

Architecture-Dependent Effects of CLAHE Enhancement Across YOLOv5, YOLOv8, YOLOv10, and YOLOv11 for Bone Fracture Detection

Febri Aldi¹⁾, Irohito Nozomi²⁾, Ronaldo Syahputra³⁾

^{1,2,3)} Universitas Putra Indonesia YPTK Padang

¹ febri_aldi@upiyptk.ac.id, ² irohito_nozomi@upiyptk.ac.id, ³ ronaldo_syahputra@upiyptk.ac.id

ABSTRACT

Bone fracture detection in X-ray images remains challenging because fracture lines often appear as low-contrast, subtle, and visually similar patterns across fracture types. Although YOLO-based detectors have been widely used for medical object detection, the effect of contrast enhancement is commonly evaluated on a single architecture, making it unclear whether preprocessing benefits are consistent across different YOLO generations. This study investigates the architecture-dependent effect of Contrast Limited Adaptive Histogram Equalization (CLAHE) on YOLOv5, YOLOv8, YOLOv10, and YOLOv11 for bone fracture detection. A total of 1,539 annotated X-ray images were prepared in YOLO bounding-box format and evaluated under two scenarios: original images and CLAHE-enhanced images. Model performance was assessed using precision, recall, mAP50, and mAP50-95, followed by paired architecture-level comparison using paired t-test, Wilcoxon signed-rank test, and bootstrap confidence intervals. The results show that CLAHE does not uniformly improve all detection metrics. YOLOv8 without CLAHE achieved the strongest mAP50 and recall, whereas YOLOv11 with CLAHE produced the highest mAP50-95, indicating better localization precision under stricter IoU thresholds. The statistical comparison suggests a positive but exploratory improvement in mAP50-95 after CLAHE, while other metrics showed no significant architecture-level difference. These findings demonstrate that image enhancement effectiveness is architecture-dependent and should be selected according to the feature extraction and localization characteristics of the detector rather than applied as a universal preprocessing step.

Keywords: Bone fracture detection; CLAHE; YOLO; X-Ray image; object detection

INTRODUCTION

The development of artificial intelligence and computer vision technology has made a significant contribution in the medical field (Aldi et al., 2024), especially in the analysis of radiological images. One of the rapidly growing applications is the use of deep learning to help the disease diagnosis process through medical imaging such as X-rays, CT-Scans, and MRI (Aldi et al., 2025; Olveres et al., 2021). This technology allows computer systems to recognize certain patterns automatically so that they can assist medical personnel in improving the speed and accuracy of diagnosis. In recent years, deep learning-based object detection methods, especially the You Only Look Once (YOLO) family, have become one of the most widely used approaches because they are able to detect objects quickly and accurately in real-time (Cai et al., 2025).

One of the medical cases that has been widely researched using the computer vision approach is the detection of bone fractures on X-ray images (Belali et al., 2025). Bone fractures are conditions of damage or cracks in bone structures that can occur due to accidents, sports injuries, or other physical trauma (Oryan et al., 2015). Fracture examination is generally performed using X-ray radiography because this method is faster, cheaper, and more accessible than other imaging modalities (Joshi, 2022). However, the process of interpreting X-ray images still has challenges, especially in cases of low-contrast fractures, thin cracks, or complex bone positions. Misinterpretation can lead to a delay in diagnosis and potentially increase the risk of complications in patients (R. Y. Ju & Cai, 2023).

Various previous studies have applied deep learning methods to automatically detect bone fractures

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

using Convolutional Neural Network (CNN) architecture and YOLO-based object detection. YOLO architectures such as YOLOv5 and YOLOv8 show quite good performance in detecting fracture areas in medical images (Shaban et al., 2024). Research by Ju and Cai using YOLOv8 on the GRAZPEDWRI-DX dataset showed that the YOLOv8 model was able to achieve state-of-the-art performance in fracture detection in pediatric X-ray images (R. Y. Ju & Cai, 2023). In addition to the development of model architectures, several studies have also begun to apply preprocessing enhancement techniques to improve the quality of X-ray images before the training process is carried out. One of the most widely used enhancement methods is the Contrast Limited Adaptive Histogram Equalization CLAHE, which functions to increase the local contrast in the image so that the details of the fracture area become clearer (Shaban et al., 2024).

Nonetheless, previous research still has some limitations. Most studies have focused only on improving the performance of one particular YOLO architecture without conducting a comparative analysis of the most recent generations of YOLO. In addition, few studies specifically analyze the effect of CLAHE preprocessing on the performance of various YOLO architectures on bone fracture detection tasks. In fact, each deep learning architecture has different feature extraction characteristics so that the response to image enhancement can also be different (Dritsas & Trigka, 2025). This condition causes the effectiveness of CLAHE in various YOLO models to remain unclear (Cai et al., 2025; R. Y. Ju, Chien, Lin, et al., 2024).

The lack of comparative analysis of preprocessing enhancement and object detection architecture can lead to the selection of less than optimal methods in the development of computer vision-based automatic diagnosis systems (Wu et al., 2025). If enhancement preprocessing is applied without proper evaluation, then the model's performance may decrease due to changes in the distribution of image characteristics. In addition, low accuracy in fracture detection can increase the risk of misdiagnosis, delay in medical treatment, and decrease the effectiveness of clinical decision support systems (R. Y. Ju, Chien, & Chiang, 2024).

Based on these problems, this study proposes a comparative analysis of several YOLO architectures for the detection of bone fractures in X-ray images using CLAHE preprocessing. The architectures used include YOLOv5, YOLOv8, YOLOv10, and YOLOv11. This study aims to analyze the effect of CLAHE preprocessing on the performance of each model based on the precision, recall, mAP50, and mAP50-95 metrics. In addition, this study also aims to determine the most effective YOLO architecture in detecting bone fractures on X-ray images. It is hoped that the results of this study can contribute to the development of a more accurate automatic fracture detection system and can support the medical diagnosis process based on artificial intelligence.

