Scientific Journal of Informatics
Vol. 12, No. 4, November 2025
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

A Hybrid YOLOv8-ResNet50 Architecture for Enhanced Cardiomegaly Prediction from Chest X-rays

Arif Nur Faudin[1*], Farikhin[2], Wahyul Amien Syafei[3]
[1,2,3]Master of Information Systems, Universitas Diponegoro, Indonesia

Abstract.
Objective: This study aim is to develop a reliable deep learning architecture for predicting Cardiomegaly from Chest X-rays images by integrating the ResNet-50 backbone into the YOLOv8 object predicting framework. The proposed hybrid approach overcomes the challenges of predicting subtle anatomical variations and low-contrast features commonly encountered in chest radiographs.
Methods: This study uses a publicly available Chest X-Ray Images dataset. Preprocessing includes adjusting the input image size to 640×640 pixels, automatic orientation correction, and real-time data augmentation applied to the training set. The data is divided 80:20 between training and testing. A hybrid model consisting of ResNet-50 for image feature extraction and YOLOv8 for image prediction was trained for 150 epochs with optimized hyperparameters (learning rate, momentum, weight decay, loss weight), and the performance of the proposed architecture has been evaluated using images metrics such as mAP, Precision, Recall, F1 Score and Confusion Matrix results.
Results: The experimental results indicate that the proposed architecture achieves improved performance in predicting Cardiomegaly, with a mAP50-95 of 0.7578, precision of 0.9955, recall of 0.9962, F1-score of 0.9959, and an inference latency of 4.5 ms per image. These results demonstrate that the model performs better than the standard YOLOv8 variant in both detection accuracy and computational efficiency.
Innovation: The integration of ResNet-50 into YOLOv8 significantly improves feature extraction capabilities for Chest X-ray images, enabling the recognition of fine anatomical details with high precision. This innovative hybrid approach advances automated cardiomegaly detection, offering future potential for large-scale, real-time implementation in clinical settings and contributing to the development of advanced AI-powered diagnostic tools.

Keywords: Cardiomegaly, Deep learning, YOLOv8, ResNet-50, Chest X-ray, Medical image analysis, Hyperparameter tuning
Received October 2025 / **Revised** November 2025 / **Accepted** December 2025

INTRODUCTION

Cardiomegaly, or abnormal enlargement of the heart, is a clinical manifestation commonly associated with serious cardiovascular disorders such as heart failure, hypertension, and coronary artery disease. Early and accurate identification of cardiomegaly is essential to prevent complications and improve patient outcomes. Globally, cardiovascular disease including cardiomegaly is projected to increase significantly, with an estimated rise of up to 90% between 2025 and 2050. Approximately 20.5 million deaths were attributed to cardiovascular disease in 2025, and this number is expected to reach 35.6 million by 2050 [1]. One of the most widely adopted diagnostic indicators is the Cardiothoracic Ratio (CTR), where a value greater than 0.5 on posteroanterior chest radiographs is considered abnormal and indicative of cardiomegaly [2].

Although chest radiography remains the primary imaging modality for assessing heart size, its interpretation is highly dependent on radiologist experience and is prone to variability. At the same time, the demand for medical imaging continues to rise, while the number of skilled radiologists remains limited, creating a diagnostic gap that can potentially be addressed through artificial intelligence (AI) and deep learning (DL) technologies [3]. Convolutional Neural Networks (CNNs) have become the dominant DL approach in medical imaging due to their strong capability in learning spatial features directly from images, enabling high performance in tasks such as classification and object detection [4]. Their multilayer representation allows robust extraction of complex visual patterns, supporting various computer-aided diagnosis systems [5].

Recent advancements in object detection models, particularly the YOLO (You Only Look Once) family, have demonstrated strong potential for real-time medical imaging applications [6]. YOLOv8 has shown

*Corresponding author.
Email addresses: arifnurfaudin@gmail.com (Faudin)*, farikhin@lecturer.undip.ac.id (Farikhin), wasyafei@elektro.undip.ac.id (Syafei)
DOI: 10.15294/sji.v12i4.35225

superior accuracy and inference speed in several medical domains, including bone fracture detection [7] and lung cancer detection [8], indicating its suitability for rapid diagnostic support. However, despite its strengths, standard YOLOv8 can face challenges when applied to chest radiographs, which typically exhibit low contrast, overlapping anatomical structures, and subtle morphological differences—conditions that often degrade feature extraction performance.

To overcome these limitations, this study proposes a hybrid YOLOv8–ResNet50 architecture specifically optimized for cardiomegaly prediction. ResNet-50, known for its deep residual learning and ability to capture fine-grained patterns in low-contrast imagery [9], is integrated as the backbone to enhance YOLOv8's feature extraction capability. This architectural enhancement, combined with targeted data augmentation and hyperparameter optimization, forms the core novelty of this research.

In summary, the contributions of this work are fourfold : (1) development of a custom hybrid architecture that integrates ResNet-50 into YOLOv8 to strengthen feature representation on Chest X-Ray Images, (2) optimization of prediction performance through systematically tuned hyperparameters and augmentation strategies, (3) comprehensive evaluation against standard YOLOv8 to demonstrate measurable improvements in cardiomegaly prediction, and (4) deployment of the model into an accessible web-based diagnostic support system for practical clinical use. This combination positions the proposed method as an enhanced object prediction based framework specifically tailored to the diagnostic characteristics of Cardiomegaly

METHODS

The methodology in this study was designed to systematically describe the steps taken from data collection to model evaluation. All stages of this research are designed to produce a comprehensive cardiomegaly detection architecture development process. This research uses a deep learning-based experimental approach to predict cardiomegaly from chest radiography images, with the stages illustrated in Figure 1.

