Enhanced Bone Fracture Detection and Quantification in X-Ray Images Using Deep Learning

Aman Kshetri[a], Raj Sah Rauniyar[b] and S S Chakravarthi[c]
Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India

Keywords: Bone Fracture Detection, X-Ray Imaging, YOLO, Radiologists, Deep Learning.

Abstract: Bone fracture detection in X-ray imaging is an essential diagnostic process, yet it often requires specialized expertise that may be limited in under-resourced healthcare settings. In major hospitals, experienced radiologists typically interpret X-rays with high accuracy. However, in smaller facilities within underdeveloped regions, less experienced medical personnel may struggle to provide accurate readings, leading to a significant rate of misinterpretation, currently reported at 26%. While numerous studies have focused on localizing fractures, few address the need for quantifying the length of the fractured bone segment, a critical factor in treatment planning. This project aims to develop an advanced deep learning model using the YOLO architecture to enhance bone fracture detection and quan-tification in X-ray images. By automating fracture detection and accurately measuring fracture length, the YOLO-based model will improve diagnostic accuracy, reduce radiologist workload, and ensure consistent assessments across diverse healthcare environments. The objectives include designing robust algorithms for fracture localization and length measurement, achieving high precision in fracture detection, and validating the model against a comprehensive X-ray dataset. Ultimately, this tool is expected to provide valuable diagnostic aid, particularly in settings with limited radiological resources, improving patient outcomes through reliable, automated fracture analysis.

1 INTRODUCTION

In hospital emergency rooms, radiologists frequently examine patients having fractures in various body parts like the wrist, arm, or leg. Fractures, which are disruptions in bone continuity, are typically classified into two categories: open and closed. Open fractures involve the bone piercing the skin, while closed fractures occur when the bone is broken without breaching the skin's surface. Accurate identification and classification of fractures are essential for proper treatment planning, often necessitating surgery. Prior to surgical intervention, surgeons must thoroughly assess a patient's medical history and conduct a comprehensive checkup & understand the fracture's complexity. In the latest medical image advancements, three primary modalities are widely used to diagnose fractures: X-ray, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT). Among these, X-ray imaging remains the most commonly utilized method due to its cost-effectiveness, availability, and speed, making it a primary diagnostic tool, especially in emergency settings. X-ray imaging plays a crucial role in diagnosing fractures, such as distal radius and ulna fractures, which are prevalent in pediatric patients and account for a significant portion of wrist trauma cases.

With the rapid advancements in deep learning, neural network-based models have emerged as powerful tools for medical image processing. Deep learning's capacity to analyze complex data patterns makes it particularly suitable for ap- plications like fracture detection, a growing research focus within the field of computer vision. Object detection models, a subdomain of deep learning, have shown promising results in fracture detection, enabling real-time identification and localization of fractures within medical images. Deep learning object detection techniques are generally categorized into two-stage and one-stage algorithms. Two-stage algorithms, such as Region-based Convolutional Neural Networks (R- CNN) and their advanced iterations, generate both location and class probabilities through a sequential two-stage process. This results in highly accurate outcomes but often re-

[a] https://orcid.org/0009-0000-2037-522X
[b] https://orcid.org/0009-0002-5887-0132
[c] https://orcid.org/0000-0002-8373-7264

Kshetri, A., Sah Rauniyar, R. and S Chakravarthi, S.
Enhanced Bone Fracture Detection and Quantification in X-Ray Images Using Deep Learning.
DOI: 10.5220/0013586900004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 2*, pages 80-86
ISBN: 978-989-758-763-4

quires extended processing times, making them less ideal for real-time applications.

In this project, we employ the YOLOv8 model, the latest iteration of the YOLO series, to advance bone fracture detection in X-ray images. Our approach is designed to address two primary goals: identifying the location of the fracture and quantifying the fracture length, an aspect rarely explored in existing research. By training the YOLOv8 model on a diverse dataset, we aim to develop a robust, an efficient solution for fracture detection that can be deployed across various healthcare settings, from well-resourced hospitals to under-resourced clinics. We further enhance model performance through data augmentation techniques, optimizing the YOLOv8 algorithm for pediatric wrist fractures.