LITERATURE REVIEW

The development of deep learning in the field of medical image analysis has prompted many studies related to the automatic detection of bone fractures using X-ray images (Meena & Roy, 2022). Various approaches have been developed, ranging from CNN-based methods to modern object detection based on the YOLO family. Previous studies have shown that the application of deep learning can improve the efficiency and accuracy of fracture detection compared to conventional manual approaches (Su et al., 2023). However, there are still challenges related to localization accuracy, model generalization, and the effect of image preprocessing on detection performance. Table 1 shows a summary of previous research related to bone fracture detection using deep learning methods and YOLO architecture.

Table 1. Previous Research

N o.	Researcher	Methods/Models	Main Finding	Limitation	Research Gap / Contribution
1	(Rutledge et al., 2023)	Machine Learning for Mandibular Fracture Detection	Achieved useful precision, recall, and mAP values.	Limited variability and generalization across fracture types.	Shows feasibility of automated fracture detection, but motivates stronger localization-oriented detectors.
2	(Ahmed et al., 2024) [2]	YOLOv5, YOLOv6, YOLOv7, YOLOv8	YOLOv8m achieved high sensitivity and mAP for wrist abnormality detection.	Focuses on YOLO comparison without evaluating contrast enhancement effects.	Supports multi-generation YOLO comparison as a relevant research direction.
3	(R.-Y. Ju & Cai, 2023)	YOLOv8	YOLOv8 improved pediatric fracture detection accuracy.	Model optimization is emphasized more than preprocessing interaction.	Shows strong YOLO performance, but leaves enhancement-architecture interaction unresolved.
4	(Chien et al., 2024) [4]	YOLOv9	Improved mAP50-95 and localization capability.	Does not compare CLAHE across multiple YOLO generations.	Highlights the importance of strict localization metrics.
5	(Zhang & Ji, 2025) [5]	YOLOv11 + Attention Mechanism	Improved precision, recall, and mAP in medical X-ray fracture detection.	Focuses on architectural modification rather than preprocessing effect.	Indicates that newer YOLO models may benefit from feature refinement.
6	(Hidayat et al., 2026) [6]	YOLOv8 + CLAHE	CLAHE-supported YOLOv8 achieved high mAP@0.5.	Uses a limited architectural comparison.	Motivates testing whether CLAHE benefit is architecture-dependent.
7	(Priyanka et al., 2025) [7]	YOLOv5 + CLAHE + SVM	Preprocessing improved hand fracture detection accuracy.	Combines methods but does not isolate cross-YOLO effects.	Supports the relevance of CLAHE while leaving comparative YOLO analysis open.
8	Proposed Research	YOLOv5, YOLOv8, YOLOv10, YOLOv11 + CLAHE	Evaluates detection accuracy, localization strictness, and visual behavior.	Limited to one dataset and one enhancement strategy.	Contributes architecture-aware evidence for selecting CLAHE in fracture detection.

The studies summarized in Table 1 indicate that YOLO-based detectors have achieved promising performance on fracture detection tasks, particularly in pediatric wrist and hand fracture datasets. However, most prior works focus on improving a selected model or adding attention mechanisms, while the relationship between preprocessing and architecture design remains underexplored. This limits the ability to determine whether the performance gain comes from the detector architecture, the enhancement process, or the interaction between both.

A second pattern in the literature is the inconsistent treatment of preprocessing. CLAHE has been reported to improve image visibility and detection performance in several studies, yet those studies do not sufficiently compare whether the same enhancement benefits older and newer YOLO generations equally. This is important because each YOLO version has different backbone, neck, feature aggregation, and localization mechanisms, which may respond differently to enhanced texture and contrast.

A third limitation is that previous studies often report final evaluation metrics without deeper interpretation of class-level errors, false positives, false negatives, or localization quality. In medical object detection, mAP50 alone may be insufficient because a model can detect the general fracture region while still producing less precise bounding boxes. Therefore, mAP50-95 and confusion matrix interpretation are needed to evaluate stricter localization behavior.

Based on this synthesis, the unresolved gap is not merely the absence of YOLO comparison, but the lack of architecture-aware analysis of preprocessing effects in fracture detection. This study addresses that gap by comparing YOLOv5, YOLOv8, YOLOv10, and YOLOv11 using original and CLAHE-enhanced X-ray images, followed by metric-level, statistical, confusion-matrix, and visual interpretation.

* Corresponding author



METHOD

This study uses a comparative experimental approach to analyze the performance of several YOLO architectures in detecting bone fractures in X-ray images. The proposed method focuses on the effect of CLAHE preprocessing on improving fracture detection performance. The research stages include dataset collection, image preprocessing, data augmentation, dataset sharing, model training, performance evaluation, statistical analysis, and comparative analysis.

Research Framework

The research framework in this study is shown in Figure 1. The research began with the collection of X-ray imagery datasets, then preprocessing was carried out using resize and CLAHE enhancement techniques. Furthermore, the image of the preprocessing results goes through the data augmentation stage to improve the generalization ability of the model. The dataset is then divided into training, validation, and testing data. Some YOLO architectures, namely YOLOv5, YOLOv8, YOLOv10, and YOLOv11 are trained and evaluated using several performance metrics. The results of the experiment were then analyzed comparatively to determine the influence of CLAHE and the best model in detecting bone fractures.



Figure 1. Research Framework

Dataset Collection

The dataset used in this study is in the form of bone X-ray images containing several types of fractures. The dataset that we have collected is 1,539 images through the annotation process using the YOLO bounding box format for object detection needs. The fracture categories in the dataset include: Comminuted, Greenstick, Linear, Oblique Displaced, Oblique, Segmental, Spiral, Transverse Displaced, Transverse. The dataset has been divided into three subsets, namely: Training data of 70%, Data validation of 20%, and Data testing of 10%. The distribution of the dataset was carried out to ensure that the training and evaluation process of the model ran optimally.