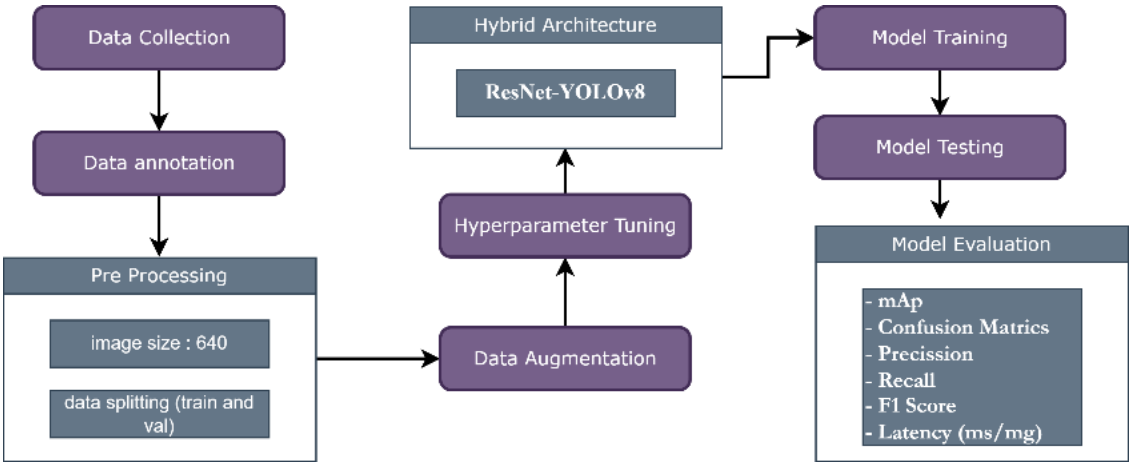


Figure 1. Flowchart of the proposed architecture Hybrid YOLOv8-ResNet50

Data collection and annotation

The data used is a collection of chest radiographs relevant to cases of cardiomegaly, obtained from a public repository, namely the Keggle platform. The training data is the Shenzhen dataset, which consists of 563 images. This dataset was collected in collaboration with Shenzhen No. 3 People's Hospital, Guangdong Medical College, Shenzhen, China. The testing data used the Montgomery dataset, which was collected in collaboration with the Department of Health and Human Services, Montgomery County, Maryland, USA. This dataset contains 131 frontal Chest X-ray images [10].

All images will be annotated on unlabelled images. The image annotation process involves providing a bounding box that carefully marks the outer boundaries of each organ on each chest radiography image. The annotation process can be seen in Figure 2.

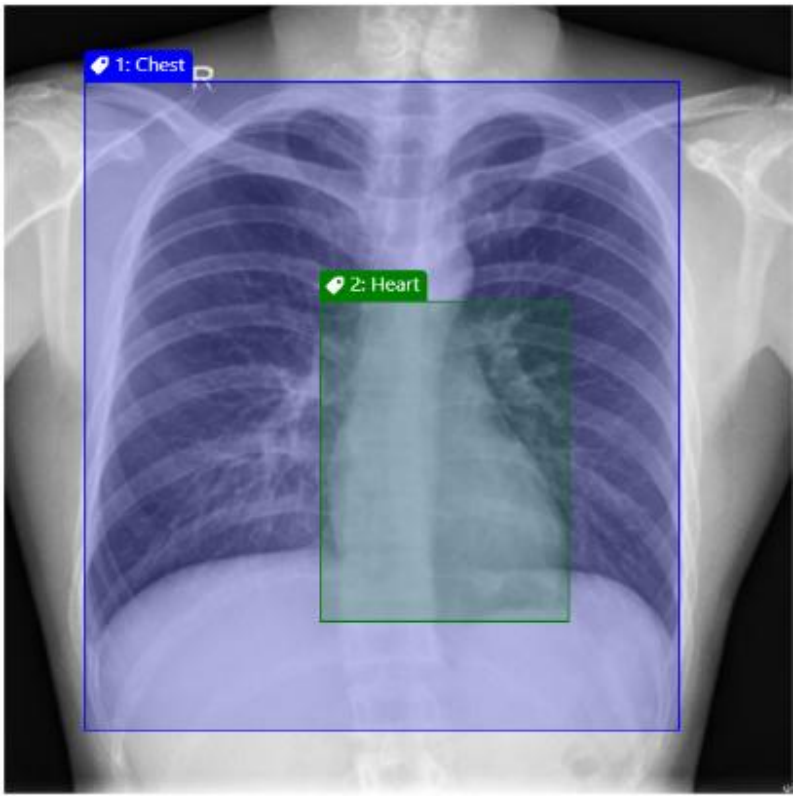


Figure 2. Annotation of chest radiograph images, 1. Chest and 2. Heart

The bounding box of the annotation results allows for the measurement of the maximum width of the heart and the maximum internal width of the chest cavity. The numerical data from these measurements is then used as the basis for automatically calculating the Cardiothoracic Ratio (CTR) value, thereby distinguishing between patients diagnosed with cardiomegaly and those who are normal.
This process is also supported by the use of annotation software, such as Label Studio, which has been proven effective in tagging large-scale radiology data for medical artificial intelligence research and model development [11]. CTR is calculated using equation (1):

