Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

Comparative Analysis of YOLOv11 with Previous YOLO in the Detection of Human Bone Fractures

Febri Aldi[1)*], Irohito Nozomi[2)], M. Hafizh[3)], Triana Novita[4)]
[1,2,3,4)]Universitas Putra Indonesia YPTK Padang, Indonesia
[1)]febri_aldi@upiyptk.ac.id, [2)]irohito_nozomi@upiyptk.ac.id, [3)]muhammad_hafizh@upiyptk.ac.id, [4)]triana_novita@upiyptk.ac.id

ABSTRACT

Accurate and rapid detection of bone fractures is an important challenge in the medical world, particularly in the field of radiology. This study aims to analyze and compare the performance of the YOLOv11 model with several previous versions of YOLO, namely YOLOv5, YOLOv8, and YOLOv10 in the task of detecting human bone fractures on X-ray and MRI images. The dataset used is the Human Bone Fractures Multi-modal Image Dataset (HBFMID) which consists of 641 raw images (510 X-rays and 131 MRIs). The four models were trained using the HBFMID dataset that had gone through a manual augmentation and annotation process, then tested using evaluation metrics such as precision, recall, mAP50, and mAP50-95. The training results showed that YOLOv11 has the most stable and consistent loss curve, with a fast convergence process. In terms of evaluation, YOLOv11 recorded a precision of 99.87%, a recall of 100%, a 99.49% mAP50, and an 84.13% increase in the number of mAP-95s, which generally outperformed other models. In addition, the visual prediction results show that YOLOv11 can detect fracture areas with the right bounding box and a balanced confidence score, without showing symptoms of overconfidence or inconsistency. When compared to previous studies, YOLOv11 shows significant improvement in detection accuracy. Thus, YOLOv11 has great potential to be applied in medical diagnosis support systems to improve the efficiency and accuracy of digital fracture identification. However, the study is limited by dataset size and modality scope. Future work may explore larger, more diverse datasets and multi-task learning.

Keywords: Bone fracture detection; HBFMID; deep learning; YOLO; YOLOv11

INTRODUCTION

Bone fractures or fractures are one of the common medical conditions that occur due to accidents, sports injuries, or pathological conditions such as osteoporosis (Dhahir et al., 2017). Early and accurate detection of fractures is essential to ensure proper treatment and prevent further complications, such as mobility disorders, paralysis, and bone deformities (Ragnarsson, 2015). In clinical practice, fractures are generally identified through a radiographic examination (X-ray) analyzed by medical personnel (Regnard et al., 2022). However, the process of interpreting radiological images manually depends heavily on the experience and expertise of the doctor, and can be influenced by fatigue factors, case complexity, and image quality (McLaughlin et al., 2022).

In Indonesia, the incidence of bone fractures reaches around 1.3 million cases annually, making it the highest in Southeast Asia (Sasmito et al., 2023). This high number is caused by various factors, including traffic accidents, work accidents, and increased physical activity of the community without adequate safety education. In addition, the growing elderly population is also a major risk factor for the increase in fractures due to osteoporosis (Bouvard et al., 2021). On the other hand, the limited number of radiology personnel, especially in remote areas, results in many cases of fractures not being detected early or experiencing delayed diagnosis (Nigatu et al., 2023). This has an impact on the quality of health services and increases the risk of complications in patients. Therefore, a technology-based solution is needed that is able to help the fracture detection process automatically, quickly, and accurately, to support more equitable and efficient medical services throughout Indonesia.

In today's digital and information technology era, the application of artificial intelligence (AI), especially deep learning in the medical field, has begun to be widely developed to help the diagnosis process (Aldi et al., 2024; Tian et al., 2024). One approach that shows great potential in the analysis of medical images is object detection using the YOLO (You Only Look Once) algorithm (Liu et al., 2024; Ragab et al., 2024). YOLO is a real-time object detection

* Corresponding author

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

method based on convolutional neural network (CNN) that is able to detect and classify objects in one efficient and fast inference process (Amin et al., 2024). This algorithm was originally designed for general tasks such as vehicle or human detection, but has now been widely adapted for medical applications, including detecting bone abnormalities from X-ray images. Previous versions such as YOLOv5, YOLOv8, and YOLOv10 have been used in various medical applications, including bone fracture detection, due to their ability to recognize objects at high speeds (Meza et al., 2024).

However, over time, the need for more precise detection of complex areas such as fine cracks in bones demands improved algorithm performance. The latest version, YOLOv11, comes with a variety of significant updates, such as improved backbone architecture, improved feature fusion mechanisms, and adaptive training strategies (Tao et al., 2025). YOLOv11 is designed to provide higher accuracy and more effective multi-scale detection, especially in cases of fractures that are difficult to recognize (Zhao, 2025). The use of YOLOv11 in detecting human bone fractures not only aims to speed up the diagnosis process, but also improve the consistency and accuracy of case identification (Wei et al., 2025). Models trained on annotated X-ray datasets, such as MURA (Musculoskeletal Radiographs) or curated local datasets, can recognize a wide range of fracture patterns such as transverse, oblique, spiral, and comminutive fractures. This is very helpful for medical personnel, especially in healthcare facilities with limited radiologists, in making decisions quickly and appropriately.

Based on these developments, there has been no study that directly compares YOLOv11 with previous versions in detecting bone fractures using the HBFMID dataset. This study addresses this gap by analyzing the comparative performance of YOLOv11 with previous versions (such as YOLOv5, YOLOv8, and YOLOv10) in detecting human bone fractures from X-ray images. The evaluation was carried out by training the model on annotated datasets and testing test images to assess accuracy, precision, recall, and mAP values. It is hoped that the results of this study can provide a deeper understanding of the effectiveness of each version of YOLO and encourage the use of deep learning technology in a more efficient and reliable computer-based medical detection system in Indonesia.

LITERATURE REVIEW

This section discusses the research basis based on a review of previous studies related to deep learning-based human fracture detection using the YOLO algorithm presented in Table 1.