Image Preprocessing

The preprocessing stage is carried out to improve the quality of X-ray images and clarify fracture areas before the model training process.

Resize Image

All X-ray images were resized to 640×640 pixels to have a uniform input size across all YOLO models.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

CLAHE Enhancement

This study applied the CLAHE method to increase the local contrast of X-ray images without exaggerating the noise excessively (Koonsanit et al., 2017). CLAHE helps clarify areas of fractures that have low contrast on medical images (by Muhammad Shahrul Zaim Ahmad). The CLAHE parameters used are Clip Limit = 2.0 and Tile Grid Size = 8×8 . CLAHE is applied to the grayscale image before the data augmentation stage is performed.

Data Augmentation

The data augmentation stage was carried out to increase the variety of the dataset and reduce the risk of overfitting during the model training process (Shorten & Khoshgoftaar, 2019). The augmentation techniques used include Flip (Horizontal and Vertical 50% probability each), Rotation (Random rotation between -5° and $+5^\circ$), Shear (Random shear of $\pm 2^\circ$ horizontally and $\pm 2^\circ$ vertically), Saturation (Random adjustment between -5% and +5%, Brightness (Random adjustment between -10% and +10%), and Zooming 2%.

YOLO Comparative Model Training

This study compares four YOLO architectures, namely YOLOv5, YOLOv8, YOLOv10, YOLOv11. Each model was tested using two experimental scenarios, namely the original dataset without CLAHE preprocessing and the dataset with CLAHE preprocessing. A comparative experiment was conducted to analyze the influence of CLAHE on the performance of each YOLO architecture in detecting bone fractures.

Model Training Configuration

The entire experiment was conducted using Google Colab with the support of Tesla's T4 GPU. The model training configuration is shown in Table 2.

Table 2. Model Training Configuration

Parameter	Value
Image size	640×640 pixels
Epoch	100
Batch size	64
Optimizer	Default optimizer of each YOLO implementation
Device	Tesla T4 GPU on Google Colab
Label format	YOLO bounding-box annotation format
Pretrained weights	Built-in pretrained weights for transfer learning
CLAHE clip limit	2.0
CLAHE tile grid size	8×8
Random seed	42

The annotation process was validated through visual rechecking of bounding-box position, class-label consistency, and object coverage. Images with unclear or inconsistent annotations were corrected before training. A fixed random seed of 42 was used to reduce run-to-run variation in data loading and training initialization.

Model Evaluation

The performance of the model was evaluated using several object detection metrics, namely Precision in equation 1, Recall in equation 2, Mean Average Precision at IoU 0.5 (mAP50) in equation 4, Mean Average Precision at IoU 0.5–0.95 (mAP50-95) in equation 5.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Where TP (True Positive) is the number of fracture objects that were detected correctly, FP (False Positive) is the number of objects that were falsely detected as fractures, and FN (False Negative) and the number of fracture objects that failed to be detected by the model. The Average Precision (AP) value is calculated from the area under the Precision-Recall curve. The AP equation is written in equation 3.

$$AP = \int_0^1 Precision(Recall) d(Recall) \quad (3)$$

Furthermore, the mAP50 value is calculated using the average AP of all classes at the IoU threshold of 0.5.

$$mAP_{50} = \frac{1}{C} \sum_{i=1}^C AP_i^{IoU=0.5} \quad (4)$$

Where C is the sum of the object class, and $AP_i^{IoU=0.5}$ is the Average Precision value of the i class at the IoU threshold of 0.5. Meanwhile, mAP50-95 is a stricter evaluation metric than mAP50 because it calculates average AP at several IoU thresholds ranging from 0.5 to 0.95 with intervals of 0.05.

$$mAP_{50:95} = \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{C} \sum_{i=1}^C AP_i^{IoU_t} \right) \quad (5)$$

Where T is the threshold sum of IoU, and $AP_i^{IoU_t}$ is the i th class Average Precision value at a given IoU threshold. In this study, the IoU threshold used in mAP50-95 is in the range 0.50, 0.55, 0.60, ..., 0.95. The mAP50-95 metric is used to evaluate the model's localization capabilities in more detail because it involves a higher and stricter IoU threshold than mAP50.

Comparative Analysis

The final stage of the study was carried out by comparing the results of experiments across YOLO architectures to analyze the influence of CLAHE on fracture detection performance, determine the best YOLO architecture, evaluate the relationship between preprocessing enhancement and detection accuracy. Comparative analysis was performed based on the precision, recall, mAP50, and mAP50-95 values of each model.

RESULT

This section presents the results of experiments and performance analysis of several YOLO architectures in detecting bone fractures in X-ray images using the original dataset and dataset with CLAHE preprocessing. Evaluation was conducted using precision, recall, mAP50, and mAP50-95 metrics to measure the detection and localization capabilities of each model.

In addition, this study also analyzes the effect of CLAHE preprocessing on the performance of each YOLO architecture through comparison of experimental results and statistical analysis. The results of the study are visualized in the form of training performance graphs, comparative tables, and sample detection results to provide a more comprehensive understanding of the performance of the proposed model.

Training Performance

In this study, the visualization of training performance was focused on two representative models, namely YOLOv8 without CLAHE preprocessing and YOLOv11 with CLAHE preprocessing. The selection of the two models was based on the results of the final experiment, where YOLOv8 without CLAHE produced the highest mAP50 value, while YOLOv11 with CLAHE produced the highest mAP50-95 value compared to the other scenarios.

