bones such as the femur, tibia, fibula, humerus, radius, ulna, metatarsals, and phalanges are particularly prone to fractures [2]. Their essential role in weight-bearing and mobility means that injuries often have serious consequences, impairing patient independence and quality of life. Failure to diagnose such fractures accurately or promptly can result in delayed healing, functional impairment, or long-term disability. A large-scale global study across 204 countries further underscores their significance, identifying long bone fractures as the most common skeletal injury worldwide [3].

The emergence of artificial intelligence (AI) has brought a paradigm shift to medical imaging, offering powerful tools for enhancing diagnostic accuracy, reducing human error, and supporting clinical decision-making. Machine learning (ML) algorithms have been applied to detect, classify, and segment medical images across diverse applications [4-6]. Within this domain, deep learning (DL) techniques have demonstrated exceptional success due to their ability to learn hierarchical features directly from imaging data. Transfer learning and pretrained convolutional models have further accelerated progress by allowing researchers to adapt existing architectures to limited medical datasets, improving model generalization and robustness [7].

Recent reviews have discussed various medical image identification approaches, highlighting the integration of ML and DL for enhanced decision support and data-driven clinical workflows [8]. Moreover, AI-based medical image processing and analysis frameworks continue to evolve, addressing challenges in image quality, computational efficiency, and clinical interpretability [9].

AI methods have also shown broad applicability across diagnostic disciplines. For example, CNN- and EfficientNet-based systems have achieved high accuracy in breast cancer detection [10], while handcrafted CNN architectures have been used for brain tumor segmentation with strong precision and recall [11]. Similarly, EfficientNet variants have demonstrated exceptional performance in COVID-19 detection from chest radiographs [12]. In dental imaging, studies comparing object detection architectures such as Faster R-CNN, YOLO, and SSD have demonstrated the adaptability of DL for anatomic localization and angle estimation tasks [13]. These examples collectively emphasize the potential of AI to streamline radiological workflows, enhance diagnostic reliability, and improve early disease detection.

Within the orthopedic field, X-ray imaging remains one of the most accessible and cost-effective diagnostic modalities. Nevertheless, subtle fracture lines, overlapping structures, and variability in image quality often lead to diagnostic uncertainty, especially in busy clinical settings [14]. The development of large annotated datasets, such as MURA, has further enabled DL research in musculoskeletal imaging by providing a benchmark for automated abnormality detection and classification [15]. Reviews of DL-based fracture detection approaches have confirmed their growing role in orthopedic diagnostics, noting improvements in sensitivity and specificity compared with conventional manual interpretation [16].

Early research employing Convolutional Neural Networks (CNNs) and feature pyramid networks (FPNs) demonstrated promising results in bone fracture detection. Studies leveraging

CNN architectures have effectively detected fractures in the thigh and arm regions, revealing their capacity to extract discriminative spatial features from X-ray images [17-20]. These methods improved classification and localization performance, setting the stage for more advanced DL models capable of handling complex fracture morphologies.

Subsequent developments introduced object detection frameworks such as Faster R-CNN, YOLO, and guided anchoring methods, which significantly enhanced detection precision and inference speed. These models achieved notable success in identifying fractures across femoral, spinal, and wrist radiographs, demonstrating their robustness in localizing small or irregular patterns [21-24]. Despite these improvements, CNN-based models still face limitations in capturing long-range dependencies, making them less effective when fractures appear subtle or distributed across large anatomical regions.

To address these challenges, transformer-based architectures have emerged as a powerful alternative. Foundation models incorporating self-attention mechanisms have reshaped medical image analysis, enabling models to capture global contextual information more effectively than traditional convolutional networks [25]. The Vision Transformer (ViT) represented a key milestone, proving that pure transformer architectures could achieve or exceed CNN performance on large-scale image classification tasks [26]. More recent innovations, such as the Contextual Transformer Network, have further refined attention mechanisms to improve feature extraction and visual recognition in complex imaging contexts [27]. These advancements have expanded to medical imaging applications, where transformers are now employed for classification, segmentation, and detection tasks across modalities, including X-rays, CT, and MRI.