Table 1. Previous research

Author	Model	Metode	Result
(Jeon et al., 2023)	YOLOv4	- Feature extraction (CSPDarkNet-53 Backbone, Neck Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PANet)) - Mapping output to 3D	Their proposed system allows intuitive visualization of fracture areas by adding a distinctive red mask displayed on top of the 3D reconstructed bone image. A fairly high average precision value (>0.60) was achieved, namely 0.71 for the PR curve of the tibia and 0.81 for the elbow. The Intersection over Union (IoU) values obtained were 0.6327 for the tibia and 0.6638 for the elbow, respectively.
(Zou & Arshad, 2024)	YOLOv4, v5, v7, v8, YOLOv7-ATT, SSD, Faster-RCNN, Mask-RCNN	Addition of attention mechanism, use of EIoU loss function	Most notably the YOLOv7-ATT model achieved an mAP of 80.2%, which shows an excellent generalization on the FracAtlas dataset, achieving an mAP of 86.2%, which is significantly better than the other models.
(Parvin & Rahman, 2024)	YOLOv8	Deep learning for classification	The proposed study effectively identifies and classifies different types of fractures in this area. Our system achieves 95% accuracy, 93% regain, and 92% average precision. The results show that this method achieves cutting-edge performance
(Medaramatla et al., 2024)	YOLO NAS	Incorporation of modern object detection architecture to improve detection accuracy and efficiency.	The proposed model obtained a good precision level of 0.989, which indicates that the proposed model is excellent for detecting fractures of the hand and joints
(Zhou et al., 2025)	Mandibula-YOLO	- MRFEM (Multiscale Residual Feature Enhancement Module)	The Mandible-YOLO model shows excellent performance in detection. The model reached 97.02%, 97.12%, 93.82%, and 95.11%, respectively

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

		- SCFHM (Spatial-Channel Feature Hybrid Module) - GLFHM (Global-Local Feature Hybrid Module)	on the mAP 0.5, Pre, Rec, and F1 indices
(Srinivasu et al., 2025)	YOLOv10	- Exploration of training parameter combinations for optimal outcomes - Image augmentation includes, image insharpness masking and CLAHE (Contrast Limited Adaptive Histogram Equalization)	The proposed model yielded an accuracy of 0.964 on the evaluation of the magnified data. Statistical analysis of classification precision on enlarged and raw images is seen as 0.98 and 0.95, respectively

Based on the above literature review, it can be concluded that to date there has been no research that explicitly utilizes the YOLOv11 algorithm to detect human bone fractures from X-ray images. Taking into account the advantages of the YOLOv11 architecture in terms of detection accuracy, computational efficiency, and its ability to recognize small-sized objects, this study has the opportunity to make a new contribution to the development of a more accurate and efficient fracture detection system. In addition, this approach also allows testing of model generalizations on different types of fractures and bones, and provides a basis for integration into AI-based clinical decision support systems.

METHOD

This section systematically describes the stages carried out in comparative analysis of bone fracture detection using the YOLO algorithm. In this case it includes the process of data collection and pre-processing, YOLO model training using performance measurement metrics. In addition, a comparison of the performance of YOLOv11 with previous versions of YOLO was also carried out to assess the performance and effectiveness of the model in detecting various types of bone fractures from X-ray and MRI images.

The YOLO model is implemented using PyTorch 2.0, a flexible and efficient deep learning library. To harness the power of the YOLOv11 architecture, the researchers used the Ultralytics YOLO framework, which provides an easy-to-use interface and supports optimal training and inference of object detection models. The training process was conducted in the Google Colab Pro environment, which provides high computing resources, including the use of the NVIDIA GPU A100. This GPU can speed up computing during model training, thereby speeding up the convergence process and optimizing the fracture detection performance of bone images significantly.

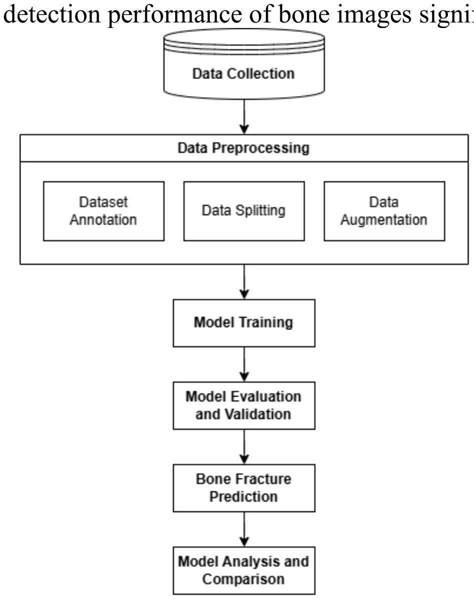


Fig. 1 Research Stages

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

Figure 1 is the research stage in the comparative analysis of YOLOv11 with the previous version of YOLO for the detection of bone fractures. The research began with the collection of X-ray imagery datasets, which then went through a data pre-processing process in the form of fracture annotations, division of the dataset into subsets of training, validation, and testing, and data augmentation using flipping techniques and brightness changes to increase data diversity. Furthermore, model training was carried out using four versions of the YOLO algorithm, namely YOLOv5, YOLOv8, YOLOv10, and YOLOv11. The trained model is then evaluated using precision, recall, mAP50, and mAP50:95 metrics to measure its performance. The results of the evaluation were used in the bone fracture prediction stage, and the entire model was compared in the model analysis and comparison stage to determine the most optimal version of YOLO in detecting human bone fractures from X-ray imaging.

Dataset Collection

The Human Bone Fractures Multi-modal Image Dataset (HBFMID) is a collection of medical images (X-ray and MRI) collected for this research in the detection of bone fractures in different parts of the human body, supporting in computer vision and deep learning for medical applications. The initial dataset consisted of 641 raw images (510 X-rays and 131 MRIs). Examples of X-ray and MRI images are presented in Figure 1.

(a) (b)

Fig. 2 Bone Fractures on X-rays (a) and MRI (b)

Figure 2 (a) is a human shin fracture from an X-ray image, while (b) is a human spinal fracture from an MRI image. These two examples are examples of datasets that have been collected from X-ray and MRI images. Among the locations of human bone fractures from this dataset are the elbow, fingers, forearm, humerus, shoulder, femur, shin bone, knee, hip bone, wrist, spinal cord, and some healthy bones.

Data Pre-processing

The X-ray and MRI images collected will go through pre-processing stages such as data annotation, data augmentation (flipping and brightness), and dataset sharing. This aims to increase the variety of training data and model performance.

Dataset annotations

Dataset annotation is the process of labeling data, such as images or videos, to help models recognize objects (Fernandes et al., 2024). In this study, annotations were carried out on human bone fracture objects using a bounding box through the Roboflow platform, so that the model could learn to recognize visual patterns accurately and efficiently.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

Fig. 3 Bone Fracture Bounding Box Using Roboflow

Data Splitting

Dataset splitting is the process of dividing data into subsets to train, evaluate, and test models (Walston et al., 2024). This division aims to enable models to learn optimally and to generalize their abilities to new data. The dataset is divided by a ratio of 70:20:10, meaning 70% for training, 20% for validation, and 10% for testing.