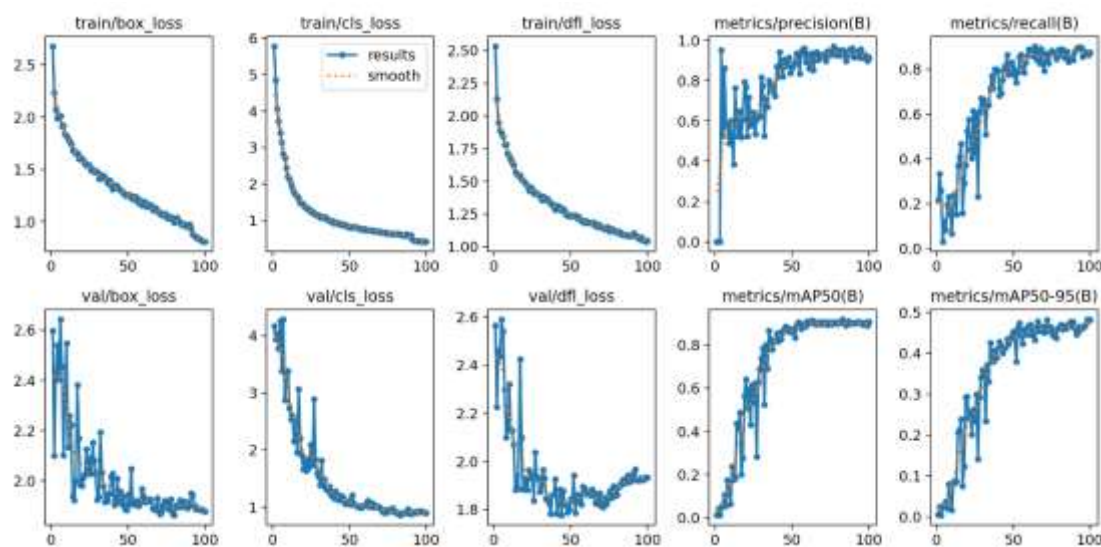


Figure 2. Original YOLOv8 training results

Figure 2 shows the training performance of the YOLOv8 model without CLAHE preprocessing. The train loss values consisting of box loss, classification loss, and distribution focal loss decreased consistently during the training process, which shows that the model is able to perform the learning process well.

In addition, validation loss also shows a stable pattern with no indication of significant overfitting. The precision and recall values gradually increased until they reached around 0.97 and 0.90. The value of the mAP50 increased significantly in the initial epoch and began to stabilize after around the 50th epoch with the final performance above 0.91. These results show that YOLOv8 is able to effectively study fracture characteristics in the original X-ray image without the need for additional enhancement.

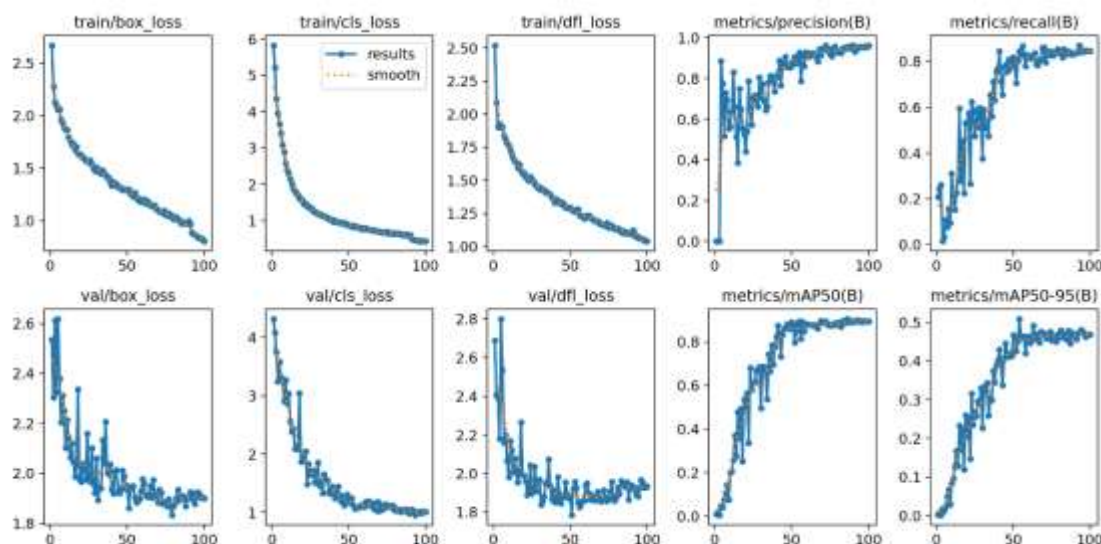


Figure 3. YOLOv11 + CLAHE training results

Figure 3 shows the training performance of the YOLOv11 model using CLAHE preprocessing. The train loss and validation loss graphs experienced a steady decline during the training process, which shows that the training process is going well.

The model produced a high precision value close to 0.96 and obtained the highest mAP50-95 value compared to all experimental scenarios. These results suggest that CLAHE preprocessing helps improve the model's localization ability to detect fracture areas in X-ray images, especially in more rigorous IoU evaluations.

Comparative evaluation

A comparative evaluation was conducted to compare the performance of each YOLO architecture on the original dataset and the CLAHE preprocessing. Performance comparisons were conducted using precision, recall, mAP50, and mAP50-95 metrics as shown in Table 3.