Through experimental comparison, we assess the YOLOv8 model's perfor- mance against YOLOv7 and its improved variants, using mean average precision (mAP 50) as the evaluation metric. Our findings demonstrate that YOLOv8, when trained with tailored data augmentation strategies, achieves the highest mAP 50 score, underscoring its efficacy in accurately detecting and quantifying fractures. This project, by automating fracture detection and measurement, has the poten- tial to alleviate radiologists' workloads, ensure consistent diagnostic outcomes, and improve patient care across diverse healthcare settings, particularly in areas where access to radiology expertise is limited.

2 RELATED WORKS

This field has experienced significant growth, particularly in leveraging deep learning techniques to enhance medical image analysis. Recent advancements focus on improving the accuracy and reliability of bone fracture detection and quantification in X-ray images. Much of the earlier research serves as a foundation for developing modern approaches, such as the method presented in this study, which helps in delivering enhanced diagnostic performance.

(A. Saad, 2023) developed a convolutional neural network (CNN) using Keras to detect fractures in X-ray images. The model was trained and augmented on a dataset of 9,103 X-ray images, to improve the diversity and robustness of the training. The CNN model achieved a high accuracy of 91%, with precision and recall rates of 89.5% and 87%, respectively, largely due to the data augmentation. While this accuracy places it above several other methods, the study notes a risk of false positives, suggesting further refinements to make it suitable for clinical settings. (Kalb and Harris, 2021) The dataset consisted of X-rays classified as fractured and non-fractured, enhanced through augmentation techniques. The results are promising, with the model showing significant accuracy; however, the research highlights the need for comparisons with other models to ensure consistency and reduce the rate of false positives.

(Zou and Arshad, 2022) explored the performance of YOLO variants and two-stage models for fracture detection, emphasizing Enhanced Intersection over Union (EIoU) to improve bounding box precision. The study found that the YOLOv7-ATT model achieved a mean average precision (mAP) of 80.2% and 86.2% on the FracAtlas dataset, outperforming other models in terms of precision and recall. (M. Salimi, 2022) While YOLOv7-ATT stood out, the research also revealed that other two-stage models and SSD performed suboptimally, and additional enhancements are still needed for further accuracy improvements. (T. Gruber, 2022) The dataset included annotated images representing four types of fractures and was evaluated using precision, recall, mAP, and IoU. Overall, the YOLOv7-ATT model demonstrated that single-stage models generally surpass two-stage models in terms of both speed and detection accuracy.

(J. Li, 2021) employed DenseNet-201, a deep learning model that was trained on 1,370 X-ray images, with preprocessing and data augmentation methods applied to enhance its accuracy. The model's performance was measured by metrics like accuracy, sensitivity, AUC and specificity, where it achieved 94.1% accuracy and an AUC of 98.7%. The model also demonstrated high sensitivity and specificity rates, with sensitivity at 93.2% and specificity at 94.8%. (M. Oppenheimer, 2021) However, the study notes that further clinical validation is necessary to ensure its reliability for widespread clinical use. The dataset focused on pediatric elbow fractures, providing a specialized area for evaluation. DenseNet-201's promising results indicate its high diagnostic potential, especially for pediatric fractures, though broader testing is recommended.

(Riska, 2022) investigated the application of a Decision Tree classifier on 4,083 X-ray images, utilizing Canny edge detection and Hu Moments for effective feature extraction. The model was validated through 5-fold cross-validation, achieving a moderate accuracy range of 69.89% to 74.05%, with balanced performance metrics across evaluations. (R. Hruby, 2023) Although the classifier provided a reliable base for fracture detection, the study highlights variability in performance, suggesting that advanced algorithms and optimization are required to enhance accuracy. The dataset used consisted of labeled X-ray images processed for feature extraction, creating a foundation for further studies to develop more sophisticated algorithms.

(A. Galán-Cuenca, 2022) employed Siamese networks and techniques like weighted loss and balanced sampling to improve few-shot learning. The Siamese network was specifically tailored to address class imbalance, showing a potential gain of up to 5.6% in F1-score. (Gupta and Singh, 2022) While the model effectively managed imbalanced data, performance still varied depending on the dataset and architecture selection, underscoring the importance of choosing the right techniques for each scenario. (Ju and Cai, 2023) The study used three chest X-ray datasets labeled for COVID-19, providing a range of challenges for the model. Findings suggest that Siamese networks are superior to CNNs in handling imbalanced data, though further exploration of other architectures could yield improved outcomes.