Building upon this evolution, the present study investigates the DETection TRansformer (DETR), a transformer-based deep learning model, for long bone fracture classification and localization in X-ray images [28-31]. DETR's self-attention mechanism allows the model to capture both global and local relationships across the entire image, enabling the detection of subtle, overlapping, or irregular fracture patterns that may be missed by region-based detectors. Unlike traditional object detection frameworks, DETR operates as a fully end-to-end model, eliminating the need for handcrafted components such as anchor boxes and non-maximum suppression. Its set-based prediction with Hungarian matching further improves the accuracy of detecting multiple fractures in a single image and reduces duplicate detections. These characteristics make DETR particularly well-suited for orthopaedic imaging, where fractures can vary in size, shape, and location. Through this approach, the current work aims to evaluate DETR's diagnostic performance on annotated long bone datasets, assessing its potential as a clinically relevant alternative to convolutional and region-based detectors for automated orthopedic imaging.

2. Materials and Methods

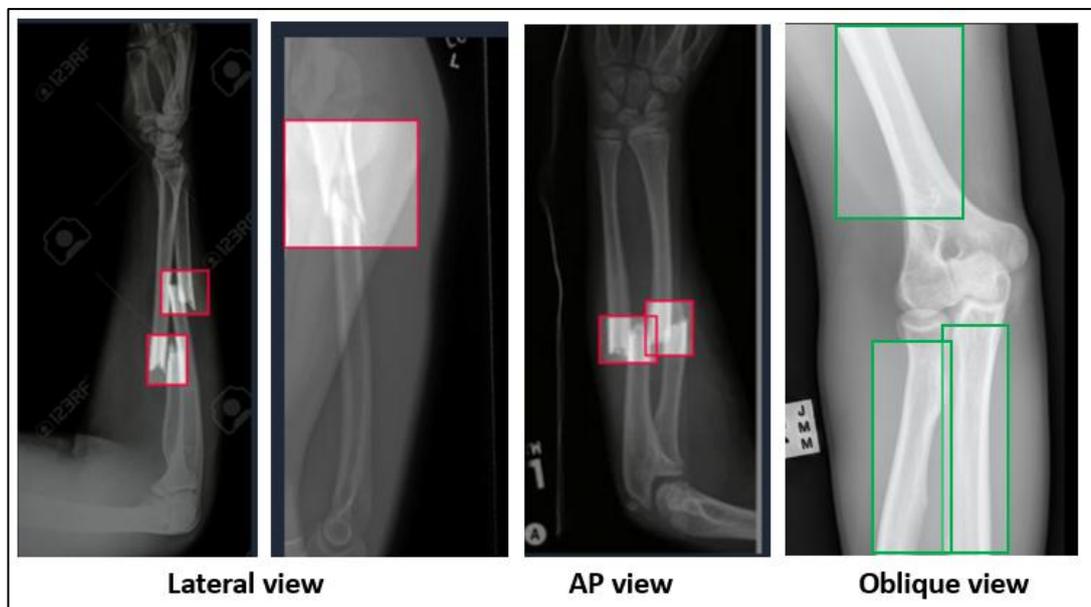
2.1. Dataset

This study employed a dataset of 3,000 long bone X-ray images collected from multiple sources. A large portion was obtained from the MURA dataset, which originally included 1,014 studies of the humerus, 1,080 studies of the forearm (radius/ulna), and 4,384 studies of the wrist. These studies were labeled as normal or abnormal (e.g., fractures, tumors, wires) [15]. After filtering to retain only normal or fractured bones, a total of 2,339 normal and 275 fractured images were selected. The remaining long bone X-ray images were collected from other publicly available sources, as summarized in Table 1.

Table 1. Number of X-ray images used in the dataset, with their sources

No of images	type	Bone name	Source
2339	normal	Humerus/radius/ulna	https://stanfordmlgroup.github.io/competitions/mura/
82	normal	Femur/tibia/fibula/Humerus/radius/ulna	http://bones.getthediagnosis.org/
275	fracture	Humerus/radius/ulna	https://stanfordmlgroup.github.io/competitions/mura/
304	fracture	Femur/tibia/fibula/phalanges	https://www.kaggle.com/datasets/orvile/human-bone-fractures-image-dataset-hbfmid

As shown in Figure 1, the dataset encompassed a variety of anatomical regions of long bones, including the femur, tibia, fibula, humerus, radius, and ulna, and incorporated multiple radiographic views such as anteroposterior, lateral, and oblique. Each image was manually annotated by trained experts to provide ground truth labels for both binary classification (fractured vs. normal) and localization (bounding box annotations). This ensured that the dataset was suitable for supervised deep learning tasks.