Table 3. Comparison of Model Training Results

Model	CLAHE	Precision	Recall	mAP50	mAP50-95
YOLOv5	No	0.95463	0.85636	0.90778	0.48255
YOLOv5	Yes	0.97912	0.8687	0.91579	0.4839
YOLOv8	No	0.97144	0.90329	0.91842	0.48295
YOLOv8	Yes	0.97097	0.87344	0.90816	0.49113
YOLOv10	No	0.93422	0.85377	0.88578	0.46914
YOLOv10	Yes	0.94379	0.86788	0.88617	0.48796
YOLOv11	No	0.96419	0.87255	0.9186	0.48408
YOLOv11	Yes	0.96298	0.87057	0.90465	0.50875

Table 4. Architecture-Level Statistical Comparison of CLAHE Effects

Metric	Mean Δ (CLAHE- Original)	Bootstrap 95% CI	Paired t-test p	Wilcoxon p	Interpretation
Precision	+0.00809	[-0.00084, 0.01825]	0.270	0.625	Small positive change, not significant
Recall	-0.00135	[-0.01930, 0.01322]	0.903	1.000	No consistent change
mAP50	-0.00395	[-0.01211, 0.00420]	0.488	0.625	No consistent improvement at IoU 0.5
mAP50-95	+0.01326	[0.00477, 0.02175]	0.085	0.125	Exploratory positive localization trend

The paired analysis shows that CLAHE produced the clearest positive effect on mAP50-95, with a mean increase of 0.01326 and a bootstrap confidence interval above zero. However, the paired t-test and Wilcoxon p-values did not reach the conventional 0.05 significance level because the comparison used only four architecture-level pairs. Therefore, the statistical result should be interpreted as an exploratory indication that CLAHE may improve stricter localization quality, while repeated training or cross-validation is needed for stronger statistical confirmation.

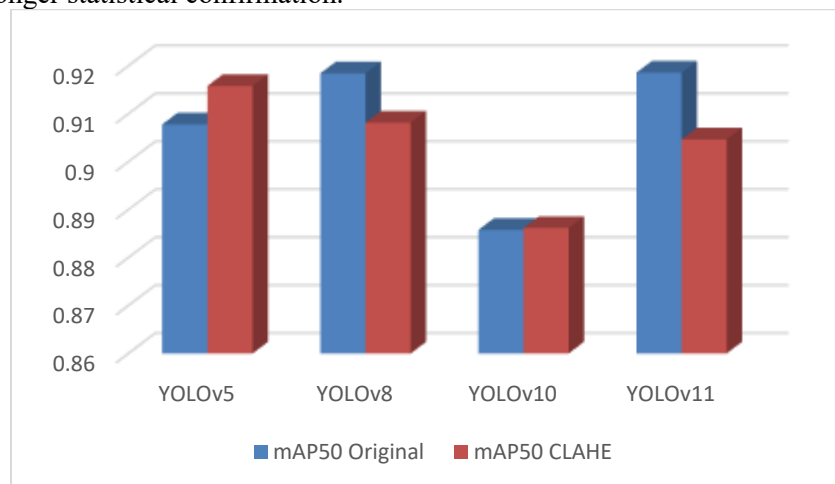


Figure 3. Comparison of the original mAP50 vs CLAHE

Figure 3 shows the comparison of mAP50 values between the original dataset and the CLAHE dataset across the entire YOLO architecture. Based on the graph, CLAHE preprocessing provides improved performance on YOLOv5 and YOLOv10, but has decreased performance on YOLOv8 and YOLOv11.

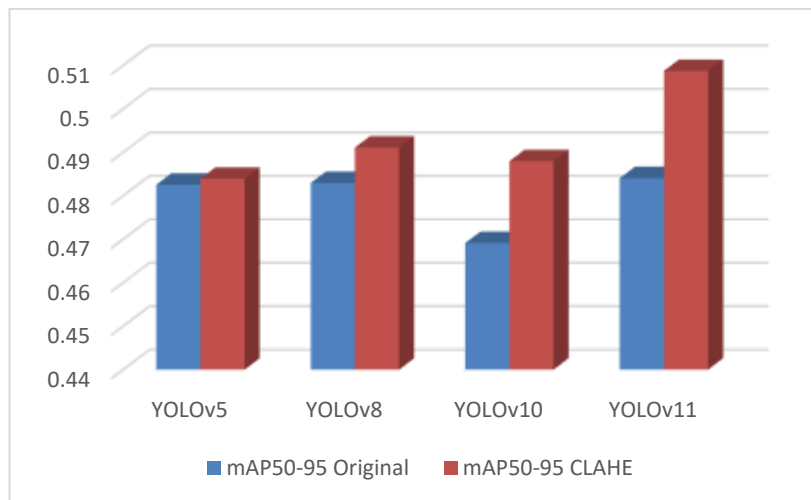


Figure 4. Comparison of mAP50-95 original vs CLAHE

Figure 4 shows the comparison of mAP50-95 values across all experimental scenarios. The visualization results showed that CLAHE preprocessing tended to increase the mAP50-95 value in almost all models, especially in YOLOv11 which experienced the most significant increase.

Confusion Matrix Analysis

The confusion matrix was used to analyze the model's ability to classify each type of fracture in the test dataset. The confusion matrix visualization provides information about the number of correct predictions, misclassification between classes, and the distribution of false positives and false negatives in each fracture category.