$$\mathrm{CTR} = \frac{\mathrm{a} + \mathrm{b}}{\mathrm{c}} \tag{1}$$

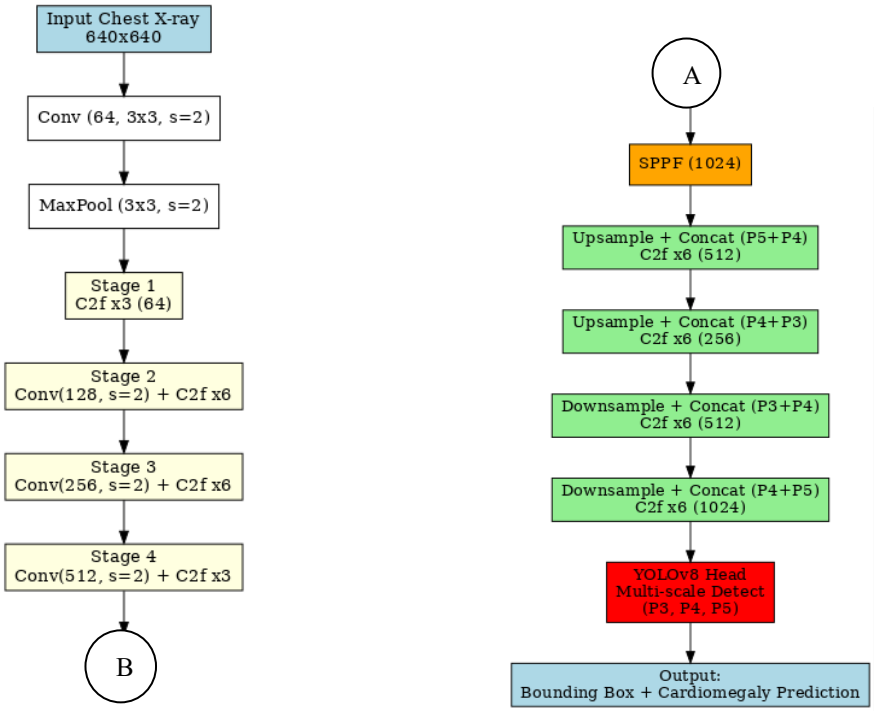
where *a* is the distance from the right border of the heart to the midline, *b* is the distance from the left border of the heart to the midline, and *c* is the maximum thoracic diameter (TD) above the costophrenic angle measured from the inner edge of the rib [2].

Data preparation
The collected image data was obtained from a public repository, namely the Kaggle platform, which consists of 694 images. The collected image data was first processed by standardizing the pixel size to 640×640 to ensure consistency of input to the object detection model [12]. Next, each image underwent an automatic orientation process to ensure that all images were displayed uniformly and ready for the training stage. After that, the dataset was divided proportionally into training data and testing data with a ratio of 80:20 [13].

Data augmentation
To improve model generalization and prevent overfitting due to limited data, data augmentation techniques are applied to the training set [14]. Data augmentation is a key technique in deep learning that significantly improves model performance and generalization [15]. This augmentation artificially creates new data variations through a series of geometric and photometric transformations, such as horizontal flipping, rotation, scaling, and color adjustment [16].

Proposed hybrid architecture (YOLOv8-ResNet50)

The basic architecture used is YOLOv8, a single-shot object detection model known for its speed and accuracy [17]. The main contribution of this research is the modification of the standard YOLOv8 architecture by integrating a backbone inspired by ResNet-50. ResNet was chosen for its proven ability to extract deep hierarchical features through the use of residual connections, which effectively overcome the vanishing gradient problem in very deep networks [18]. This integration aims to improve the model's ability to recognize specific features in medical images. The YOLOv8- ResNet50 model architecture can be seen in Figure 3.



(a) (b)

Figure 3. Visualization of the YOLOv8-ResNet50 model architecture. (a) Backbone structure (Feature extractor). (b) Neck structure (PANet feature fusion) and detection head

Figure 3 illustrates the YOLOv8-ResNet50 architecture for predicting Cardiomegaly through Chest X-ray images. The input image, measuring 640x640 pixels, is processed through several stages of convolution with increasingly larger filters (64, 128, 256, 512) and the Czf x n technique, which strengthens features with residual connections. After that, SPPF (Spatial Pyramid Pooling Fusion) is used to combine multi-scale spatial information and generate a 1024-sized feature vector. Upsampling and downsampling techniques combine features from various levels (P3, P4, P5) to improve prediction accuracy. At the end of the architecture, YOLO Head is used to detect bounding boxes and classify Cardiomegaly, enabling real-time and accurate object detection in medical images [19].

Hyperparameter tuning

Adjustments to key hyperparameters such as learning rate, batch size, optimizer, and augmentation intensity were performed iteratively based on evaluation of the validation data, resulting in the most optimal model configuration for the Cardiomegaly prediction task [20].

Model training & testing

The model training process begins with the initialization of the YOLOv8-ResNet50 architecture, which is trained for 150 epochs. This training is configured using a series of hyperparameters that have been optimized through the previous tuning stage, including learning rate, momentum, and loss function weights [21]. During this iterative process, real-time data augmentation such as flipping, scaling, and color variation was applied to each batch of images to improve model generalization. Model performance was monitored

after each epoch using the validation data set, and the weights from the epoch with the best mAP score were saved as the final model. After the entire training process is complete, a testing phase is conducted as a final objective evaluation, in which the best model is tested on a separate test set that has never been used before to measure final performance metrics such as mAP, Precision, and Recall [22].