**Figure 1.** Samples of the dataset.

The diversity of bones, imaging conditions, and projection angles included in the dataset was intended to reflect the variability encountered in real-world clinical settings. Such diversity enhances the robustness and generalizability of the trained models, especially for subtle or ambiguous fracture presentations.

2.2. Image Pre-processing and Augmentation

All images underwent a series of preprocessing steps to standardize inputs and improve diagnostic quality. Manual cropping was performed to isolate long bone regions, removing irrelevant background structures to reduce noise and improve feature focus. Images stored in RGB format were converted to grayscale, simplifying feature representation and reducing computational redundancy.

To ensure uniform input dimensions, all images were resized to 512×512 pixels, balancing computational efficiency with the preservation of subtle fracture details. Additional enhancements, including contrast adjustment and sharpening, improved edge clarity and structural visibility, ensuring consistent quality across the dataset.

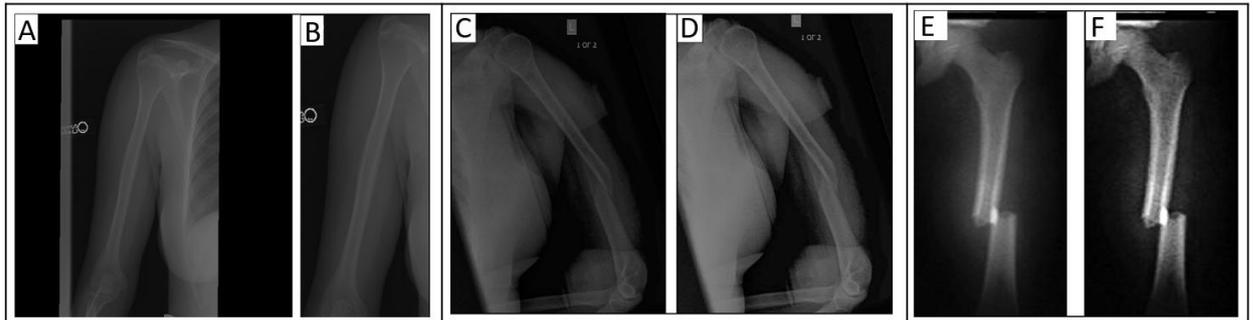


Figure 2. Preprocessing examples of long bone X-rays: (A, B) before and after cropping, (C, D) before and after contrast adjustment, and (E, F) before and after sharpening.

To further increase dataset diversity and improve model generalization, data augmentation was applied, expanding the dataset from 3,000 to 6,000 images. Transformations included horizontal and vertical flipping, brightness adjustments ($\pm 5\%$), and random rotations within $\pm 15^\circ$, simulating natural variability in X-ray acquisition while preserving diagnostic integrity. All augmentations were implemented using the Albumentations library, which is widely adopted for efficient medical imaging preprocessing.

3. Model Architecture

This study employed the DEtection TRansformer (DETR) as the primary architecture for long bone fracture detection and localization. DETR is a cutting-edge deep learning framework that introduces a fully end-to-end design, eliminating the need for manually defined components such as anchor boxes and non-maximum suppression (NMS) [32]. Unlike traditional region-based detectors, DETR formulates object detection as a direct set prediction problem, simplifying the pipeline and reducing hyperparameter dependence.

As illustrated in Figure 3, DETR operates in three stages:

1. **Backbone (ResNet-50):** Extracts both low-level and high-level spatial features from X-ray inputs, capturing structural bone patterns relevant for fracture detection.
2. **Transformer Encoder–Decoder:** The encoder leverages multi-head self-attention to model long-range dependencies across the image. The decoder introduces a fixed number of learned object queries that interact with the encoded features through cross-attention, allowing simultaneous prediction of class labels (fractured vs. normal) and bounding box coordinates.
3. **Prediction Heads:** Feed-forward networks output bounding boxes and class probabilities for each object query. A bipartite matching strategy (Hungarian algorithm) ensures a one-to-one alignment between predictions and ground-truth annotations.