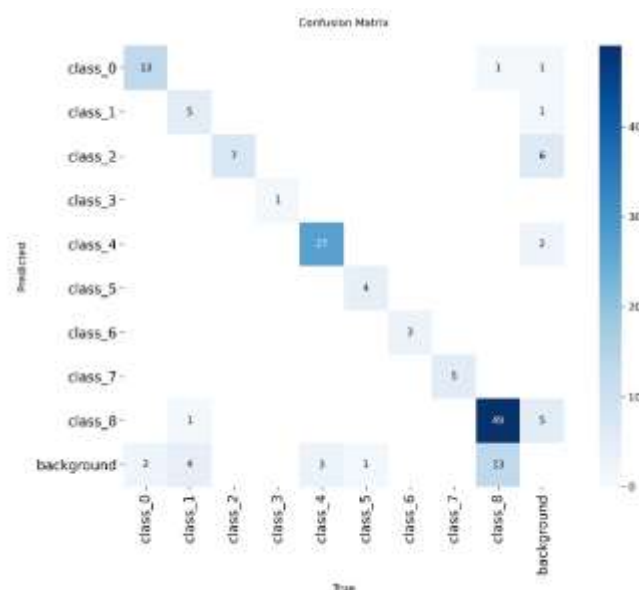


Figure 5. Confusion matrix YOLOv11 original

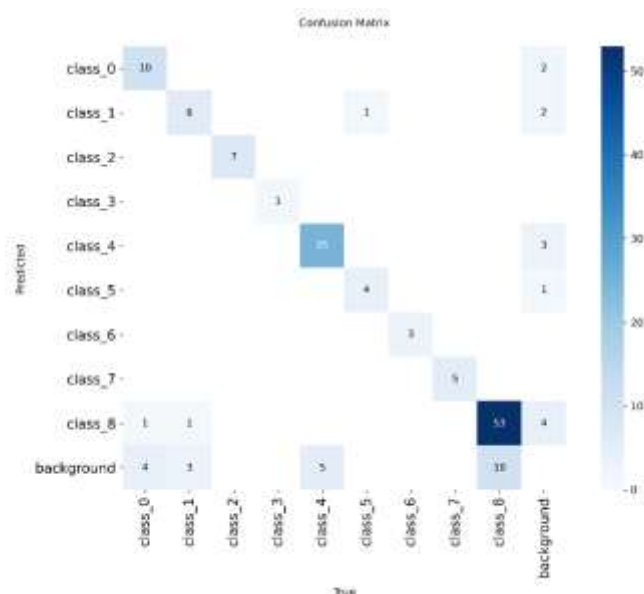


Figure 6. Confision matrix YOLOv11 CLAHE

Figure 5 shows the confusion matrix of the YOLOv11 model without CLAHE preprocessing, while Figure 6 shows the confusion matrix of the YOLOv11 model with CLAHE preprocessing. Both models were chosen for their best performance on key evaluation metrics, specifically mAP50 and mAP50-95.

Detection Result

Visualization of detection results was carried out to evaluate the model's ability to localize fracture areas in X-ray images visually. In this study, the detection visualization was focused on the YOLOv11 model without CLAHE preprocessing and YOLOv11 with CLAHE preprocessing because both models showed the best performance in the previous evaluation results.



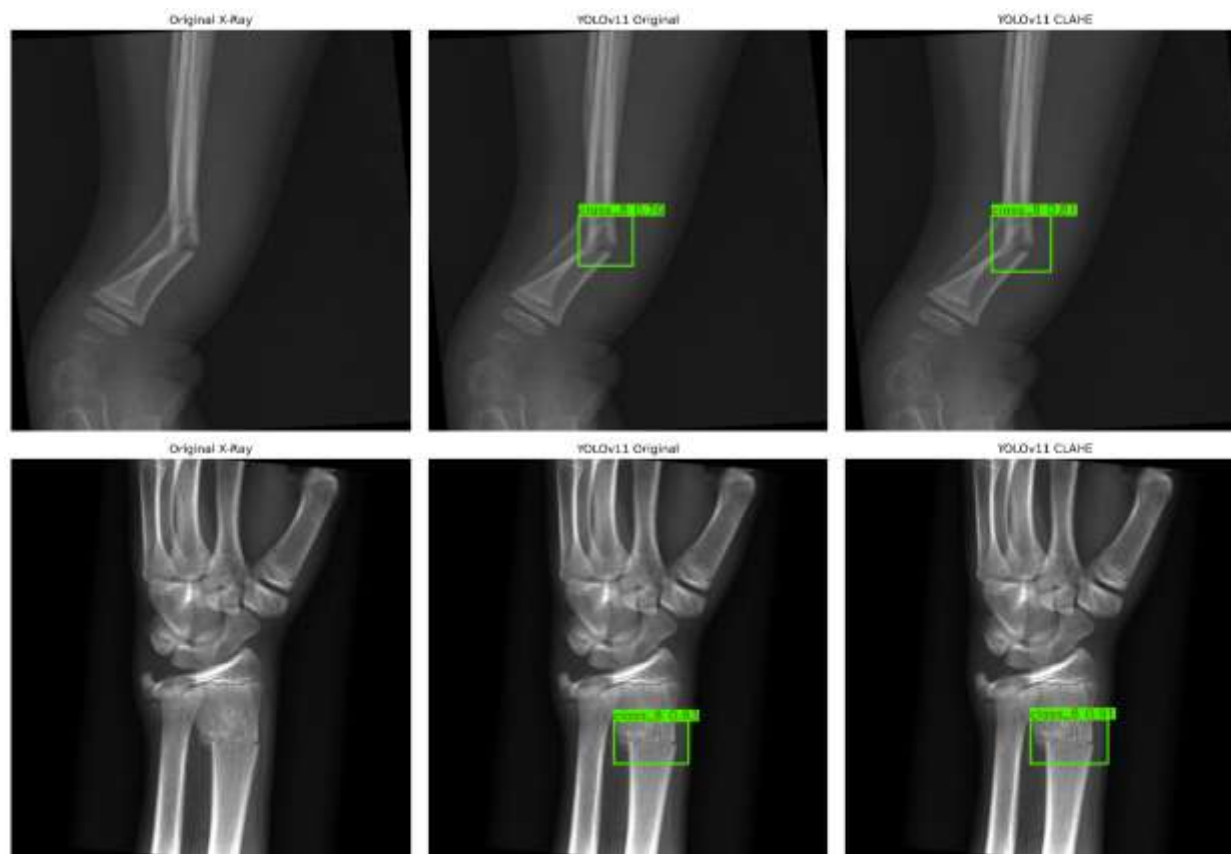


Figure 7. Detection results of the original YOLOv11 vs CLAHE model

Figure 7 shows a comparison of detection results between the original image, the detection results of YOLOv11 without CLAHE, and the results of the detection of YOLOv11 with CLAHE in several fracture X-ray image samples.

DISCUSSIONS

The experimental results show that all YOLO architectures were able to detect fracture regions with relatively high performance, but the effect of CLAHE differed across models and evaluation metrics. This finding indicates that preprocessing enhancement should not be evaluated only based on final accuracy values, but should also be interpreted in relation to architecture design, feature extraction behavior, and localization requirements.