Model evaluation

The evaluation of model prediction results was carried out using a confusion matrix, which consists of several key metrics such as precision, recall, and F1-score [23]. Precision indicates the proportion of correct positive predictions to the total positive predictions generated by the model. Meanwhile, recall measures how many correct positive predictions there are compared to the total number of positive data available. The F1-score is the harmonic mean between precision and recall, which are analysed together. Accuracy and confusion matrix can be formulated through a number of equations, namely equations (2), (3), and (4).

$$Precision = \frac{\text{TP}}{TP+FP} \tag{2}$$

$$Recall = \frac{\text{TP}}{TP+FN} \tag{3}$$

$$F1 = \frac{\text{2 x Precision X Recall}}{Precision + Recall} \tag{4}$$

Where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative serve as key parameters that indicate whether the model can accurately identify and classify objects in each detection result.

Meanwhile, mean average precision (mAP) is used as the main measure of accuracy in object detection tasks, where mAP is obtained by calculating the average precision value across all tested classes [24]. This value is calculated by integrating the precision-recall curve for each category, then taking the average of all average precision (AP) values obtained [25], where N is the total number of classes evaluated. The formula for mAP is described in Equation (5).

$$mAP = \frac{1}{N} \sum_{k=1}^{n} APc \tag{5}$$

RESULTS AND DISCUSSION

Data preprocessing

This study used a dataset from the Keggle website containing frontal Chest X-ray images of patients, with a total of 694 image datasets. The resolution and pixel size in the dataset was 640x640 pixels for each image. Data annotation was performed on unlabeled images by placing bounding boxes on the images [26]. The labels given to the images consisted of two classes, namely the heart and chest cavity, which had been defined at the data collection stage. At the image augmentation stage, augmentation techniques (e.g., horizontal flip, blur, mosaic, saturation, and brightness adjustment) were applied sequentially:

1. Flip: Horizontal (65% probability).
2. Hue: Small color shift, ±2.8%.
3. Saturation: Range of approximately –67% to +67%.
4. Brightness (Value/Intensity): Range of approximately –44% to +44%.
5. Translate (Shift): Maximum ±5.7% from the original position.
6. Scale (Zoom in/out): Change in object size up to ±61%.
7. Mosaic: Always applied (100%) – combines 4 images in 1 frame.
8. Close Mosaic: Mosaic is disabled in the last 10 epochs to stabilize the model.

The data augmentation process in YOLOv8 does not increase the number of image files, so the dataset still consists of 563 training images and 131 validation images. However, each batch of images that enters the model undergoes on-the-fly augmentation such as flip, scale, translate, hue, saturation, brightness, and mosaic. Thus, even though the number of files does not change, the variety of images seen by the model increases significantly. In one epoch, all 563 training images are processed with random augmentation, so that in 150 epochs the model has the potential to see up to 84,450 unique image variations. This mechanism helps improve model generalization and reduces the risk of overfitting. Conversely, augmentation is not

applied to the validation data (only resizing and normalization), so that the model performance evaluation remains objective and reflects the prediction capabilities on the original images.

Hyperparameter tuning

In this study, the model training process was carried out by optimizing a number of hyperparameters that play an important role in determining the quality of learning and model convergence. Hyperparameter adjustments include optimization, regularization, and weights on the loss function used by YOLOv8 [27]. The tuning results in Table 1 show the optimal configuration used in the final training, which significantly contributed to improving the model's performance in detecting cardiomegaly in Chest X-ray images.

Table 1. Hyperparameters optimized in YOLOv8-ResNet50 training

Hyperparameter	Value	Function
Learning Rate Awal (lr0)	0.00795	Set the size of the weight update step at the beginning of training.
Learning Rate Final (lrf)	0.00727	The learning rate value after the decay process maintains stability at the end of training.
Momentum	0.84997	Maintain the stability of the gradient direction so that it does not easily get stuck at *local minima*.
Weight Decay	0.00078	Regularization to prevent overfitting by reducing excessive weights.
Warmup Epochs	3.08136	The initial number of epochs for gradually increasing the learning rate.
Warmup Momentum	0.69663	Momentum values during the warmup phase help make the initial transition to training smoother.
Box Loss Gain	8.6415	The weight of the *loss* function for bounding box prediction accuracy.
Class Loss Gain (cls)	0.59359	The weight of the *loss* function for object class classification.
Distribution Focal Loss (dfl)	1.61848	The weight of the IoU-based bounding box distribution loss function for higher precision.

CTR computation pipeline after detection output

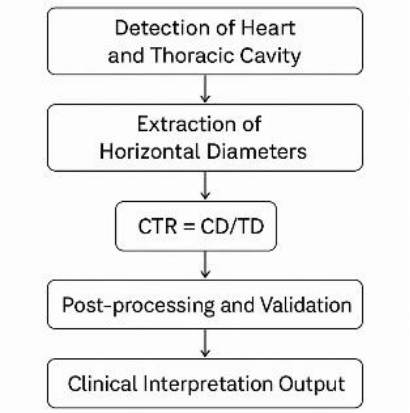


Figure 4. CTR Computation pipeline after detection output from proposed YOLOv8-ResNet50 architecture

Figure 4. illustrates the computational pipeline used to automatically derive the Cardiothoracic Ratio (CTR) from the bounding box predictions produced by the proposed YOLOv8-ResNet50 architecture. After inference, the model generates two key bounding boxes corresponding to the cardiac silhouette (heart) and thoracic cavity (chest). Each bounding box consists of coordinates $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$ that are extracted directly from the detection output.