Dataset Augmentation

To improve the model's performance in detecting bone fractures from X-ray images, a data augmentation process was carried out before training (Kumar et al., 2023). Data augmentation aims to enrich the variation of images in the dataset, so that the model becomes more robust against differences in position, lighting, and image conditions that may be encountered in the real world (Islam et al., 2024). In this study, the two main augmentation techniques used were flipping and brightness adjustment.


(a) Asli (b) Flipping (c) Brightness

Fig. 4 Example of Augmentation Image

Figure 4 is one of the augmented training images. Figure 4 (b) and (c) are the result of augmentation from the original image (a) using flipping and brightness techniques. All of these augmented training images were used for the model training process with a total of 1,347 trained images.

Model Training

In this section, we discuss the training process of human bone fracture detection models using YOLOv5, YOLOv8, YOLOv10 and especially YOLOv11 algorithms as a comparison reference. The YOLOv11 architecture is a significant improvement over previous versions, specifically YOLOv8. YOLOv11 incorporates new layers, blocks, and optimizations that improve computing efficiency and detection accuracy, making it ideal for real-time tasks such as human detection. The YOLOv11 architecture consists of three main parts, namely the backbone, neck, and head (Ali & Zhang, 2024).

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

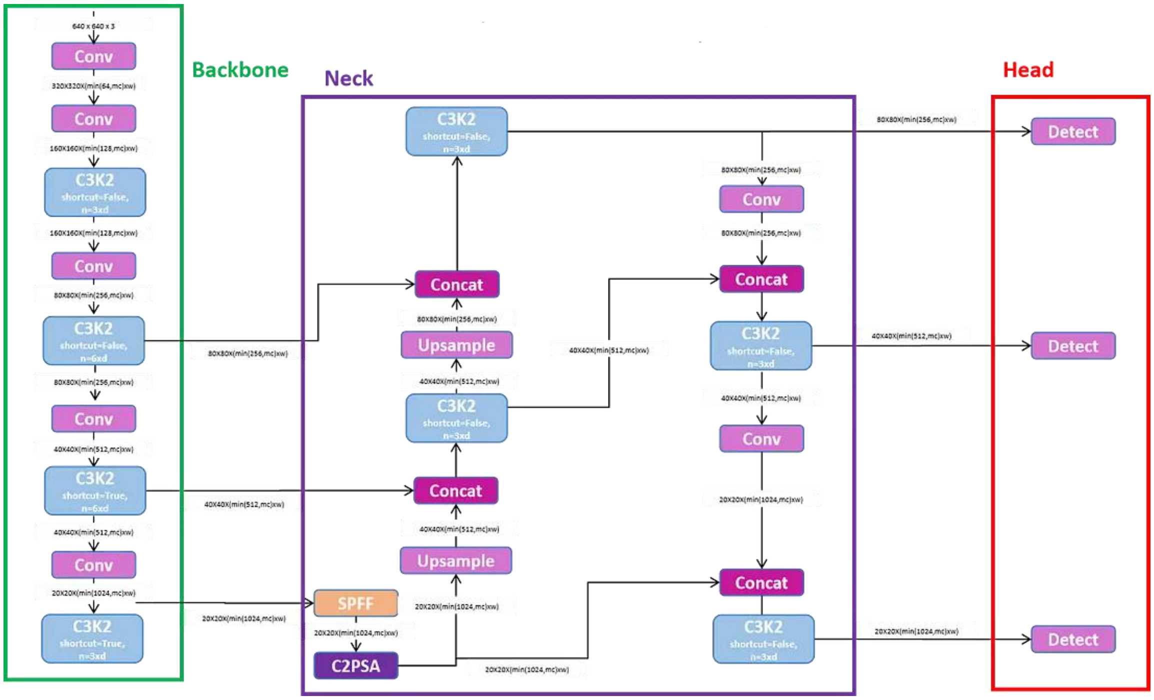


Fig. 5 YOLOv11 Architecture

The YOLOv11 architecture is designed to optimize speed and accuracy, by refining various developments that have been introduced in previous versions such as YOLOv5, YOLOv8, and YOLOv10. The main innovation in the YOLOv11 architecture lies in the use of the C3K2 block, SPFF module, and C2PSA block, which significantly improve the model's ability to process spatial information without sacrificing inference speed. The backbone of YOLOv11 functions as a feature extractor from input images at various scales, through a series of convolution layers and special blocks that produce feature maps with different resolutions. YOLOv11 introduces the C3K2 block and retains the Spatial Pyramid Pooling Fast (SPPF) module from the previous version, as well as adding a new improvement in the form of the C2PSA block (Ankan Ghosh, 2024). The initial part of the YOLOv11 convolution layer (in Equations 1 and 2) begins with several convolution layers that are tasked with downsampling the input image.

$$Conv_1 = Conv(I, 64, 3, 2) \quad (1)$$
$$Conv_2 = Conv(Conv_1, 128, 3, 2) \quad (2)$$

This layer progressively reduces the spatial resolution while increasing the depth of the feature map. The C3k2 Block section, YOLO11 introduces the more efficient C3k2 block, which is based on the Cross-Stage Partial (CSP) network. The C3k2 block consists of two smaller convolutions (kernel size = 2) to reduce compute costs while maintaining performance. This block is explained in equation 3.

$$C3k2(X) = Conv\ (Split(X)) + Conv\ (Merge(Split(X))) \quad (3)$$

Where Split(X) divides the feature map into two parts, one is processed through bottlenecks, and Merge combines its outputs. SPPF and C2PSA Blocks perform spatial collection at various scales. This block is described in equation 4.

$$SPPF(X) = Concat(MaxPool(X, 5), MaxPool(X, 3), MaxPool(X, 1)) \quad (4)$$

YOLO11 introduces the C2PSA block, which improves spatial attention across the feature map. This helps the model to focus on the specific regions in the image that are most relevant for detection, thereby improving performance on small and enclosed objects.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

$$C2PSA(X) = Attention(Concat(X_{path1}, X_{path2})) \quad (5)$$

The YOLO11 neck is designed to combine feature maps of various resolutions and pass them to the detection head. YOLO11 integrates C3k2 blocks into the neck to improve the speed and performance of feature aggregation. Feature Aggregation on the Neck applies the upsampling and merge layers to combine feature maps of various scales as in equations 6, 7 and the use of C3k2 blocks after the merge ensures efficient feature aggregation on equation 8.