YOLOv8 without CLAHE achieved the best mAP50 and recall. This suggests that YOLOv8 can learn discriminative fracture features directly from the original X-ray images when fracture textures are already sufficiently visible. In this condition, additional contrast enhancement may not be necessary and may slightly disturb the natural intensity distribution learned by pretrained features.

In contrast, YOLOv11 with CLAHE achieved the highest mAP50-95. Since mAP50-95 evaluates localization consistency across multiple IoU thresholds, this result indicates that CLAHE contributes more clearly to bounding-box precision than to general detection performance. By improving local contrast around fracture boundaries, CLAHE may help YOLOv11 refine localization when fracture regions are thin, low contrast, or partially blended with surrounding bone texture.

The architecture-dependent effect of CLAHE can be explained by differences in backbone design, feature aggregation, receptive field, and localization mechanisms. Models with feature extractors that are

* Corresponding author



already sensitive to subtle texture may gain limited benefit from CLAHE, while models that rely more on clearer edge contrast may show improved localization. Therefore, the same preprocessing technique may improve one YOLO generation while reducing or only minimally affecting another.

The decrease of mAP50 in some CLAHE scenarios also indicates that enhanced local contrast can introduce additional visual patterns. In medical X-ray images, these patterns may be useful when they emphasize fracture lines, but they may also amplify irrelevant trabecular or anatomical textures. As a result, the model may produce higher localization precision in certain cases while losing some general detection stability.

The confusion matrix analysis supports this interpretation. Misclassification is more likely to occur in fracture types with similar morphology, such as Greenstick and Linear or Spiral and Oblique. These errors are not caused only by model weakness, but also by visual overlap between fracture class definitions. CLAHE can make crack boundaries clearer, but it cannot fully separate classes whose anatomical patterns are intrinsically similar.

Detection visualization also shows that CLAHE is most useful for low-contrast fracture regions. In samples with unclear crack boundaries, enhanced local contrast supports more precise bounding boxes and sometimes improves confidence scores. However, in already clear images, the gain is limited because additional enhancement does not add meaningful structural information.

Overall, this study demonstrates that CLAHE is an architecture-dependent preprocessing strategy. Its strongest contribution is not universal improvement across all metrics, but improved localization strictness in selected architectures, especially YOLOv11. This finding is important for medical object detection because precise localization is essential for interpreting fracture position and reducing ambiguous detection results.

The practical implication is that preprocessing selection should be validated for each detector rather than assumed beneficial by default. For systems prioritizing high recall at IoU 0.5, YOLOv8 without CLAHE may be preferable. For systems prioritizing stricter localization, YOLOv11 with CLAHE provides a stronger option based on the mAP50-95 result.

This study has several limitations. First, the evaluation used one dataset, so generalization across imaging devices, hospitals, and patient populations still needs further validation. Second, the comparison was limited to four YOLO architectures and one preprocessing method. Third, the statistical analysis was based on architecture-level pairs rather than repeated runs or cross-validation, which limits inferential strength.

Future work should evaluate adaptive CLAHE parameter selection, transformer-based detectors, attention-enhanced YOLO variants, and explainable AI methods to identify which image regions influence fracture detection. Multi-dataset evaluation and repeated training are also needed to confirm the robustness of the architecture-dependent preprocessing effect.

CONCLUSION

This study compared YOLOv5, YOLOv8, YOLOv10, and YOLOv11 for bone fracture detection on X-ray images using original and CLAHE-enhanced datasets. The results show that all models achieved competitive detection performance, but the effect of CLAHE was not uniform across architectures and metrics.

The main scientific contribution of this study is the finding that CLAHE enhancement is architecture-dependent. YOLOv8 without CLAHE produced the strongest mAP50 and recall, while YOLOv11 with CLAHE achieved the highest mAP50-95. This indicates that CLAHE contributes more clearly to stricter localization precision than to general detection accuracy at IoU 0.5.

The confusion matrix and detection visualization further show that errors mainly occur in fracture classes with similar visual morphology, such as Greenstick versus Linear and Spiral versus Oblique.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

CLAHE can clarify low-contrast fracture lines and improve bounding-box quality in selected samples, but it cannot fully resolve class overlap caused by similar anatomical patterns.

This study is limited by the use of one dataset, four YOLO architectures, and one preprocessing method. In addition, the statistical analysis was exploratory because it used one final score per architecture rather than repeated experimental runs. Therefore, the generalizability of the findings should be strengthened through multi-dataset evaluation and repeated experiments.

Future research should evaluate larger and more diverse datasets, adaptive CLAHE parameters, additional preprocessing methods, transformer-based object detectors, attention-enhanced YOLO models, and explainable AI techniques. These directions can strengthen understanding of how preprocessing, architecture design, and fracture morphology interact in medical object detection systems.