In the first processing stage, the cardiac diameter is computed by measuring the horizontal distance between the leftmost and rightmost points of the predicted heart bounding box. Similarly, the thoracic diameter is obtained from the width of the chest bounding box. Because CTR is a dimensionless ratio, both measurements are derived in pixel units without the need for calibration to physical dimensions. The second stage calculates the CTR value using the standard clinical formula on equations (6):

$$CTR = \frac{Cardiac\ Diameter}{Thoracic\ Diameter} \tag{6}$$

The computed CTR is then compared with the established clinical threshold of 0.50. A ratio greater than 0.50 indicates cardiomegaly, while values at or below the threshold represent normal heart size. This classification output is automatically generated alongside the numerical CTR value, enabling real-time decision support [28].

Finally, the CTR result, bounding-box visualization, and classification are transmitted to the user interface or API endpoint. This automated CTR pipeline ensures consistency, eliminates inter-observer variability, and enhances the model's role not only as an object detector but also as a quantitative diagnostic support tool. This integrated capability strengthens the clinical relevance of the proposed ResNet-YOLOv8 architecture.

Model evaluation

The results of the experiments in this study show the effectiveness of the ResNet-YOLOv8 architecture in detecting cardiomegaly in chest X-ray images. The evaluation was carried out through statistical analysis of the dataset annotations, training curve analysis, confusion matrix, and comparison with several standard variants of YOLOv8.

Figure 5 presents four plots, each of which provides unique insights into the dataset. The first plot on the top left shows the distribution of the number of instances for each class. It can be seen that this dataset is very balanced, with exactly the same number of instances for the 'Chest' and 'Heart' classes, namely 563 instances. This class balance is ideal because it can prevent the model from becoming biased towards the majority class during the training process.

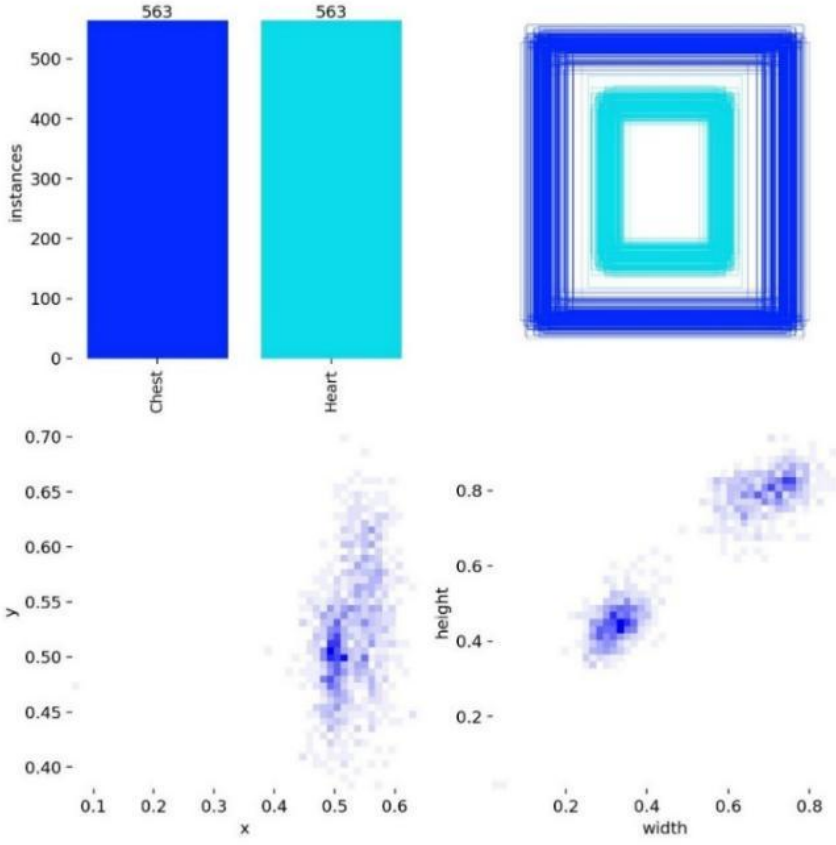


Figure 5. Statistical visualization of Heart and Chest cavity data annotations

Visual analysis of the annotation data in Figure 5 confirms the quality and suitability of the dataset for training. The results show that the dataset is perfectly balanced with 563 instances for each class ('Chest' and 'Heart'), which prevents potential bias in the model. The spatial distribution of objects is also consistent with radiographic images, where objects are centered and the 'Heart' class is located within the 'Chest'.

Furthermore, bounding box dimension analysis shows two distinct clusters of different sizes between the two classes, validating this dataset as a strong basis for training a cardiomegaly detection model.

Next, Figure 6 shows the training and validation curves, where the box loss, classification loss, and distribution focal loss values show a progressive downward trend as the number of epochs increases. This decrease in loss values indicates that the model is able to learn feature representations well, reduce prediction errors, and produce stable generalization. In addition, evaluation metrics such as precision, recall, mAP50, and mAP50-95 also show consistent improvements, confirming the stability of the model's convergence.