$$Feature_{upsample} = Upsample(Feature_{previous}) \quad (6)$$
$$Feature_{concat} = Concat(Feature_{upsample}, Feature_{lower}) \quad (7)$$
$$C3k2_{neck} = Conv_{small}(Concat(Feature_{concat})) \quad (8)$$

The Neck component in YOLOv11 is equipped with a spatial attention mechanism through the integration of the C2PSA block, which improves the model's ability to focus attention on important areas in the image, especially in complex situations such as overlapping objects. Meanwhile, the head section in YOLOv11 is responsible for producing the final prediction of the model. Similar to the previous version, this head produces bounding boxes, class probabilities, and confidence scores. In the Detection Layer section, YOLOv11 uses three levels of detection scale, namely small (P3), medium (P4), and large (P5), which allow detection of objects of various sizes. Each scale processes feature maps from different resolution levels, so that the model can work effectively in recognizing objects both near and far from the camera.

$$Detect(P3, P4, P5) = BoundingBoxes + ClassLabels \quad (9)$$

The YOLOv11 model will be trained using the processed training data. The training parameters are at a learning rate of 0.01, a batch size of 64, and a number of epoches of 100 to get optimal results. The training process is carried out using a 40 GB GPU for time efficiency.

Evaluation and Validation

At this evaluation stage, the YOLOv11 model that has been trained using the HBFMID dataset will be tested to measure its performance in detecting objects accurately and efficiently. Model evaluation is an important step to determine the extent to which the model is able to recognize and detect various objects in the image, especially under conditions that may differ from the training data. A number of evaluation metrics used are Precision and Recall, and Mean Average Precision (mAP).

Precision measures the proportion of true positives to all positive detections made by the model, as presented in equation 10. Recall measures the proportion of true positives to the total number of objects present, as presented in equation 11.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \quad (10)$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \quad (11)$$

mAP is the average of the Average Precision (AP) for each class of object. AP is calculated by integrating the precision-recall curve for each class at various thresholds (IoU - Intersection over Union), which is typically valued at 0.5 (mAP@0.5) or a range from 0.5 to 0.95 (mAP@[0.5:0.95]). mAP is presented at equation 12.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \quad (12)$$

Where N is the number of classes, and AP_i is *Average Precision* for each class *i*. Precision, recall, and mean Average Precision (mAP) provide an overview of the accuracy and precision of the model in correctly detecting bone fracture objects. Thus, the results of this evaluation will be a benchmark for the effectiveness of the YOLOv11 model and provide an idea of the extent to which this model can be implemented in real scenarios.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

Bone Fracture Prediction

This section discusses the implementation of the YOLO model training results in the process of predicting bone fractures on X-ray and MRI images. After going through the training and evaluation stage, all models are used to automatically detect fracture areas. This process aims to test the model's ability to recognize fracture patterns directly on test images, as well as evaluate the accuracy and consistency of detection under real conditions. This prediction is expected to be a quick and reliable diagnostic support solution in clinical practice.

Model Analysis and Comparison

At the model analysis and comparison stage, an evaluation was carried out on the performance of the YOLOv11 model that has been trained in detecting bone fractures based on medical images. The model performance assessment was carried out using several main evaluation metrics, namely mean Average Precision (mAP), precision, and recall. These values are calculated to assess how well the model identifies fracture locations, how many fractures are detected correctly, and how accurate the bounding box predictions are. Visualization of detection results is also used as part of the analysis to see how accurately the model is marking fracture areas.

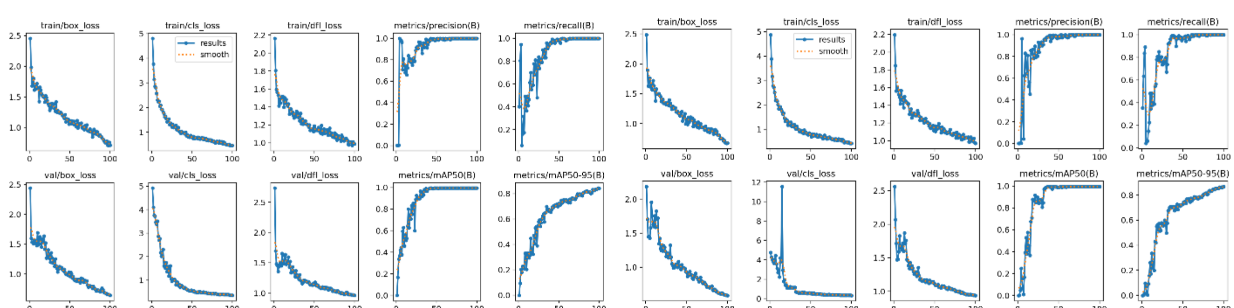
Furthermore, the performance of the YOLOv11 model was compared with the previous generation of YOLO models, such as YOLOv5, YOLOv8, and YOLOv10. The results of the comparison were used to assess the improvements achieved by YOLOv11, both in terms of detection accuracy and inference time efficiency. The analysis also includes a discussion of the advantages and limitations of each model, so that a comprehensive picture of the effectiveness of YOLOv11 can be obtained in the context of bone fracture detection. If the proposed model is proven to have superior performance compared to the previous model, then this can be a strong basis for recommending its use in the development of deep learning-based fracture detection systems in the medical field.

RESULT

This section presents the results of the research which includes three main stages, namely the model training process, testing of test data, and comparative analysis of performance between models. At the training stage, the YOLOv11 model was trained using a dataset of bone images that had gone through a pre-processing and augmentation process. Next, tests were carried out to evaluate the model's performance against previously unseen data, using evaluation metrics such as precision, recall, and mAP. The results of this test were then analyzed and compared with the results of the previous version of the YOLO model, to assess the improved performance and effectiveness of the YOLOv11 model in detecting bone fractures.

Model Training Results

This section presents training results from four bone fracture detection models, namely YOLOv5, YOLOv8, YOLOv10, and YOLOv11. Each model was trained using an augmented and annotated X-ray and MRI imagery dataset, with the aim of comparing the effectiveness of each model's learning process in recognizing bone fractures. The training results are visualized in the form of a graph to show the trend of decreasing loss values and increasing evaluation metrics during the training. This visualization is the initial basis for assessing the performance and stability of the model before entering the testing and prediction stage.