REFERENCES

- Ahmed, A., Imran, A. S., Manaf, A., Kastrati, Z., & Daudpota, S. M. (2024). Enhancing wrist abnormality detection with YOLO: Analysis of state-of-the-art single-stage detection models. *Biomedical Signal Processing and Control*, 93, 106144. <https://doi.org/10.1016/J.BSPC.2024.106144>
- Aldi, F., Nozomi, I., Hafizh, M., & Novita, T. (2025). Comparative Analysis of YOLOv11 with Previous YOLO in the Detection of Human Bone Fractures. *Journal of Computer Networks, Architecture and High Performance Computing*, 7(3), 777–790. <https://doi.org/10.47709/CNAHPC.V7I3.6051>
- 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>
- Belali, M. H., Khan, F. R., Rajin, N. M., Sarkar, H., & Uddin, K. M. M. (2025). Bone Osteoporosis Fractures Detection with Deep Learning: An X-ray Image Analysis Approach. *Biomedical Materials & Devices* 2025 4:1, 4(1), 533–551. <https://doi.org/10.1007/S44174-025-00285-6>
- Cai, Z., Zhou, K., & Liao, Z. (2025). A Systematic Review of YOLO-Based Object Detection in Medical Imaging: Advances, Challenges, and Future Directions. *Computers, Materials & Continua*, 85(2), 2255–2303. <https://doi.org/10.32604/CMC.2025.067994>
- Chien, C. T., Ju, R. Y., Chou, K. Y., & Chiang, J. S. (2024). YOLOv9 for fracture detection in pediatric wrist trauma X-ray images. *Electronics Letters*, 60(11), e13248. <https://doi.org/10.1049/ELL2.13248>; REQUESTEDJOURNAL:JOURNAL:1350911X; WEBSITE:W EBSITE:IETRESEARCH; WGROUP:STRING:PUBLICATION
- Dritsas, E., & Trigka, M. (2025). Machine Learning in Intelligent Networks: Architectures, Techniques, and Use Cases. *IEEE Access*, 13, 102724–102745. <https://doi.org/10.1109/ACCESS.2025.3577968>
- Hidayat, J. J., Anshor, A. H., Anwar, M. S., & Syaibani Anwar, M. (2026). Pemodelan Deteksi dan Klasifikasi Fraktur Tulang pada Radiografi X-Ray Menggunakan YOLOv8 dan Preprocessing CLAHE. *JURNAL FASILKOM*, 16(1), 31–45. <https://doi.org/10.37859/JF.V16I1.11241>
- Joshi, S. (2022). *Why Bounding Box Annotation is Crucial: The Ultimate Guide*. https://www.hitechbpo.com/blog/bounding-box-annotation-guide.php?utm_source=chatgpt.com
- Ju, R.-Y., & Cai, W. (2023). Fracture Detection in Pediatric Wrist Trauma X-ray Images Using YOLOv8 Algorithm. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-47460-7>
- Ju, R. Y., & Cai, W. (2023). Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. *Scientific Reports*, 13(1), 20077. <https://doi.org/10.1038/S41598-023-47460-7>
- Ju, R. Y., Chien, C. T., & Chiang, J. S. (2024). YOLOv8-ResCBAM: YOLOv8 Based on An Effective Attention Module for Pediatric Wrist Fracture Detection. *Communications in Computer and Information Science*, 2290 CCIS, 403–416. https://doi.org/10.1007/978-981-96-6972-1_28
- Ju, R. Y., Chien, C. T., Lin, C. M., & Chiang, J. S. (2024). Global Context Modeling in YOLOv8 for Pediatric Wrist Fracture Detection. *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS, 2024*.

* Corresponding author



[Creative Commons Attribution-NonCommercial-ShareAlike 4.0
International License.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

<https://doi.org/10.1109/ISPACS62486.2024.10869064>

- Koonsanit, K., Thongvigitmanee, S., Pongnapang, N., & Thajchayapong, P. (2017). Image enhancement on digital X-ray images using N-CLAHE. *BMEiCON 2017 - 10th Biomedical Engineering International Conference, 2017-January*, 1–4. <https://doi.org/10.1109/BMEiCON.2017.8229130>
- Meena, T., & Roy, S. (2022). Bone Fracture Detection Using Deep Supervised Learning from Radiological Images: A Paradigm Shift. *Diagnostics 2022, Vol. 12, Page 2420, 12(10)*, 2420. <https://doi.org/10.3390/DIAGNOSTICS12102420>
- Olveres, J., González, G., Torres, F., Moreno-Tagle, J. C., Carbajal-Degante, E., Valencia-Rodríguez, A., Méndez-Sánchez, N., & Escalante- Ramírez, B. (2021). What is new in computer vision and artificial intelligence in medical image analysis applications. *Quantitative Imaging in Medicine and Surgery, 11(8)*, 3830–3853. <https://doi.org/10.21037/QIMS-20-1151>
- Oryan, A., Monazzah, S., & Bigham-Sadegh, A. (2015). Bone Injury and Fracture Healing Biology. *Biomedical and Environmental Sciences, 28(1)*, 57–71. <https://doi.org/10.3967/BES2015.006>
- Priyanka, M., Reddy, C. L., Aymon, U., & Jipeng, T. (2025). FractureNet: A Fusion Approach for Automated Hand Fracture Detection and Classification. *Procedia Computer Science, 258*, 1131–1146. <https://doi.org/10.1016/J.PROCS.2025.04.348>
- Rutledge, M., Yap, M., & Chai, K. (2023). Plain film mandibular fracture detection using machine learning – Model development. *Advances in Oral and Maxillofacial Surgery, 11*, 100436. <https://doi.org/10.1016/J.ADOMS.2023.100436>
- Shaban, W., El-Behaidy, W., & badawy, W. (2024). *YOLOv8-Based Fracture Detection System*. <https://doi.org/10.21203/RS.3.RS-5455810/V1>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data 2019 6:1, 6(1)*, 60-. <https://doi.org/10.1186/S40537-019-0197-0>
- Su, Z., Adam, A., Nasrudin, M. F., Ayob, M., & Punganan, G. (2023). Skeletal Fracture Detection with Deep Learning: A Comprehensive Review. *Diagnostics 2023, Vol. 13, Page 3245, 13(20)*, 3245. <https://doi.org/10.3390/DIAGNOSTICS13203245>
- Wu, P., He, X., Dai, W., Zhou, J., Shang, Y., Fan, Y., & Hu, T. (2025). A Review on Research and Application of AI-Based Image Analysis in the Field of Computer Vision. *IEEE Access, 13*, 76684–76702. <https://doi.org/10.1109/ACCESS.2025.3565300>
- Zhang, W., & Ji, S. (2025). Rehabilitation Driven Optimized YOLOv11 Model for Medical X-Ray Fracture Detection. *Sensors 2025, Vol. 25, Page 5793, 25(18)*, 5793. <https://doi.org/10.3390/S25185793>