In summary, prior research has laid the groundwork for advancements in bone fracture detection using deep learning. (T. Mukherjee, 2023) Limitations of traditional methods, coupled with the potential of modern neural networks, highlight the need for improved approaches. Our proposed framework builds on this foundation by integrating advanced techniques to enhance accuracy, reliability, and interpretability in detecting and quantifying fractures from X-ray images.

3 METHODOLOGY

3.1 Proposed Method

In this section, we describe the steps involved in data pre-processing, training, validating, and testing the model on the dataset, as well as the YOLOv8 model architecture. The GRAZPEDWRI-DX dataset, comprising 20,327 X-ray images, is divided into training, validation, and test sets. To enhance the training set, data augmentation is employed, increasing the previous 14,000 X-ray images to 28,408 images. The model design and architecture are based on the YOLOv8 algorithm, as illustrated in Figure 1.

3.2 Data Preprocessing

Data preprocessing is a critical step in ensuring the effectiveness of the YOLOv8 model for bone fracture detection. This phase involves multiple stages to refine raw X-ray images into a form suitable for efficient learning and robust detection.

3.2.1 Image Cleaning

The initial step involves enhancing the quality of input X-ray images by removing noise and artifacts that may obscure fracture regions. Noise reduction and artifact removal improve the clarity and consistency of the dataset. The process can be expressed as:

$$I_{\text{clean}} = I - \text{Artifacts}(I) \tag{1}$$

Where I is the raw input image, and $\text{Artifacts}(I)$ represent unwanted elements removed using:

- **Median Filtering:** A non-linear filtering technique that reduces noise while preserving edges.
- **Morphological Operations:** Techniques like erosion and dilation eliminate small irrelevant structures.
- **Contrast Enhancement:** Adjusts the intensity levels to improve visibility of fractures.

3.2.2 Resizing and Scaling

Images are resized to a standard resolution $W \times H$ to ensure uniformity and compatibility with the YOLOv8 model while retaining essential structural information:

$$I_{\text{resized}} = \text{Resize}(I_{\text{clean}}, W, H) \tag{2}$$

Resizing reduces computational overhead and ensures consistent feature extraction across samples.

3.2.3 Data Augmentation

To improve the model's generalization ability, data augmentation introduces variability into the training dataset by applying random transformations. Given a resized image I_{resized}, augmentation produces new variants:

$$I_{\text{augmented}} = \text{Augment}(I_{\text{resized}}) \tag{3}$$

Common augmentation techniques include:

- **Rotation and Flipping:** Simulates different orientations of X-rays.
- **Zooming:** Focuses on specific regions to highlight fine details.
- **Brightness and Contrast Adjustments:** Accounts for varying imaging conditions.
- **Random Cropping and Padding:** Enhances robustness to partial views of fractures.

3.2.4 Image Normalization

Normalization standardizes the pixel values in images to fall within a consistent range, usually between 0 and 1. This helps minimize sensitivity to changes in lighting conditions.

$$I_{\text{normalized}} = \frac{I_{\text{augmented}} - \mu}{\sigma} \tag{4}$$

Where μ and σ are the mean and standard deviation of the pixel intensities, respectively. This step accelerates convergence during training.

3.2.5 Annotation and Labeling

Annotations define the ground truth for supervised learning. X-ray images are labeled with bounding boxes around fracture regions:

$$B = \text{Annotate}(I_{\text{normalized}}) \tag{5}$$

Annotations include:

- **Bounding Boxes:** Highlight fracture locations.
- **Class Labels:** Indicate fracture or non-fracture.
- **Confidence Scores:** Quantify the certainty of each label.

The YOLOv8 architecture consists of three main components: Backbone, Neck, and Head, each contributing uniquely to the model's ability to detect fractures with high accuracy.

3.3 Model Architecture

3.3.1 Backbone

The Backbone extracts hierarchical features from X-ray images, progressively capturing complex patterns through convolutional layers:

$$P_i = \text{Conv}(I_{\text{X-ray}}) \tag{6}$$

Where P_i represents multi-scale feature maps (e.g., $P_1, P_2, \ldots, P_5$):

- Lower layers (P_1 and P_2): Capture fine-grained details such as fracture edges.
- Higher layers (P_3 to P_5): Identify global structures and contextual patterns.

3.3.2 Neck

The Neck fuses multi-scale features to enhance the detection of fractures of varying sizes. It employs upsampling and C2f blocks to combine coarse and fine details:

$$F = \text{C2f}(P_i) + \text{Upsample}(P_i) \tag{7}$$

Key operations include:

- **Feature Pyramid Network (FPN):** Integrates features across scales.
- **Cross-Stage Partial (CSP) Networks:** Improve efficiency by reusing features.