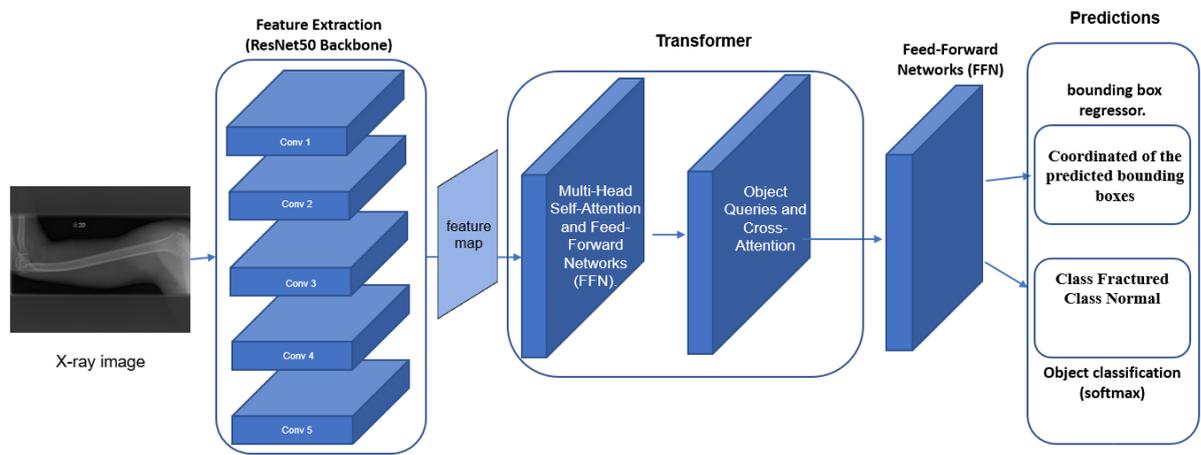


Figure 3. DETR model architecture used.

The advantages of DETR are particularly relevant for medical imaging. By modeling global spatial context, DETR improves fracture detection in challenging cases such as subtle hairline fractures, irregular fracture patterns, and overlapping anatomical structures. The self-attention mechanism enhances detection accuracy in cluttered radiographs where local features alone may be insufficient. Moreover, DETR eliminates complex post-processing, making the workflow simpler and more interpretable.

However, DETR also presents challenges. It typically requires longer training times and larger datasets to achieve convergence, and performance can be dataset-sensitive. Despite these limitations, DETR's architectural elegance and robustness highlight its potential as a powerful alternative to conventional region-based detectors, especially in tasks demanding high localization precision, such as fracture detection in radiographs.

4. Training and Evaluation

The DETR model was trained on the augmented dataset of 6,000 X-ray images, divided into training (70%), validation (20%), and testing (10%) subsets. Training employed the Adam optimizer with an initial learning rate of $1e-4$, along with a dynamic scheduler that reduced the rate when validation performance plateaued. To prevent overfitting, early stopping with a patience of 20 epochs was applied. Hyperparameters such as batch size, weight decay, and dropout rate were tuned experimentally to ensure stable convergence.

The loss function combined three components:

- **Cross-entropy loss** for binary classification (fractured vs. normal),
- **L1 regression loss** for bounding box localization, and
- **Generalized Intersection over Union (IoU) loss** to encourage precise spatial alignment.

Model performance was evaluated using several complementary metrics. The primary benchmark was mean Average Precision (mAP) at $\text{IoU} = 0.5$, reflecting the overlap between predicted and ground-truth bounding boxes. Along with the confusion matrix analysis, which was performed to visualize the distribution of true positives, true negatives, false positives, and false negatives. This analysis provided valuable insight into the clinical trade-off between minimizing missed fractures and limiting overdiagnosis.

5. Result and discussion

5.1. Performance of DETR

The DETR model was trained and evaluated on the long bone X-ray dataset, achieving a mAP of 80.9% at IoU = 0.5 with a total loss of 2.52. This performance indicates that DETR is highly competitive for fracture detection despite its end-to-end design, which eliminates handcrafted components such as anchor boxes and non-maximum suppression.