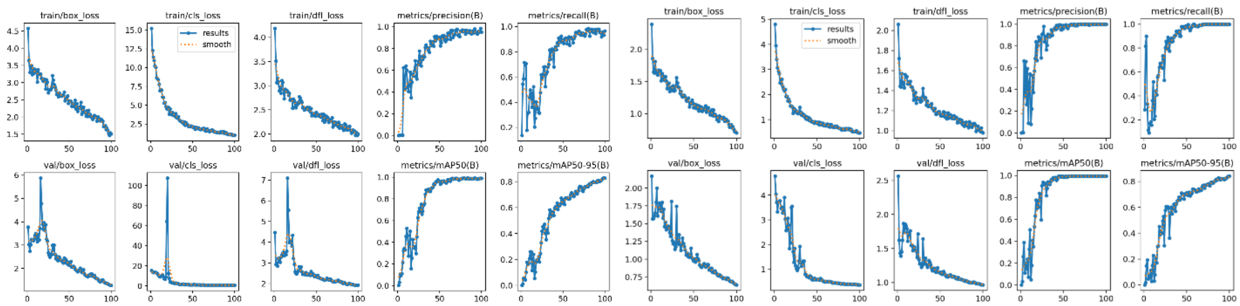


(a) (b)

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025



(c) (d)

Fig. 6 Training Results Model (a) YOLOv5, (b) YOLOv8, (c) YOLOv10, and (d) YOLOv11

Figure 6 shows a graph of the training results and evaluation of the YOLOv5, YOLOv8, YOLOv10, and YOLOv11 models over 100 epochs on bone fracture detection tasks. Each sub-figure (a) through (d) represents the training curve of each model, showing the development of metrics during the training process such as decreased loss values and increased detection accuracy. This curve provides important information about the stability and effectiveness of the learning process of each model against the manually annotated X-ray image dataset. In the YOLOv5 model (a), it can be seen that the decrease in losses occurs gradually, but with slight fluctuations that indicate a convergence process that is not completely stable. This model takes longer to reach the optimal point. Meanwhile, the YOLOv8 (b) model shows better performance, with more stable loss reduction and faster precision improvement. YOLOv10(c) appears to show better stability than YOLOv5 and YOLOv8, but there are still slight fluctuations at some stages of training. Most notable are the training results of the YOLOv11(d) model, which shows the most stable and consistent training curve. The decrease in losses occurred quickly and evenly, and was accompanied by a gradual increase in evaluation metrics without sharp fluctuations. This shows that YOLOv11 has a better learning ability to recognize bone fracture features in X-ray images than previous versions. Thus, the results of this training provide an initial indication that YOLOv11 is the most optimal model candidate to be used in the prediction and further evaluation stage of this study.

Bone Fracture Prediction Results

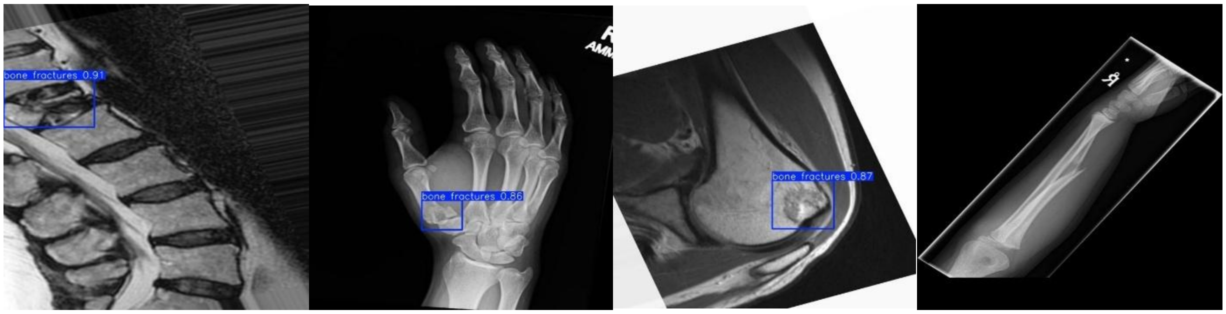
This section presents the results of bone fracture prediction using YOLOv5, YOLOv8, YOLOv10, and YOLOv11 models on X-ray and MRI images that have been prepared as test data. Predictions are made to observe the ability of each model to recognize and mark fracture areas directly through the bounding box. The visualization of the prediction results aims to provide a real picture of the accuracy of the model's detection of bone fractures, as well as the basis for qualitative evaluation in comparing the effectiveness of each version of YOLO tested.

(a)

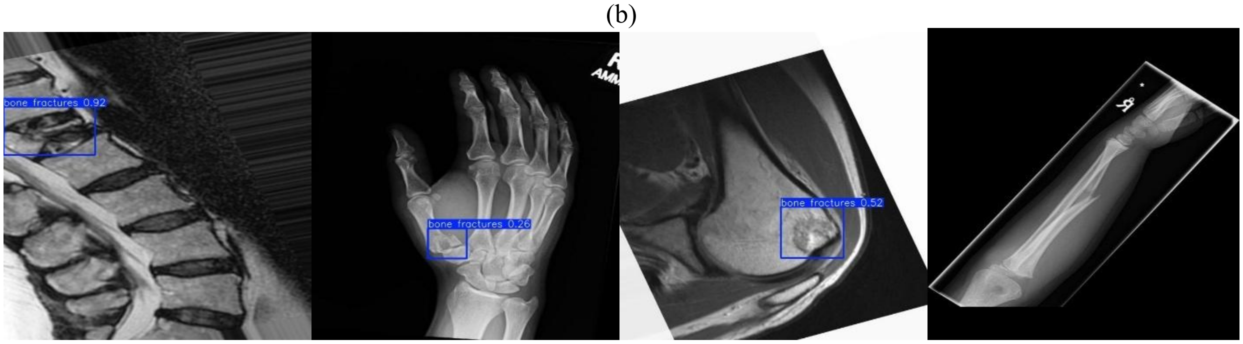
* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025



(b)



(c)