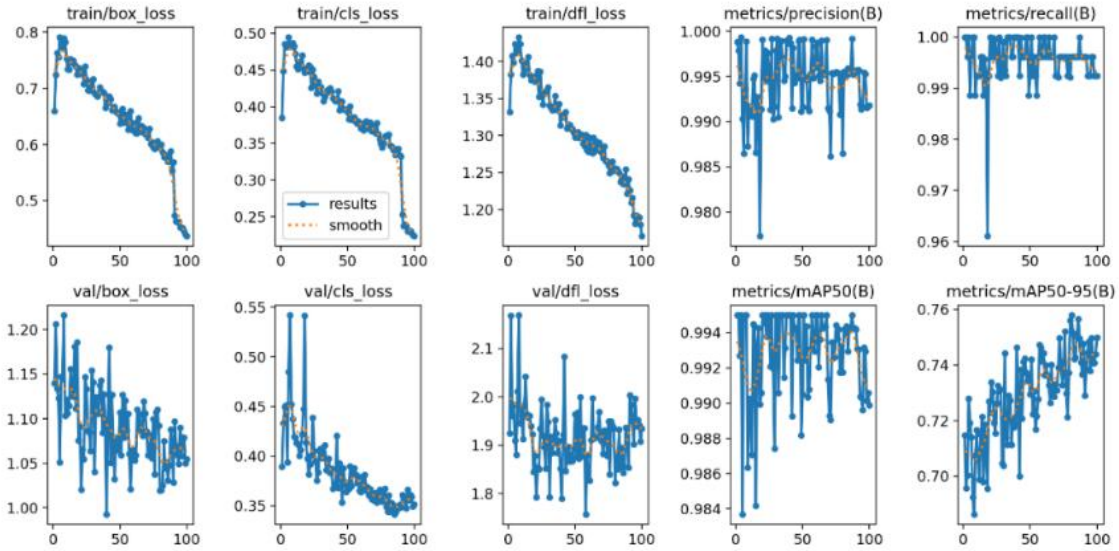


Figure 6. Training results curve for the ResNet-YOLOv8 model

Further evaluation is shown in Figure 7, which is the confusion matrix of the validation results. The model was able to correctly classify all samples into two classes (chest and heart), each with 131 data points. These results indicate excellent predictive capabilities with minimal classification errors (false positives and false negatives), proving that the proposed model is reliable in detecting cardiomegaly.

A comparison of the performance of the proposed architecture with the standard YOLOv8 variants is shown in Table 2. The results show that ResNet-YOLOv8 (best) outperforms YOLOv8n, YOLOv8s, YOLOv8m, and YOLOv8l, particularly in terms of precision (0.9955), recall (0.9962), F1-score (0.9959), and mAP50-95 (0.7578). Although YOLOv8l has a comparable mAP50-95 score (0.7582), the model is slower with a latency of 7.27 ms/img, while the proposed model is much more efficient with only 4.5 ms/img.

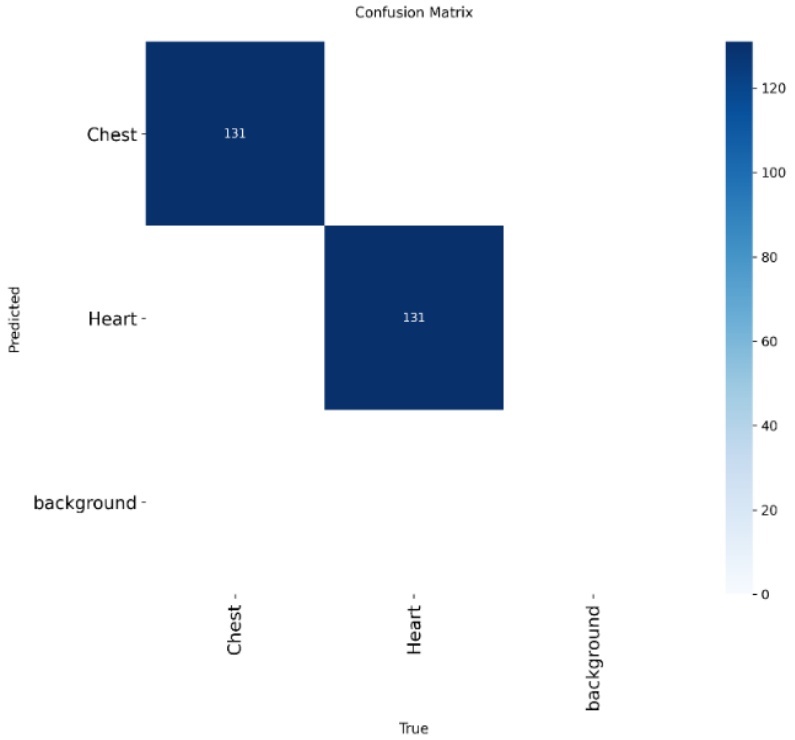


Figure 7. Confusion matrix of the evaluation results of the ResNet-YOLOv8 model on the validation data

Table 2. Comparison of ResNet-YOLOv8 performance with standard YOLOv8 variants

Model	mAP50-95	mAP50	Precision	Recall	F1-Score	Latensi (ms/img)
YOLOv8n	0.7561	0.9934	0.9939	0.996	0.9949	6.31
YOLOv8s	0.7529	0.9898	0.9922	0.9924	0.9923	5.22
YOLOv8m	0.7537	0.9884	0.9902	0.9885	0.9894	6.32
YOLOv8l	0.7582	0.9931	0.9915	0.9924	0.9919	7.27
Proposed model (ResNet-YOLOv8)	**0.7578**	**0.9942**	**0.9955**	**0.9962**	**0.9959**	**4.5**

This improvement in accuracy and inference efficiency confirms the great potential of the proposed architecture to support the automatic early diagnosis of cardiomegaly, assist radiologists in clinical decision-making, and reduce manual workload [29].

Qualitative detection results

To complement the quantitative evaluation, Figure 8 presents qualitative examples of cardiomegaly detection using the proposed ResNet50-YOLOv8 model. The visualizations show the predicted bounding boxes for the cardiac silhouette and thoracic cavity along with their confidence scores. The model consistently localizes both anatomical structures accurately across different chest X-ray samples, including images with low contrast and varying patient anatomy.