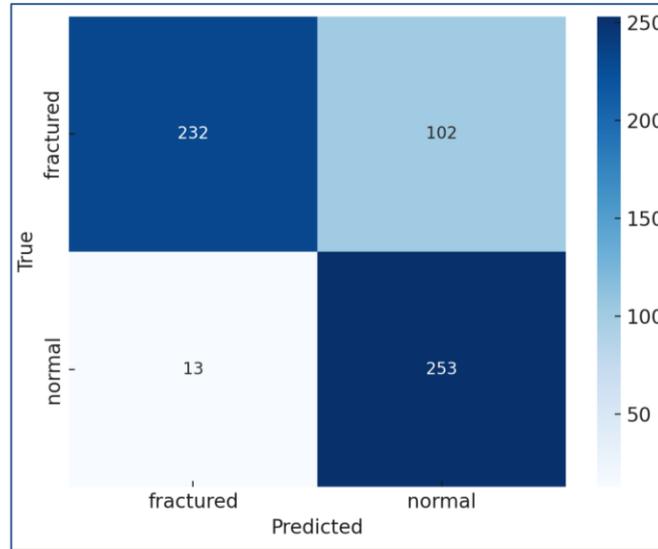


Figure 4. The confusion matrix of the DETR model.

The confusion matrix presented in Figure 4 offers a detailed assessment of DETR’s diagnostic performance. The model accurately classified 232 fractured and 253 normal radiographs, resulting in an overall accuracy of 80.8%. DETR demonstrated high sensitivity (94.7%), successfully identifying nearly all true fracture cases, which is critical for clinical decision-making. Although approximately 29% of normal cases were incorrectly flagged as fractured, this tendency toward conservative over-detection reflects a model bias favouring patient safety—prioritizing the minimization of missed fractures over the risk of additional follow-up imaging. Such a trade-off is clinically acceptable in early diagnostic settings, where sensitivity is paramount, and further refinement with expanded datasets could enhance specificity without compromising detection reliability.

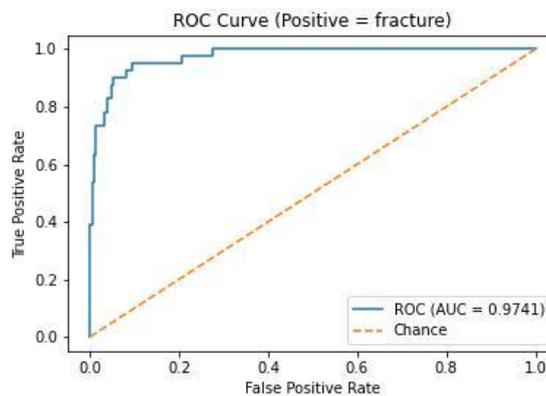


Figure 5. Receiver Operating Characteristic (ROC) curve for the DETR model in fracture detection.

To further evaluate the model's discriminatory power, a Receiver Operating Characteristic (ROC) analysis was conducted shown in Figure 5. The DETR model achieved an Area Under the Curve (AUC) of 0.9741, demonstrating an excellent capacity to differentiate between fractured and non-fractured long bones. This high AUC value provides a more comprehensive measure of performance than accuracy alone, as it accounts for the model's behavior across all classification thresholds.

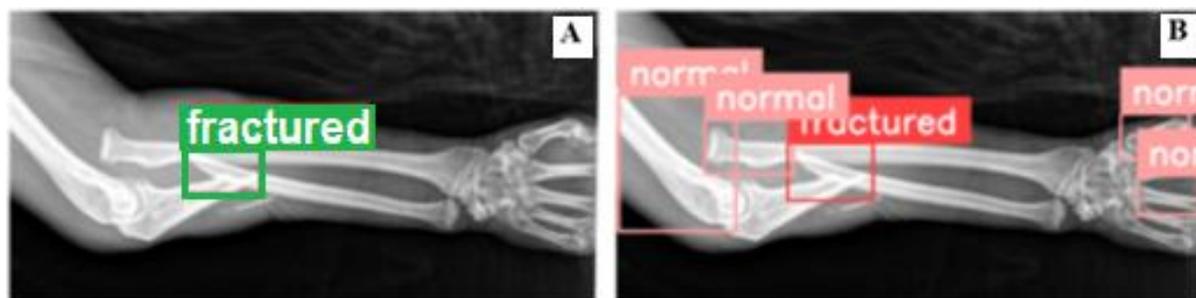


Figure 6. Model predictions : (A) ground truth, (B) predicted boxes.

Figure 6 illustrates examples of the DETR model's predictions, where the ground truth bounding box is shown on the left side and the corresponding predicted bounding box is displayed on the right.

