ASPIRE-YOLOV12: Adaptive SimAM Attention and Contrast Enhanced for Multi Disease Pulmonary Detection

Pandu Pangestu
Departement of Informatics
Faculty of Engineering
Siliwangi University
Tasikmalaya, Indonesia
pangestup60@gmail.com

Irfan Darmawan
Department of Information Systems
Faculty of Industrial Engineering
Telkom University
Bandung, Indonesia
irfandarmawan@telkomuniversity.ac.id
**corresponding author*

Alam Rahmatulloh
Departement of Informatics
Faculty of Engineering
Siliwangi University
Tasikmalaya, Indonesia
alam@unsil.ac.id

***Abstract*—Lung disease remains one of the most prevalent diseases in the world, including in Indonesia. The complexity of radiographic patterns on X-ray images often complicates the automatic diagnosis process due to overlapping tissue structures and high intensity variations. This study employs deep learning technology by proposing ASPIRE- YOLOv12, an adaptive lung disease detection architecture based on YOLOv12 equipped with two main components, namely Contrast Limited Adaptive Histogram Equalization (CLAHE) and Simple Attention Module (SimAM). CLAHE functions as a pre-processing stage to normalize image contrast so that lung structures and lesion boundaries become clearer, while SimAM is integrated into the YOLOv12 backbone as a parameter-free attention mechanism that strengthens the model's focus on important areas without adding to the complexity of the network. The combination of these two modules improves the sharpness and clarity of features extracted from medical images. Experimental results show that ASPIRE-YOLOv12 provides a significant performance improvement over the baseline model. The YOLOv12n model produced a mAP50 of 65%, and YOLOv12s model produced a mAP50 of 68%, while ASPIRE-YOLOv12 achieved a mAP50 of 77%. This improvement shows that the synergy between contrast normalization and adaptive attention can improve feature representation quality and detection accuracy in highly complex medical images.**

Keywords—Attention Module, CLAHE, Deep Learning, Image Enchancement, X-ray Image

I. INTRODUCTION

Respiratory illnesses represent a major source of global morbidity and mortality. The World Health Organization (WHO) identifies conditions such as pneumonia, tuberculosis, and chronic obstructive pulmonary disease (COPD) as continuing significant health challenges, especially in developing countries, including Indonesia [1]. Early diagnosis is an important aspect in increasing the chances of recovery and reducing mortality rates. However, the manual diagnosis process, which relies on the interpretation of chest X-rays by radiologists, often faces limitations such as fatigue, subjectivity in assessment, and differences in results between examiners [2].

X-ray images themselves have unique characteristics, namely being monochromatic, having low contrast, and containing complex anatomical structures such as bones, soft tissues, and overlapping organs. This makes it difficult to visually detect small lesions, such as pulmonary nodules or subtle infiltrates [3]. Therefore, the use of deep learning technology has become a promising alternative in assisting the automatic diagnosis process for detecting lung abnormalities [4].

You Only Look Once (YOLO) is recognized as a premier deep learning framework, distinguished by its capacity for real-time object detection without compromising accuracy [5]. Its latest version, YOLOv12 [6], introduces an architecture that is more focused on attention-centric feature learning, which allows the network to better understand spatial and semantic context compared to previous versions[7]. However, implementing YOLO for medical image diagnostics is not without limitations. The model's performance is often constrained by the inconsistent quality of X-ray datasets, stemming from differences in acquisition equipment and patient status. Such variations, specifically noise and poor contrast can degrade the network's ability to perform optimal feature extraction [8].

Various studies have been conducted in the field of lung disease detection using deep learning- based X-ray images. Object detection methods such as YOLO have been widely applied due to their advantages in identifying the location of abnormalities at high speed [9]. For example, research conducted by [10] developed a Fast-YOLO model to detect pneumonia in X-ray images. The results of this study show that the use of YOLO can increase inference speed without reducing detection accuracy. Meanwhile, studies [11] and [12] applied YOLOv8 to detect pulmonary nodules and reported an increase in mAP after pre-processing the images.