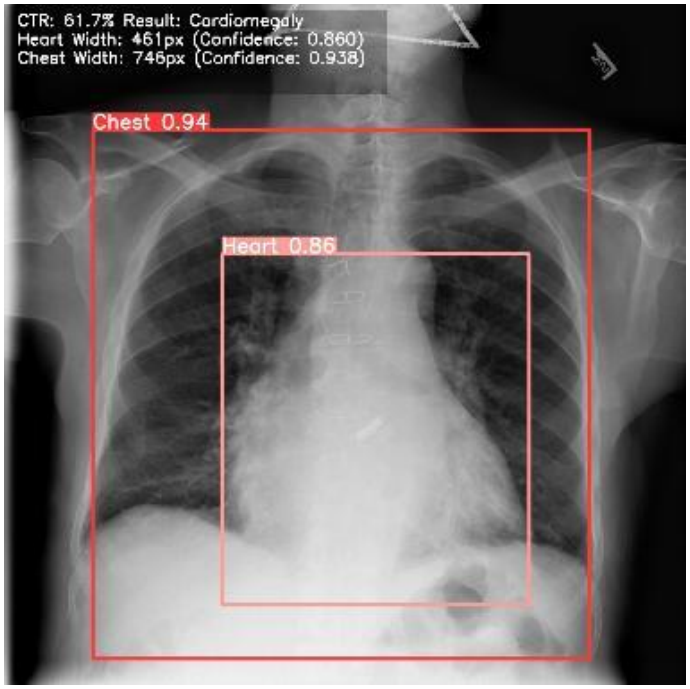


Figure 8. Example of prediction results on Chest X-ray Images using ResNet50-YOLOv8

The detection results demonstrate that the model is able to distinguish the heart and chest boundaries clearly, which is essential for reliable Cardiothoracic Ratio (CTR) computation. These qualitative findings reinforce the quantitative performance reported earlier and confirm the model's capability to generalize to unseen radiographs.

Although the proposed ResNet-YOLOv8 architecture demonstrates improved performance in detecting cardiomegaly, several important limitations must be considered. First, the dataset employed in this study was sourced exclusively from two public repositories (Shenzhen and Montgomery), restricting the diversity of image acquisition protocols, patient demographics, and clinical presentations [30]. This limitation may affect the model's generalizability to broader and heterogeneous clinical populations. Second, the relatively modest dataset size constrains the models ability to robustly learn rare or atypical anatomical patterns, while data augmentation was implemented, synthetic variations cannot fully replicate the complexity and diversity observed in real-world clinical practice. Third, the evaluation in this study relied largely on object detection metrics such as precision, recall, and mean average precision (mAP), which may not fully capture clinically relevant decision factors or reflect radiologist level agreement. Lastly, rigorous clinical validation including multi-center trials or integration with hospital PACS/EMR systems was beyond the present scope and represents a necessary future direction for upcoming deployment.

CONCLUSION

This study proposes a ResNet-YOLOv8 architecture for detecting cardiomegaly in chest X-ray images. By integrating the ResNet-50 backbone into YOLOv8 and performing hyperparameter tuning and data augmentation strategies, the resulting model shows significant improvement over the standard YOLOv8 variant. The experimental results show that the proposed model achieves a precision of 0.9955, recall of 0.9962, F1-score of 0.9959, and mAP50 of 0.9936, with an inference latency of only 4.5 ms/img, making it superior in terms of both accuracy and efficiency.

These findings confirm that ResNet-YOLOv8 is capable of improving visual feature extraction capabilities and producing more precise cardiomegaly detection in medical images. With its high performance and good computational efficiency, this model has great potential for use as a clinical decision support system in the early diagnosis of cardiomegaly. However, this study is still limited to a small-scale, single-source dataset. Therefore, future research can be directed towards validation with larger, multi-center datasets and the development of integration with other architectures, such as transformer-based models, to further improve accuracy and generalization.

REFERENCES

[1] B. Chong *et al.*, "Global burden of cardiovascular diseases: projections from 2025 to 2050," *Eur J Prev Cardiol*, vol. 32, no. 11, pp. 1001–1015, Aug. 2025, doi: 10.1093/eurjpc/zwae281.

[2] W. Kusakunniran *et al.*, "Automatic measurement of cardiothoracic ratio in chest x-ray images with ProGAN-generated dataset," *Applied Computing and Informatics*, Apr. 2023, doi: 10.1108/ACI-11-2022-0322.

[3] B. P. Doppala, A. Al Bataineh, and B. Vamsi, "An Efficient, Lightweight, Tiny 2D-CNN Ensemble Model to Detect Cardiomegaly in Heart CT Images," *J Pers Med*, vol. 13, no. 9, p. 1338, Aug. 2023, doi: 10.3390/jpm13091338.

[4] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[5] M. Swapna, Dr. Y. K. Sharma, and Dr. B. Prasad, "CNN Architectures: Alex Net, Le Net, VGG, Google Net, Res Net," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 6, pp. 953–959, Mar. 2020, doi: 10.35940/ijrte.F9532.038620.

[6] Z. Cai, K. Zhou, and Z. Liao, "A Systematic Review of YOLO-Based Object Detection in Medical Imaging: Advances, Challenges, and Future Directions," *Computers, Materials and Continua*, vol. 85, no. 2, pp. 2255–2303, 2025, doi: 10.32604/cmc.2025.067994.