5.2. Comparison with Previous Studies

Table 2 summarizes DETR's results in relation to previously reported detection models applied to musculoskeletal X-rays. Earlier studies using region-based detectors such as Faster R-CNN reported mAP values ranging from 68.8% on femoral fractures [20] to 79.8% on custom datasets [34]. Variants like GA_Faster R-CNN achieved 70% on hand fractures [23], while single-stage models such as YOLOv8 reported 77% on larger wrist datasets [33]. Heterogeneous datasets like MURA produced lower results, with an mAP of 62% using FPN [21].

Table 2. The comparison between Detection models

Ref	Dataset / Image Type	Task Type	Model	No. of Images	Performance
[20]	Femoral X-ray(10 classes)		Faster R-CNN	2333	mAP = 68.8%
[21]	MURA X-ray(2 Classes)		FPN	4000	mAP = 62%
[23]	Hand X-ray(2 Classes)	Classification	GA_Faster R-CNN	3067	mAP = 70%
[33]	Wrist X-ray (10 classes)	+ Detection	YOLOv8	10308	mAP = 77%
[34]	Custom X-ray(4 classes)		Faster R-CNN	3000	mAP = 79.8%
This study	Custom X-ray(2 Classes)		DETR	3000	mAP = 80.9%

In comparison, the proposed DETR model attained an mAP of 80.9% on a two-class custom X-ray dataset, slightly outperforming Faster R-CNN under identical conditions [34]. These results indicate that DETR delivers competitive or superior detection accuracy while offering the advantages of an end-to-end transformer architecture that eliminates handcrafted components such as anchor boxes and non-maximum suppression. DETR's capacity to model long-range spatial dependencies further enhances its robustness in identifying subtle or overlapping fracture patterns. Overall, these findings position DETR as a strong and scalable alternative to conventional convolution-based detectors for automated fracture localization and classification in radiographic imaging.

6. Conclusion and Future Work

This study evaluated the DETection TRansformer (DETR) for automated detection of long bone fractures in X-ray images. DETR achieved a mAP of 80.9% with strong sensitivity, correctly identifying nearly all fractures and minimizing false negatives. While the model showed a higher false positive rate, this is clinically preferable to missed fractures, underscoring DETR's value for patient safety.

The results highlight DETR's strength as an end-to-end transformer-based detector, capable of modeling global dependencies and handling subtle or complex fracture patterns. However, performance remains limited by dataset size and variability. Expanding the dataset, applying advanced augmentation techniques, and incorporating interpretability tools like Grad-CAM could further improve robustness and clinical trust.

Future work should explore transformer variants such as ViT or Swin Transformers and hybrid CNN-transformer approaches, as well as developing an integrated diagnostic pipeline that combines classification and detection within a user-friendly system. With continued optimization, DETR and related architectures hold strong potential to advance automated musculoskeletal imaging in real-world clinical practice.

ORCID iDs

Soaad Nasser Eldin  <https://orcid.org/0009-0001-4482-8868>

Funding: This research received no external funding.

Data Availability Statement: This study is based on the publicly available MURA (Musculoskeletal Radiographs) dataset, provided by the Stanford ML Group and accessible at <https://stanfordmlgroup.github.io/competitions/mura/>. For this research, the original MURA images were filtered, annotated, and preprocessed to focus on long bone fracture detection using the Roboflow platform. In accordance with the licensing terms of the original MURA dataset, the processed and annotated images cannot be publicly redistributed. However, details of the preprocessing and annotation pipeline are available from the authors upon reasonable request.

Acknowledgments: I would like to express my sincere appreciation to my supervisor, Prof. Ashraf El Marakby, for his invaluable guidance, expertise, and continuous support throughout this work. I also extend my gratitude to my co-supervisor, Prof. Hala Maghraby, for her constructive support and for providing essential medical insights that significantly contributed to this study. I am also thankful to my colleagues and friends for their support and encouragement. This work was made possible through the contributions of many individuals, and I sincerely acknowledge their efforts.

Conflicts of Interest: The author declares no conflicts of interest.