As YOLO versions evolved, several studies also adapted the model architecture to better suit the medical domain. Furthermore, [13] proposed the YOLO-CXR model, which is specifically designed for multi-disease detection in the lungs. The model introduces refined convolutional layers to address the problem of small lesion sizes and low contrast in X-ray images. Research [14] then introduced NSEC-YOLO, which combines noise suppression and global feature aggregation to improve the model's robustness against low-quality images. These two studies show a trend that adapting the YOLO structure is increasingly necessary in the context of medical images. In addition to architectural development, improving image quality is an important step in improving detection model performance. CLAHE is one of the contrast enhancement methods widely used in medical images. Research [15], which applied CLAHE to X-ray images to detect diseases in medical images, reported an increase in classification accuracy. This technique has been proven to clarify the differences between healthy tissue and pathological lesions such as opacity and fine infiltration. This emphasizes that histogram-based pre-processing stages greatly affect detection results, especially in datasets with high lighting variation.

On the other hand, attention mechanisms have been developed to enable deep-learning models to capture dependencies among input elements and represent contextual relationships [16]. SimAM is a simple attention module that does not require additional parameters [17]. SimAM works based on an energy function that assesses the importance of each neuron in the feature map, allowing the model to focus its learning on relevant areas [18]. This approach has been shown to improve model performance on various computer vision tasks, including segmentation and object detection.

Based on previous research, it can be concluded that both image quality improvement through CLAHE and feature focus enhancement through attention modules such as SimAM contribute significantly to improved detection performance. However, most previous studies have focused on only one aspect, either pre-processing or network attention, without combining the two in a single framework. Therefore, this study presents a more comprehensive approach by integrating CLAHE and SimAM into the YOLOv12 model. The ASPIRE-YOLOv12 model is expected to overcome these limitations by offering a more adaptive, accurate, and efficient multi-disease detection system.

To address these issues, this study proposes the application of two complementary approaches, namely Contrast Limited Adaptive Histogram Equalization (CLAHE) and Simple Attention Module (SimAM). CLAHE is an image contrast enhancement method that works by normalizing the local intensity distribution so that the details of the lung anatomy are more prominent without increasing noise [19]. This technique has been proven effective in improving the visual quality of medical images, especially chest radiographs. Meanwhile, SimAM is a parameter-free attention module developed to help neural networks focus on important areas adaptively based on neuron energy [18]. This module can strengthen spatial and semantic features in the YOLOv12 backbone without significantly increasing the computational load. By combining these two approaches, this study introduces the ASPIRE-YOLOv12 (Adaptive SimAM Attention and Contrast-Enhanced YOLOv12) model, which is expected to provide both theoretical and practical added value in the development of X-ray image-based lung disease detection systems, namely:

1. YOLOv12 was adopted as the baseline one-stage detector for multi-lung disease detection tasks because of its superior accuracy and computational efficiency compared to previous generations.
2. Integrating the CLAHE adaptive contrast enhancement method to normalize the local intensity of X-ray images so that the structure of the lung tissue and lesions become clearer without introducing excessive noise.
3. Applying the SimAM module to the YOLOv12 backbone enables a parameter-free attention mechanism to adaptively strengthen the network's focus on pathological areas without adding parameter load.
4. Improving the model's adaptability to variations in medical image quality, particularly in X-ray conditions with low contrast or uneven intensity distribution, which were previously major limitations of deep learning-based automatic detection.

II. METHOD

The research methodology was systematically designed to develop and evaluate our proposed lung disease detection architecture, ASPIRE-YOLOv12. This section outlines the chronological steps, from dataset acquisition to model evaluation, ensuring reproducibility and scientific rigor. The overall workflow, illustrated in Figure 1, encompasses data preparation, preprocessing, architectural modifications, training, and performance assessment.