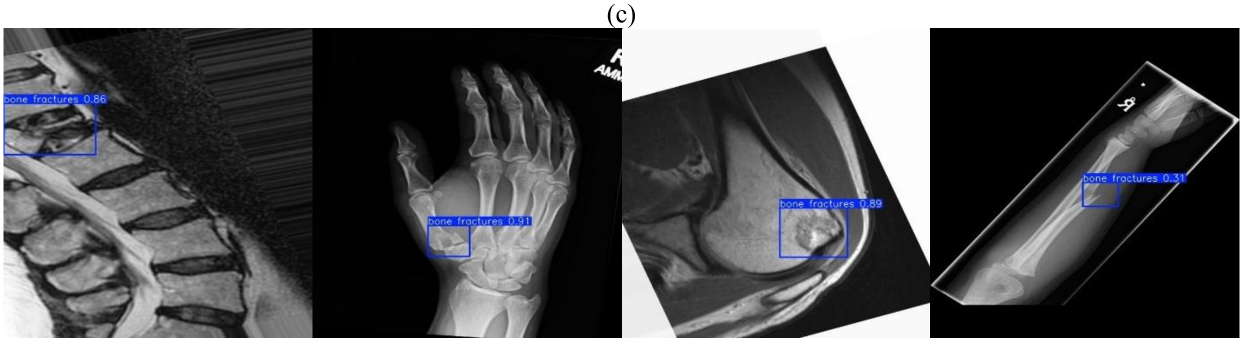


Fig. 7 Bone Fracture Prediction Results Model (a) YOLOv5, (b) YOLOv8, (c) YOLOv10, and (d) YOLOv11

Figure 7 is the result of fracture detection prediction from 64 test images taken randomly, using YOLO, namely YOLOv5, YOLOv8, YOLOv10, and YOLOv11. The results of the test image predictions are presented in Table 2 which informs the average confidence values, inference time, and FPS speed of each model.

Table 2. Comparison of Bone Fracture Test Image Prediction Results

Model	Average Confidence Score	Inference Time (seconds)	Speed (FPS)
YOLOv5	73%	**1,04**	**61,25**
YOLOv8	66%%	1,21	52,82
YOLOv10	42,5%	7,64	8,37
YOLOv11	**74,25%**	1,34	47,78

Table 2 shows the results of a comparison of bone fracture test image predictions from four YOLO models, namely YOLOv5, YOLOv8, YOLOv10, and YOLOv11, based on three aspects: average confidence score, inference time, and detection speed (FPS). In terms of average confidence score, the YOLOv11 model obtained the highest score of 74.25%, slightly superior to YOLOv5 with 73%, while YOLOv10 showed the lowest score of only 42.5%, indicating less stable predictive confidence. In terms of inference time, the fastest model is YOLOv5 with an average of 1.04 seconds, followed by YOLOv8 (1.21 seconds) and YOLOv11 (1.34 seconds). Meanwhile, YOLOv10 showed the slowest inference time, which was 7.64 seconds, which shows a lack of efficiency in prediction. In line with the

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

inference time, the predicted speed in frames per second (FPS) also shows that YOLOv5 has the fastest performance at 61.25 FPS, followed by YOLOv8 (52.82 FPS), YOLOv11 (47.78 FPS), and the slowest is YOLOv10 (8.37 FPS).

Model Analysis and Comparison

This section discusses the analysis and comparison of the performance of four bone fracture detection models, namely YOLOv5, YOLOv8, YOLOv10, and YOLOv11. The results of this study are also compared with previous studies. The analysis was carried out based on the results of training, evaluation of metrics, and prediction visualization of each model. The goal is to assess the accuracy, consistency, and effectiveness of each model in detecting bone fractures in X-ray images, as well as determine the most optimal model for use in deep learning-based automated detection systems.

Table 3. Comparison of YOLO Model Evaluation Metrics Results

Model	**Precision**	**Recall**	**mAP50**	**mAP50-95**
YOLOv5	99,93%	**100%**	99,47%	84,22%
YOLOv8	**99,94%**	**100%**	99,48%	**86,65%**
YOLOv10	95,29%	96,47%	99,02%	82,77%
YOLOv11	99,87%	**100%**	**99,49%**	84,13%

Table 4. Comparison of Evaluation Metrics Results with Previous Research

Author	**Model**	**Precision**	**Recall**	**mAP50**	**mAP50-95**	**IoU**
(Jeon et al., 2023)	YOLOv4	-	-	-	-	66,38%
(Zou & Arshad, 2024)	YOLOv7	-	-	86,2%	-	-
(Parvin & Rahman, 2024)	YOLOv8	95%	93%	-	-	-
(Medaramatla et al., 2024)	YOLO NAS	98,9%	-	-	-	-
(Zhou et al., 2025)	Mandibula-YOLO	97,12%	93,82%	97,02%	-	-
(Srinivasu et al., 2025)	YOLOv10	98%	-	-	-	-
Proposed (2025)	YOLOv11	**99,87%**	**100%**	**99,49%**	**84,13%**	-

Table 3 shows a comparison of the results of evaluation metrics from the four latest YOLO models used for bone fracture detection, namely YOLOv5, YOLOv8, YOLOv10, and YOLOv11. Based on these results, YOLOv8 recorded the highest precision of 99.94% and 100% recall, as well as achieving the second highest mAP50 of 99.48% and the highest mAP50-95 of 86.65%, indicating excellent performance in detecting fractures with high accuracy. Meanwhile, YOLOv11 as the proposed model showed very competitive results, with a precision of 99.87%, a recall of 100%, and an mAP50 of 99.49% (which is the highest among all models) and an mAP50-95 of 84.13%. YOLOv5 also gives good results with a 100% recall, but has a lower mAP50-95 (84.22%). On the other hand, the YOLOv10 has relatively lower precision and recall than other models, although the mAP50 is still relatively high.

When compared with previous studies as shown in Table 4, it can be seen that the proposed YOLOv11 model outperforms most of the previous approaches. For example, the Mandibula-YOLO model of Zhou et al. (2025) recorded an mAP50 of 97.02%, while YOLOv11 achieved an mAP50 of 99.49%, indicating a significant increase. The YOLOv10 reported by Srinivasu et al. (2025) has an accuracy of 98%, which is still below the precision of YOLOv11. Likewise, the YOLO NAS and YOLOv8 models from previous studies have an accuracy of 98.9% and 95%, respectively, which is lower than the proposed model. This reinforces the conclusion that YOLOv11 provides highly superior evaluation results in terms of accuracy and detection capabilities, and consistently surpasses previous research achievements. As such, YOLOv11 deserves to be considered as a cutting-edge solution in deep learning-based fracture detection systems.

DISCUSSIONS

Based on the results obtained, this discussion emphasizes the advantages of the YOLOv11 model in detecting bone fractures, both in terms of training performance, prediction accuracy, and quantitative evaluation results. The training process carried out on all four models (YOLOv5, YOLOv8, YOLOv10, and YOLOv11) showed that the YOLOv11

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

model had the most stable learning curve among all. The YOLOv11 training graph shows a rapid and consistent decrease in loss values, accompanied by a gradual improvement in evaluation metrics without sharp fluctuations. This reflects that this model is able to learn efficiently and effectively in recognizing fracture patterns in annotated X-ray and MRI images. Compared to YOLOv5 which shows slow fluctuations and convergence, or YOLOv10 which although stable but has little uncertainty in some phases of training, YOLOv11 shows superior stability.