References

- [1] Ou, X.; et al. Recent Development in X-Ray Imaging Technology: Future and Challenges. *Res.* 2021, 2021, 1–14. doi:10.34133/2021/9892152.
- [2] Wu, A.M.; et al. Global, Regional, and National Burden of Bone Fractures in 204 Countries and Territories, 1990–2019: A Systematic Analysis from the Global Burden of Disease Study 2019. *Lancet Healthy Longev.* 2021, 2(9), 1–12. doi:10.1016/S2666-7568(21)00172-0.
- [3] Meinberg, E.G.; Agel, J.; Roberts, C.S.; Karam, M.D.; Kellam, J.F. Fracture and Dislocation Classification Compendium–2018. *J. Orthop. Trauma* 2018, 32, 1–10. doi:10.1097/BOT.0000000000001063.
- [4] Christopher, I. Machine Learning: A Review. *Semicond. Sci. Inf. Devices* 2020, 2, 1–8. doi:10.30564/ssid.v2i2.1931.
- [5] Panayides, A.S.; et al. AI in Medical Imaging Informatics: Current Challenges and Future Directions. *IEEE J. Biomed. Health Inform.* 2020, 1–10. doi:10.1109/JBHI.2020.2991043.
- [6] Adugna, T.D.; Ramu, A.; Haldorai, A. A Review of Pattern Recognition and Machine Learning. *J. Mod. Comput.* 2024, 4(2), 1–9. doi:10.53759/7669/jmc202404020.
- [7] Kim, H.E.; Cosa-Linan, A.; Santhanam, N.; Jannesari, M.; Maros, M.E.; Ganslandt, T. Transfer Learning for Medical Image Classification: A Literature Review. *BMC Med. Imaging* 2022, 22, 1–20. doi:10.1186/s12880-022-00793-7.
- [8] Li, J.; Jiang, P.; An, Q.; Wang, G.G.; Kong, H.F. Medical Image Identification Methods: A Review. *Comput. Biol. Med.* 2024, 170, 107777. doi:10.1016/j.compbiomed.2023.107777.
- [9] Alghazo, J.; Latif, G. AI/ML-Based Medical Image Processing and Analysis. *Diagnostics* 2023, 13, 3671. doi:10.3390/diagnostics13243671.
- [10] Khalid, A.; et al. Breast Cancer Detection and Prevention Using Machine Learning. *Diagnostics* 2023, 13(19), 3113. doi:10.3390/diagnostics13193113.
- [11] Ullah, F.; et al. Brain Tumor Segmentation from MRI Images Using Handcrafted Convolutional Neural Network. *Diagnostics* 2023, 13(16), 2650. doi:10.3390/diagnostics13162650.
- [12] Kurt, Z.; Işık, Ş.; Kaya, Z.; Anagün, Y.; Koca, N.; Çiçek, S. Evaluation of EfficientNet Models for COVID-19 Detection Using Lung Parenchyma. *Neural Comput. Appl.* 2023, 35(16), 1–11. doi:10.1007/s00521-023-08344-z.
- [13] Vilcapoma, P.; et al. Comparison of Faster R-CNN, YOLO, and SSD for Third Molar Angle Detection in Dental Panoramic X-rays. *Sensors* 2024, 24(18), 6053. doi:10.3390/s24186053.
- [14] Pinto, A.; et al. Traumatic Fractures in Adults: Missed Diagnosis on Plain Radiographs in the Emergency Department. *Ann. Biomed.* 2018, 89(1-S), 7015. doi:10.23750/abm.v89i1-S.7015.
- [15] Rajpurkar, P.; et al. MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. Available online: <https://arxiv.org/abs/1712.06957>.
- [16] Kalmet, P.H.S.; et al. Deep Learning in Fracture Detection: A Narrative Review. *Acta Orthop.* 2020, 91(2), 1–8. doi:10.1080/17453674.2019.1711323.
- [17] Yadav, D.P.; Rathor, S. Bone Fracture Detection and Classification Using Deep Learning Approach. In Proceedings of the 2020 International Conference on Power Electronics and IoT Applications in Renewable Energy and Its Control (PARC 2020), Mathura, India, 28–29 February 2020. doi:10.1109/PARC49193.2020.236611.