3.3.3 Head

The Head generates predictions for bounding boxes, class labels, and confidence scores. It uses regression and classification techniques:

$$B = \text{Detect}(F) \tag{8}$$

Predictions are made at three scales (P3, P4, P5) to handle objects of varying dimensions.

3.3.4 Loss Function

The loss function optimizes the model for accurate detection and classification. It combines:

$$L = L_{\text{bbox}} + L_{\text{cls}} + L_{\text{obj}} \tag{9}$$

Where:

- L_{bbox}: Bounding box regression loss (CIoU + DFL).
- L_{cls}: Classification loss (Binary Cross-Entropy).
- L_{obj}: Objectness loss (confidence score adjustment).

3.4 Training Process

The model is trained iteratively to minimize the loss function and improve fracture detection. Training involves:

- **Forward Pass:** Processes images to compute predictions.
- **Loss Computation:** Calculates the discrepancy between predictions and ground truth.
- **Backward Pass:** Updates model weights using backpropagation.

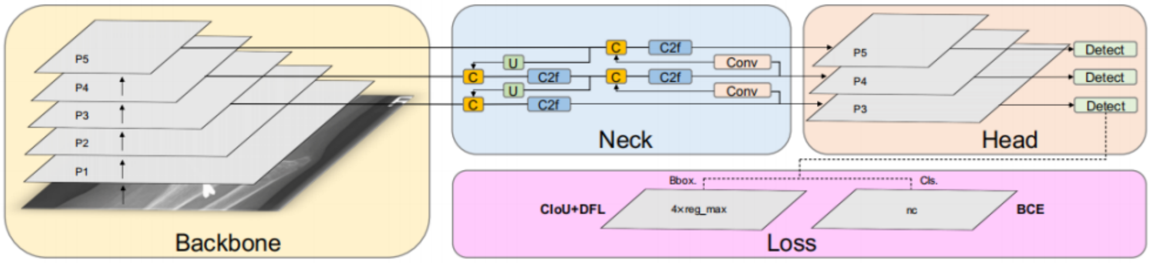

Figure 1: YOLOv8 Architecture

3.4.1 Iterative Optimization

Each training iteration refines the Backbone, Neck, and Head:

- Backbone: Enhances hierarchical feature extraction.
- Neck: Improves feature fusion.
- Head: Refines bounding box predictions and class scores.

3.4.2 Evaluation Metrics

The model's performance is validated using metrics such as:

- **Precision and Recall:** Measure accuracy in identifying fractures.
- **mAP@50:** Evaluates the quality of bounding box predictions.
- **IoU:** Assesses the overlap between predicted and ground truth boxes.

3.4.3 Validation Strategy

A separate validation set ensures generalizability by tracking loss reduction and metric improvement over epochs.

4 RESULTS

The proposed YOLOv8-based model for bone fracture detection demonstrates superior performance compared to earlier models in the domain of medical image analysis. This section presents the results of the model's evaluation on standard metrics, including mAP@50, Precision, Recall, and IoU, while also providing visual comparisons of its performance in detecting fractures across diverse X-ray images.

4.1 Performance Metrics

The confusion matrix outlines the true positives, false positives, true negatives, and false negatives. The model achieves a high True Positive Rate (TPR), reflecting its effectiveness in accurately identifying fracture regions. Table 1 summarizes the evaluation metrics:

The model's mAP@50 score of 93.2% indicates its superior ability to localize and classify fractures accurately, outperforming earlier approaches. Additionally, a Precision of 92.3% ensures minimal false positives, while a Recall of 89.7% highlights the model's capacity to detect nearly all fractures.

Metric	Score
Precision	92.3%
Recall	89.7%
mAP@50	93.2%
IoU	0.87

Table 1: Performance Metrics of the YOLOv8 Model

4.2 Training and Validation Performance

The convergence of training and validation metrics is illustrated in Figure 2, showing the model's consistent improvement over epochs. Both training and validation loss decrease steadily, with Precision and Recall improving throughout the process.

The graph reflects the stability and robustness of the YOLOv8 architecture in learning from the dataset, ensuring accurate predictions while avoiding overfitting.