TABLE I. PSEUDO-CODE FOR CONTRAST LIMITED ADAPTIVE HISTOGRAM EQUALIZATION

Algorithm: CLAHE-Based Image Enhancement

Input: Grayscale image I, tile size T, clip limit C, and gray levels L

Output: Enhanced image $I_{(CLAHE)}$

 # Tile Division
Split I into non-overlapping tiles T_i with size $T \times T$.

 # Histogram Processing
For each tile T_i:
$H_i \leftarrow$ Histogram(T_i)
$H_i \leftarrow$ ClipAndRedistribute(H_i, C)
$CDF_i \leftarrow$ CumulativeHistogram(H_i)
$T_i^{(eq)} \leftarrow$ Equalize(T_i, CDF_i, L)
End for

 # Boundary Smoothing
For each pixel p in image I:
$I_{(CLAHE)}(p) \leftarrow$ BilinearInterpolation($T_i^{(eq)}$, p)
End for

Return $I_{(CLAHE)}$

Definitions:
$ClipAndRedistribute(H_i, C)$ = clips histogram values above C and redistributes the excess uniformly across all bins.
$CumulativeHistogram(H_i)$ = cumulative histogram of tile T_i.
$T_i^{(eq)}$ = equalized tile T_i.
L = number of gray levels, usually 256 for 8-bit images.

This study uses the public medical image dataset "Detection of pulmonary diseases," obtained from the Roboflow platform [20]. The dataset consists of chest X-ray (CXR) images annotated with bounding boxes for object detection. It contains 800 images grouped into five classes, namely Bacterial Pneumonia, Coronavirus, Tuberculosis, Viral Pneumonia, and Normal. To ensure experimental validity, the dataset was divided into three subsets, with 80% allocated for model training, 20% used for validation and hyperparameter tuning, and 10% reserved for final testing on unseen data. All images have varying resolutions and were standardized to 640 × 640 pixels during training and inference, following the standard input size of YOLOv12.

The image pre-processing stage is a crucial step to standardize input and improve model robustness. It begins with converting the input image using resize to 640x640. After that, data augmentation is applied during training to prevent overfitting and improve generalization. A series of

augmentations are applied, namely geometric transformations such as Horizontal Flip, Rotation, Translate, and Scale. Then, photometric changes are made to Hue, Saturation, Value (HSV).