- [18] Chung, S.W.; et al. Automated Detection and Classification of the Proximal Humerus Fracture by Using Deep Learning Algorithm. *Acta Orthop.* 2018, 89(4), 1–8. doi:10.1080/17453674.2018.1453714.
- [19] Guan, B.; Yao, J.; Zhang, G.; Wang, X. Thigh Fracture Detection Using Deep Learning Method Based on New Dilated Convolutional Feature Pyramid Network. *Pattern Recognit. Lett.* 2019, 125, 1–7. doi:10.1016/j.patrec.2019.06.015.
- [20] Guan, B.; Zhang, G.; Yao, J.; Wang, X.; Wang, M. Arm Fracture Detection in X-Rays Based on Improved Deep Convolutional Neural Network. *Comput. Electr. Eng.* 2020, 81, 106530. doi:10.1016/j.compeleceng.2019.106530.
- [21] Qi, Y.; et al. Ground Truth Annotated Femoral X-Ray Image Dataset and Object Detection Based Method for Fracture Types Classification. *IEEE Access* 2020, 8, 1–10. doi:10.1109/ACCESS.2020.3029039.
- [22] Sha, G.; Wu, J.; Yu, B. Detection of Spinal Fracture Lesions Based on Improved YOLOv2. In *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA 2020)*, Dalian, China, 27–29 June 2020. doi:10.1109/ICAICA50127.2020.9182582.
- [23] Xue, L.; et al. Detection and Localization of Hand Fractures Based on GA_Faster R-CNN. *Alex. Eng. J.* 2021, 60(5), 4513–4524. doi:10.1016/j.aej.2021.03.005.
- [24] Hardalaç, F.; et al. Fracture Detection in Wrist X-ray Images Using Deep Learning-Based Object Detection Models. *Sensors* 2022, 22(3), 31285. doi:10.3390/s22031285.
- [25] Zhang, S.; Metaxas, D. On the Challenges and Perspectives of Foundation Models for Medical Image Analysis. *Med. Image Anal.* 2024, 91, 102996. doi:10.1016/j.media.2023.102996.
- [26] Tanzi, L.; Audisio, A.; Cirrincione, G.; Aprato, A.; Vezzetti, E. Vision Transformers for Femur Fracture Classification. *arXiv* 2021, abs/2108.03414. Available online: <https://arxiv.org/abs/2108.03414>
- [27] Li, Y.; Yao, T.; Pan, Y.; Mei, T. Contextual Transformer Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2023, 45(2), 1077–1092. doi:10.1109/TPAMI.2022.3164083.
- [28] Wang, Y.; Lee, J.; Alam, N.; Patel, A. FractuVision: Enhancing Bone Fracture Diagnostics. *Cornell Univ. Libr.* 2024, Preprint.
- [29] Ruhi, S.S.F.; Nahar, F.; Ashrafi, A.F. A Novel Approach Towards the Classification of Bone Fracture from Musculoskeletal Radiography Images Using Attention Based Transfer Learning. *arXiv* 2024, abs/2410.14833. Available online: <https://arxiv.org/abs/2410.14833>
- [30] Islam, M.; Fathima, A.; Ghosh, D. An Improved Technique for Bone Fracture Classification Using Visual Transformers and Convolutional Neural Networks. *SSRN Electron. J.* 2025, Jul., 1–9. doi:10.2139/ssrn.5191555.
- [31] González, G.; et al. Classification and Segmentation of Hip Fractures in X-rays: Highlighting Fracture Regions for Interpretable Diagnosis. *Insights Imaging* 2025, 16(1), 86. doi:10.1186/s13244-025-01958-y.
- [32] Ickler, M.K.; Baumgartner, M.; Roy, S.; Wald, T.; Maier-Hein, K.H. Taming Detection Transformers for Medical Object Detection. In *Proceedings of Informatik Aktuell 2023*, Berlin, Germany, 2023. doi:10.1007/978-3-658-41657-7_39.
- [33] Ahmed, A.; Imran, A.S.; Manaf, A.; Kastrati, Z.; Daudpota, S.M. Enhancing Wrist Abnormality Detection with YOLO: Analysis of State-of-the-Art Single-Stage Detection Models. *Biomed. Signal Process. Control* 2024, 93, 106144. doi:10.1016/j.bspc.2024.106144.

- [34] Soaad, N.; Hala, M.; Sabry, M.; Ashraf, E.M. Long Bones X-Ray Fracture Classification Using Machine Learning. *J. Al-Azhar Univ. Eng. Sect.* 2024, *19(72)*, 121–133. doi:10.21608/aej.2024.259630.1577.