[7] V. Shree, "Detection and Classification of Bone Fractures in X-Ray Images using Yolov8," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 10, no. 4, pp. 2454–132, 2024, [Online]. Available: https://www.ijariit.com

[8] M. Elavarasu and K. Govindaraju, "Unveiling the Advancements: YOLOv7 vs YOLOv8 in Pulmonary Carcinoma Detection," *Journal of Robotics and Control (JRC)*, vol. 5, no. 2, pp. 459–470, Feb. 2024, doi: 10.18196/jrc.v5i2.20900.

[9] T. Alam *et al.*, "'An Integrated Approach using YOLOv8 and ResNet, SeResNet & Vision Transformer (ViT) Algorithms based on ROI Fracture Prediction in X-ray Images of the Elbow,'" *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 20, Nov. 2024, doi: 10.2174/0115734056309890240912054616.

[10] N. Gaggion *et al.*, "CheXmask: a large-scale dataset of anatomical segmentation masks for multi-center chest x-ray images," *Sci Data*, vol. 11, no. 1, p. 511, May 2024, doi: 10.1038/s41597-024-03358-1.

[11] W. Fan *et al.*, "A deep-learning-based framework for identifying and localizing multiple abnormalities and assessing cardiomegaly in chest X-ray," *Nat Commun*, vol. 15, no. 1, p. 1347, Feb. 2024, doi: 10.1038/s41467-024-45599-z.

[12] Z. Fatima, M. H. Tanveer, H. Mariam, R. C. Voicu, T. Rehman, and R. Riaz, "Performance Comparison of Object Detection Models for Road Sign Detection Under Different Conditions," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 12, 2024, doi: 10.14569/IJACSA.2024.0151299.

[13] A. Rácz, D. Bajusz, and K. Héberger, "Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification," *Molecules*, vol. 26, no. 4, p. 1111, Feb. 2021, doi: 10.3390/molecules26041111.

[14] N. I. Fardana, R. R. Isnanto, and O. D. Nurhayati, "Pneumothorax Detection System in Thoracic Radiography Images Using CNN Method," *Scientific Journal of Informatics*, vol. 11, no. 4, pp. 981–990, Jan. 2025, doi: 10.15294/sji.v11i4.16635.

[15] T. Islam, Md. S. Hafiz, J. R. Jim, Md. M. Kabir, and M. F. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," *Healthcare Analytics*, vol. 5, p. 100340, Jun. 2024, doi: 10.1016/j.health.2024.100340.

[16] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[17] A. Widayani, A. M. Putra, A. R. Maghriebi, D. Z. C. Adi, and Moh. H. F. Ridho, "Review of Application YOLOv8 in Medical Imaging," *Indonesian Applied Physics Letters*, vol. 5, no. 1, pp. 23–33, May 2024, doi: 10.20473/iapl.v5i1.57001.

[18] V. Lakide and V. Ganesan, "Precise Lung Cancer Prediction using ResNet – 50 Deep Neural Network Architecture," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 1, pp. 38–46, Nov. 2024, doi: 10.35882/jeeemi.v7i1.518.

[19] R. Patel and A. Chaware, "Comparative Analysis of Hyperparameter Tuned Convolutional Neural Networks for classification of Diabetic Retinopathy," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 9s, pp. 185–197, 2023.

[20] S. S. A. P. Ece, S. J. Prabu, D. S. Stephen, S. Sutharsan, and P. Vishnu, "Diagnosis of Cardiovascular Diseases Using YOLOv8 on MRI Heart Imaging Data Abstract :," vol. 8, no. 2, pp. 2154–2159, 2025.
[21] D. Altınel, "Development of Deep Learning Optimizers: Approaches, Concepts, and Update Rules," Sep. 2025, [Online]. Available: http://arxiv.org/abs/2509.18396
[22] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci Rep*, vol. 14, no. 1, p. 6086, Mar. 2024, doi: 10.1038/s41598-024-56706-x.
[23] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.
[24] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach Learn Knowl Extr*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, doi: 10.3390/make5040083.
[25] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A Survey on Performance Metrics for Object-Detection Algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, IEEE, Jul. 2020, pp. 237–242. doi: 10.1109/IWSSIP48289.2020.9145130.
[26] R. Rianto, V. Purwayoga, Aradea, A. A. Mikail, and I. Yumna, "SOCA-YOLO: Smart Optic with Coordinate Attention Model for Vision System-Based Eye Disease Detection," *Scientific Journal of Informatics*, vol. 12, no. 3, pp. 465–476, Sep. 2025, doi: 10.15294/sji.v12i3.29293.
[27] M. S. Mas, S. Saidah, and N. Ibrahim, "Detection and counting of wheat ear using YOLOv8," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 5, p. 5813, Oct. 2024, doi: 10.11591/ijece.v14i5.pp5813-5823.
[28] L. Zhou *et al.*, "Detection and Semiquantitative Analysis of Cardiomegaly, Pneumothorax, and Pleural Effusion on Chest Radiographs," *Radiol Artif Intell*, vol. 3, no. 4, p. e200172, Jul. 2021, doi: 10.1148/ryai.2021200172.
[29] A. M. Ayalew, B. Enyew, Y. A. Bezabh, B. M. Abuhayi, and G. S. Negashe, "Early-stage cardiomegaly detection and classification from X-ray images using convolutional neural networks and transfer learning," *Intelligent Systems with Applications*, vol. 24, p. 200453, Dec. 2024, doi: 10.1016/j.iswa.2024.200453.
[30] M. Raghu and E. Schmidt, "A Survey of Deep Learning for Scientific Discovery," Mar. 2020, [Online]. Available: http://arxiv.org/abs/2003.11755