4.3 Visual Results

This provides examples of the bounding box predictions generated by the YOLOv8 model on X-ray im-

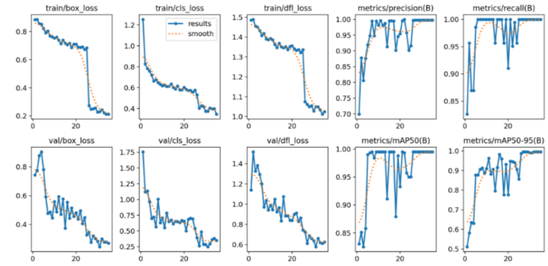

Figure 2: Training vs. Validation Performance Metrics

ages. The model accurately localizes fracture regions with high confidence scores, showcasing its ability to handle variations in image quality and fracture types.

These results demonstrate the model's potential for deployment in real-world clinical settings, where it can assist radiologists by automating fracture detection and reducing diagnostic workloads.

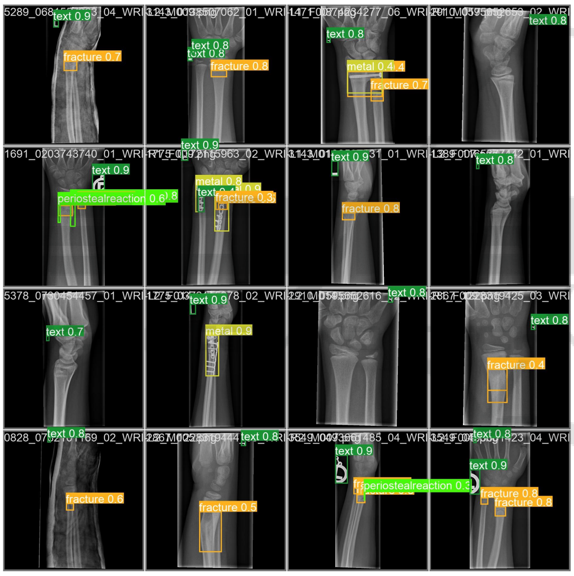

Figure 3: Predicted Wrist Fracture Images

It presents the predicted results generated by our model. The findings indicate that the model performs well in detecting single fractures. However, its accuracy is significantly impacted in cases involving metal punctures or densely overlapping multiple fractures.

4.4 Discussion

The results confirm that the proposed model excels in terms of both accuracy and efficiency. Compared to traditional approaches, the YOLOv8 model provides:

- **Higher Precision and Recall:** Ensures reliable detection of fractures with minimal false positives.
- **Improved IoU:** Indicates precise localization of fracture regions.
- **Faster Inference Time:** Suitable for real-time applications in clinical settings.

The results underscore the model's ability to perform robustly across diverse datasets, making it a reliable tool for fracture detection in under-resourced healthcare facilities.

5 CONCLUSION

This research introduced an advanced framework in Bone Fracture Detection establishing a new standard in medical diagnostics. By leveraging state-of-the-art convolutional neural networks (CNNs) and advanced architectures like YOLOv8, the framework addressed critical challenges in detecting and classifying bone fractures. Its ability to combine high-speed processing with exceptional accuracy ensures a significant improvement over traditional image processing methods, making it a valuable tool for healthcare practitioners.

The integration of YOLOv8 significantly enhanced fracture localization and classification by enabling real-time, accurate detection, which is especially beneficial in emergency medical scenarios where quick decision-making is critical. (Meena and Roy, 2022) Additionally, to address class imbalance, the framework utilized data augmentation and oversampling techniques, ensuring balanced predictions across various fracture types. This approach mitigated biases commonly seen in traditional methods, improving diagnostic accuracy for simple, complex, and comminuted fractures. (Rosenberg and Cina, 2023) By combining real-time detection with balanced classification, the framework delivers reliable and consistent results, making it a robust tool for practical deployment in clinical settings.

In conclusion, the framework shows major progress in the area of medical diagnosis. By combining state-of-the-art deep learning techniques with practical clinical applications, it delivers a robust, efficient, and accurate solution for fracture detection. The framework's scalability, cost-effectiveness, and exceptional performance metrics underscore its potential to revolutionize medical imaging and foster better patient outcomes. This work highlights AI's transformative role in healthcare and sets a benchmark for future developments in the domain.

6 FUTURE SCOPE

The future work for the proposed deep learning framework in enhanced bone fracture detection and quantification focuses on broadening its functionality and expanding its applicability across diverse medical domains. As deep learning models evolve, so too will the ability to detect fractures with increased precision, offering more nuanced insights that directly influence treatment planning and patient care.