At the prediction stage, the visualization of the results showed that the YOLOv11 model was able to detect bone fractures well and consistently in various parts of the body, including the spine, forearm, and fingers. The highest confidence score was generated from YOLOv11 with an average score of 74.25%. As for the inference time and the highest FPS speed are generated from YOLOv5. Based on these results, it can be concluded that although YOLOv11 is not the fastest model, it does provide the highest confidence score, which indicates a better quality of prediction. With a combination of high confidence accuracy and still relatively good speed, YOLOv11 is the most balanced model in terms of accuracy and efficiency for automatic bone fracture detection.

The quantitative analysis in Table 3 further reinforces the prominence of YOLOv11. This model records an accuracy of 99.87% and a recall of 100%, which indicates that almost all existing fractures can be detected correctly without many errors. The mAP50 value of 99.49% is the highest among all models, and the mAP50-95 of 84.13% puts the model in very competitive performance. Although YOLOv8 is slightly higher at mAP50-95 (86.65%), overall YOLOv11 still excels in the balance between prediction accuracy and reliability. Furthermore, when compared to some of the previous studies (Table 4), YOLOv11 was shown to outperform most of the approaches that have been taken. For example, Mandibula-YOLO by Zhou et al. (2025) only achieved mAP50 of 97.02%, and YOLO NAS by Medaramatla et al. (2024) had an accuracy of 98.9%, both of which are still below the YOLOv11 achievement. Even the YOLOv10 model reported by Srinivasu et al. (2025) recorded a precision of 98%, lower than the YOLOv11 precision which almost touched the perfect figure. Taking into account all aspects, from training efficiency, prediction accuracy, to the results of quantitative comparisons and previous literature, it can be concluded that YOLOv11 is the most superior model for detecting bone fractures automatically. Consistent performance, high precision, and stability of training and prediction make YOLOv11 very potential to be implemented in deep learning-based medical decision support systems, especially in the field of orthopedic radiology.

CONCLUSION

Based on the discussions that have been conducted, it can be concluded that the YOLOv11 model shows the most superior performance compared to previous versions of YOLO (YOLOv5, YOLOv8, and YOLOv10) in the detection task of human bone fractures. In terms of training, YOLOv11 exhibits the most stable learning curve, with rapid and consistent loss reductions and significant improvement in evaluation metrics without sharp fluctuations. Predictive results on X-ray and MRI images also show that YOLOv11 is able to detect fractures with a high and consistent level of confidence, while avoiding the overconfidence and inconsistencies seen in some other models.

Quantitatively, the YOLOv11 recorded a precision of 99.87%, a recall of 100%, and an mAP50 of 99.49%, which is the highest value among the models compared. Although YOLOv8's mAP50-95 is slightly higher, YOLOv11 still strikes the best balance between accuracy, stability, and generalizability. In addition, compared to models from previous studies, YOLOv11 managed to surpass most of the performance that has been achieved in the literature, including in terms of precision and mAP.

Thus, through this comprehensive analysis, it can be affirmed that YOLOv11 is the most optimal and reliable model for automatically detecting human bone fractures. This model not only combines the advantages of previous versions of YOLO, but also presents a noticeable improvement in accuracy and consistency, making it very potential to be implemented in deep learning-based diagnosis support systems in the field of medical radiology.

This research contributes to the growing body of work on real-time, AI-assisted fracture diagnosis, and paves the way for further integration of YOLOv11 into edge-based radiology systems. Future research should include larger datasets, multimodal testing, and integration with clinical workflows.

REFERENCES

Aldi, F., Yuhandri, & Tajuddin, M. (2024). Enhanced U-Net Architecture for Glottis Segmentation with VGG-16. *JOIV : International Journal on Informatics Visualization*, *8*(4), 2173–2180. https://doi.org/10.62527/JOIV.8.4.3088

Ali, M. L., & Zhang, Z. (2024). *The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection*. https://doi.org/10.20944/PREPRINTS202410.1785.V1

Amin, R. Al, Hasan, M., Wiese, V., & Obermaisser, R. (2024). FPGA-Based Real-Time Object Detection and

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

Classification System Using YOLO for Edge Computing. *IEEE Access*, *12*, 73268–73278. https://doi.org/10.1109/ACCESS.2024.3404623

Ankan Ghosh. (2024). *Ultralytics YOLO11: Faster Than You Can Imagine!* https://learnopencv.com/wp-content/uploads/2024/10/feature.gif

Bouvard, B., Annweiler, C., & Legrand, E. (2021). Osteoporosis in older adults. *Joint Bone Spine*, *88*(3), 105135. https://doi.org/10.1016/J.JBSPIN.2021.105135

Dhahir, B. M., Hameed, I. H., & Jaber, A. R. (2017). Prospective and Retrospective Study of Fractures According to Trauma Mechanism and Type of Bone Fracture. *Research Journal of Pharmacy and Technology*, *10*(11), 3810. https://doi.org/10.5958/0974-360X.2017.00692.8

Fernandes, R., Pessoa, A., Salgado, M., De Paiva, A., Pacal, I., & Cunha, A. (2024). Enhancing Image Annotation With Object Tracking and Image Retrieval: A Systematic Review. *IEEE Access*, *12*, 79428–79444. https://doi.org/10.1109/ACCESS.2024.3406018

Islam, K., Zaheer, M. Z., Mahmood, A., & Nandakumar, K. (2024). *DiffuseMix: Label-Preserving Data Augmentation with Diffusion Models* (pp. 27621–27630).