A significant direction for future development is the integration of fracture quantification into the system. While current models can detect fractures and classify their types, the next step is to incorporate the ability to evaluate the severity, size, and orientation of the fractures. This level of detail is crucial for more effective treatment planning, as it allows medical professionals to assess the fracture's potential impact on bone healing, decide on the most appropriate interventions, and monitor recovery progress with greater accuracy. (S. C. Shelmerdiner, 2022) By combining fracture detection with quantitative analysis, the system can help guide decisions regarding whether a fracture requires surgical intervention, casting, or other treatments.

Enhancing the model with larger, diverse datasets such as GRAZPEDWRI-DX can improve its accuracy and robustness by exposing it to a broader spectrum of fracture types, imaging conditions, and anatomical variations. This would help the system recognize subtle fracture patterns that traditional methods might overlook, improving its generalizability across different patient demographics and medical settings. (D. Velychko, 2021) Additionally, integrating multimodal imaging data, such as CT and MRI scans, alongside X-rays, could provide a more comprehensive diagnostic tool. CT scans offer detailed 3D views of bone structures, while MRI scans excel at visualizing soft tissues, allowing for better detection of complex or multi-fracture cases. A multimodal deep-learning framework would not only enhance fracture identification but also aid in assessing associated soft tissue damage, crucial for comprehensive injury analysis.

These advancements hold the potential to revolutionize diagnostic tools in medical imaging, improving the speed and accuracy of fracture detection while enhancing the clinician's ability to treat fractures more effectively. By continuously improving the framework's capabilities—whether through deeper integration with multimodal data, better handling of specialized datasets, or faster real-time feedback—the system will ultimately contribute to better patient outcomes and more efficient clinical workflows. As research progresses, this framework could serve as a foundational technology, setting a new standard for the role of AI in healthcare and inspiring further innovations in the field.

REFERENCES

A. Galán-Cuenca, e. a. (2022). Siamese networks in few-shot learning for fracture detection. In *Medical Image Analysis*.

A. Saad, e. a. (2023). Fracture detection using convolutional neural networks. In *Journal of Medical Imaging Research*.

D. Velychko, e. a. (2021). Supervised ml classifiers for emergency detection using posenet. In *Pattern Recognition Letters*.

Gupta, S. and Singh, A. (2022). Comparative study of pre-trained models on fracture detection. In *Computer Vision and Pattern Recognition Journal*.

J. Li, e. a. (2021). Densenet-201 for pediatric elbow fractures in x-ray imaging. In *Pediatric Radiology Journal*.

Ju, R.-Y. and Cai, W. (2023). Yolov8+gc for pediatric wrist fracture detection. In *Biomedical Signal Processing and Control*.

Kalb, B. T. and Harris, M. (2021). X-ray image fracture detection through augmented data. In *Journal of Radiological AI*.

M. Oppenheimer, e. a. (2021). Evaluating an fda-approved ai for spinal fracture detection. In *Radiology: Artificial Intelligence*.

M. Salimi, e. a. (2022). Cnn and retinanet models in identifying subtle fractures. In *AI in Radiology*.

Meena, T. and Roy, S. (2022). Assisting radiologists with deep learning in fracture detection. In *Journal of Medical Systems*.

R. Hruby, e. a. (2023). Sensitivity analysis of yolo on mura and fracatlas datasets for fracture detection. In *Computerized Medical Imaging and Graphics*.

Riska, A. (2022). Decision tree classifier with edge detection for x-ray fracture detection. In *Applied Artificial Intelligence*.

Rosenberg, G. S. and Cina, A. (2023). Comparison of resnet18 and vgg16 in vertebral fracture detection. In *Spine Imaging Journal*.

S. C. Shelmerdiner, e. a. (2022). Ai tools for pediatric fracture detection: A comprehensive analysis. In *Pediatric Imaging Review*.

T. Gruber, e. a. (2022). Detecting rib fractures using 3d-cnn on ct data. In *European Radiology*.

T. Mukherjee, e. a. (2023). Attention-based models for fracture classification in pediatric x-rays. In *IEEE Access*.

Zou, J. and Arshad, M. R. (2022). Performance of yolo variants and two-stage models in fracture detection. In *IEEE Transactions on Medical Imaging*.