Jeon, Y. D., Kang, M. J., Kuh, S. U., Cha, H. Y., Kim, M. S., You, J. Y., Kim, H. J., Shin, S. H., Chung, Y. G., & Yoon, D. K. (2023). Deep Learning Model Based on You Only Look Once Algorithm for Detection and Visualization of Fracture Areas in Three-Dimensional Skeletal Images. *Diagnostics 2024, Vol. 14, Page 11*, *14*(1), 11. https://doi.org/10.3390/DIAGNOSTICS14010011

Kumar, T., Mileo, A., Brennan, R., & Bendechache, M. (2023). *Image Data Augmentation Approaches: A Comprehensive Survey and Future directions*. https://doi.org/10.1109/ACCESS.2024.3470122

Liu, Y., Li, Y., Jiang, M., Wang, S., Ye, S., Walsh, S., & Yang, G. (2024). SOCR-YOLO: Small Objects Detection Algorithm in Medical Images. *International Journal of Imaging Systems and Technology* , *34*(4), e23130. https://doi.org/10.1002/IMA.23130;JOURNAL:JOURNAL:10981098;PAGE:STRING:ARTICLE/CHAPTER

McLaughlin, L., Johnstone, G., Nesbitt, L., McFadden, S. L., Hughes, C., Bond, R., & McConnell, J. (2022). *An evaluation of a training tool and study day in chest image interpretation*. ROC Events. https://hdl.handle.net/10468/16904

Medaramatla, S. C., Samhitha, C. V., Pande, S. D., & Vinta, S. R. (2024). Detection of Hand Bone Fractures in X-Ray Images Using Hybrid YOLO NAS. *IEEE Access*, *12*, 57661–57673. https://doi.org/10.1109/ACCESS.2024.3379760

Meza, G., Ganta, D., & Gonzalez Torres, S. (2024). Deep Learning Approach for Arm Fracture Detection Based on an Improved YOLOv8 Algorithm. *Algorithms 2024, Vol. 17, Page 471*, *17*(11), 471. https://doi.org/10.3390/A17110471

Nigatu, A. M., Yilma, T. M., Gezie, L. D., Gebrewold, Y., Gullslett, M. K., Mengiste, S. A., & Tilahun, B. (2023). Medical imaging consultation practices and challenges at public hospitals in the Amhara regional state, Northwest Ethiopia: a descriptive phenomenological study. *BMC Health Services Research*, *23*(1), 1–15. https://doi.org/10.1186/S12913-023-09652-9/FIGURES/1

Parvin, S., & Rahman, A. (2024). A real-time human bone fracture detection and classification from multi-modal images using deep learning technique. *Applied Intelligence*, *54*(19), 9269–9285. https://doi.org/10.1007/S10489-024-05588-7/METRICS

Ragab, M. G., Abdulkadir, S. J., Muneer, A., Alqushaibi, A., Sumiea, E. H., Qureshi, R., Al-Selwi, S. M., & Alhussian, H. (2024). A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *IEEE Access*, *12*, 57815–57836. https://doi.org/10.1109/ACCESS.2024.3386826

Ragnarsson, K. T. (2015). Bone loss and fractures in limbs paralyzed by spinal cord injury: Prevention, diagnosis, and treatment. *Journal of Spinal Cord Medicine*, *38*(1), 10–12. https://doi.org/10.1179/2045772314Y.0000000200;CTYPE:STRING:JOURNAL

Regnard, N. E., Lanseur, B., Ventre, J., Ducarouge, A., Clovis, L., Lassalle, L., Lacave, E., Grandjean, A., Lambert, A., Dallaudière, B., & Feydy, A. (2022). Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays. *European Journal of Radiology*, *154*, 110447. https://doi.org/10.1016/J.EJRAD.2022.110447

Sasmito, P., innekerisnawati, D., Hutasoid, D. L., Rafi, R., Aprila, R., & cindy, A. A. (2023). HEALTH EDUCATION ON FRACTURES EMERGENCIES TO INCREASE KNOWLEDGE AMONG SMK KESEHATAN LETRIS INDONESIA 1 STUDENTS. *Jurnal Pengabdian Kolaborasi Dan Inovasi IPTEKS* , *1*(6), 905–908. https://doi.org/10.59407/JPKI2.V1I6.194

Srinivasu, P. N., Kumari, G. L. A., Narahari, S. C., Ahmed, S., & Alhumam, A. (2025). Exploring the impact of

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 7, Number 3, July 2025
https://doi.org/10.47709/cnahpc.v7i3.6051

Submitted : Jun 9, 2025
Accepted : Jun 28, 2025
Published : Jul 05, 2025

hyperparameter and data augmentation in YOLO V10 for accurate bone fracture detection from X-ray images. *Scientific Reports*, *15*(1), 1–24. https://doi.org/10.1038/S41598-025-93505-4;SUBJMETA=1042,114,117,631,639,705;KWRD=COMPUTATIONAL+BIOLOGY+AND+BIOINFORMATICS,COMPUTATIONAL+SCIENCE,COMPUTER+SCIENCE

Tao, H., Zhang, L., Sun, Z., Cui, X., & Yi, W. (2025). Multi-Strategy Improvement of Coal Gangue Recognition Method of YOLOv11. *Sensors 2025, Vol. 25, Page 1983*, *25*(7), 1983. https://doi.org/10.3390/S25071983

Tian, J., Li, H., Qi, Y., Wang, X., & Feng, Y. (2024). Intelligent medical detection and diagnosis assisted by deep learning. *Applied and Computational Engineering*, *64*(1), 120–125. https://doi.org/10.54254/2755-2721/64/20241356

Walston, S. L., Seki, H., Takita, H., Mitsuyama, Y., Sato, S., Hagiwara, A., Ito, R., Hanaoka, S., Miki, Y., & Ueda, D. (2024). Data set terminology of deep learning in medicine: a historical review and recommendation. *Japanese Journal of Radiology*, *42*(10), 1100–1109. https://doi.org/10.1007/S11604-024-01608-1/METRICS

Wei, W., Huang, Y., Zheng, J., Rao, Y., Wei, Y., Tan, X., & OuYang, H. (2025). YOLOv11-based multi-task learning for enhanced bone fracture detection and classification in X-ray images. *Journal of Radiation Research and Applied Sciences*, *18*(1), 101309. https://doi.org/10.1016/J.JRRAS.2025.101309

Zhao, Q. (2025). *An Improved Yolov11 Architecture with Multi-Scale Attention and Spatial Fusion for Fine-Grained Residual Detection*. https://doi.org/10.2139/SSRN.5263718

Zhou, T., Wang, H., Chen, K. X., Zhang, Z., Chai, W., & Lu, H. (2025). Mandible-YOLO: The fracture region is detected only once. *Biomedical Signal Processing and Control*, *106*, 107724. https://doi.org/10.1016/J.BSPC.2025.107724

Zou, J., & Arshad, M. R. (2024). Detection of whole body bone fractures based on improved YOLOv7. *Biomedical Signal Processing and Control*, *91*, 105995. https://doi.org/10.1016/J.BSPC.2024.105995

* Corresponding author