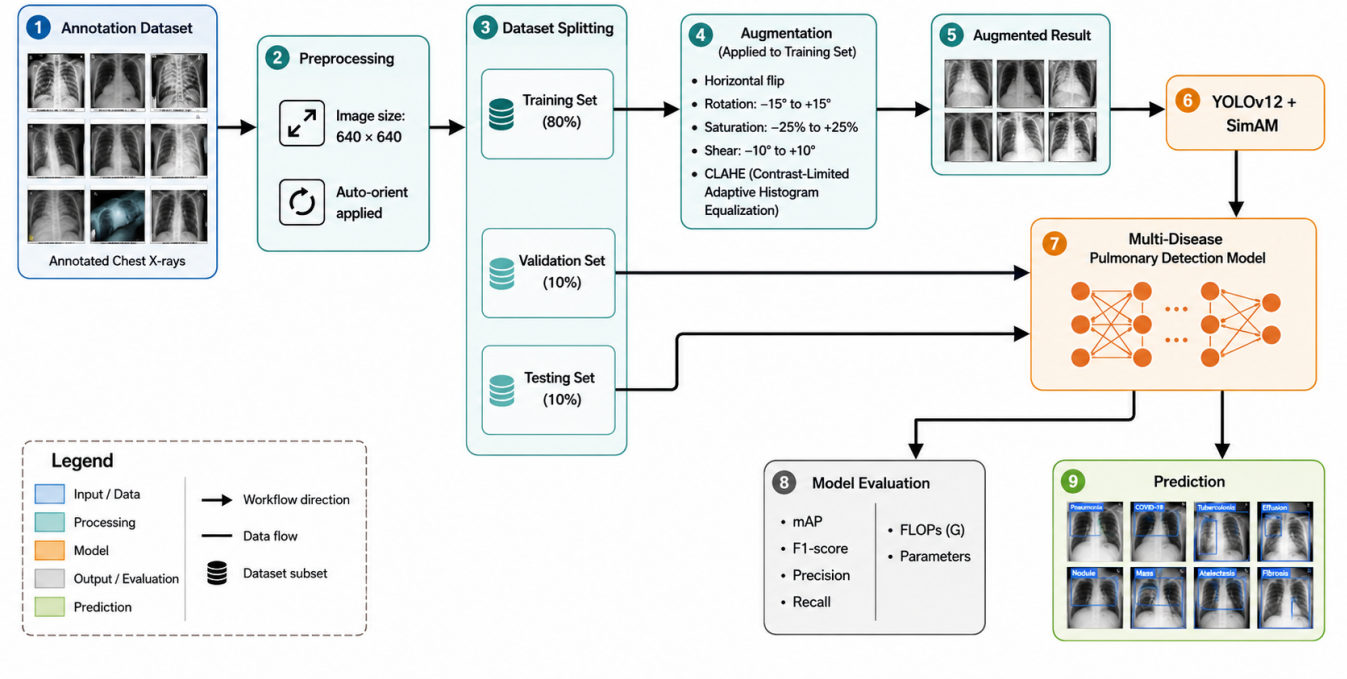

Fig. 1. General Architecture of The Proposed Model

Specifically, to address the main challenge in X-ray images, namely low contrast, based on Table 1 implemented Contrast Limited Adaptive Histogram Equalization (CLAHE) after grayscale conversion. Unlike global histogram equalization, CLAHE works adaptively by dividing the image into tiles (small regions) and applying histogram equalization locally to each tile. A "clip limit" is used to prevent excessive noise amplification in homogeneous areas

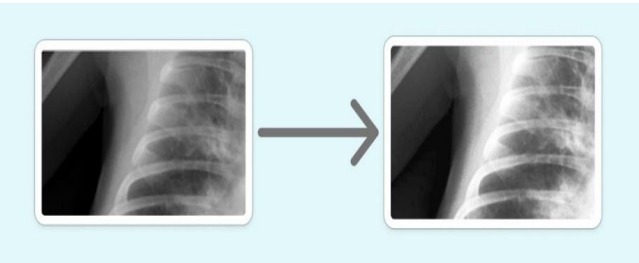

Fig. 2. Comparison of X-ray images before (left) and after (right) applying CLAHE enhancement

Figure 2 Show, this process significantly improves local contrast, sharpens the boundaries of anatomical structures (such as ribs and lungs), and clarifies lesions or infiltrates that may be hidden. The image that has gone through all these steps (Processed Image) then becomes a clean, feature-rich input that is ready to be processed by the ASPIRE-YOLOv12 architecture.

As a baseline, three scaled variants of the standard YOLOv12 architecture were evaluated: YOLOv12-Nano (n), YOLOv12-Small (s). This analysis aims to establish a baseline for performance and identify trade-offs between accuracy, inference speed, and computational complexity (measured in GFLOPs and number of parameters) on the dataset.

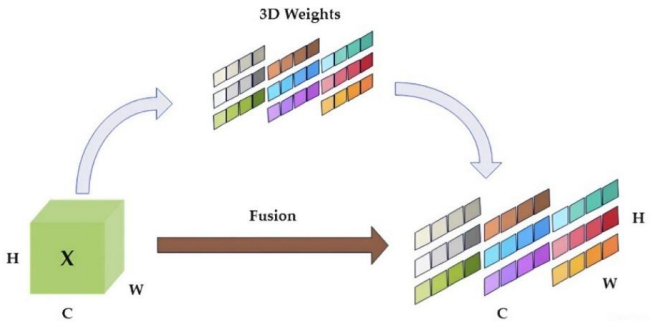

Fig. 3. The SimAM Module Structure [21]

The first modification is the integration of a parameter-free attention module which is illustrated on Figure 3. It uses SimAM (Simple, Parameter-Free Attention Module). Unlike popular attention mechanisms such as SE (Squeeze and Excitation) [22] or CBAM [23], SimAM does not require additional parameters to be trained. This module adopts neuroscience principles, specifically the theory of spatial suppression, to reduce the energy function for each neuron in the feature map. By minimizing this energy function, SimAM analytically calculates the 3D attention weights for each neuron, effectively highlighting the most informative neurons and suppressing less relevant (noise) neurons.

Based on architecture analysis and tuning processes, Figure 4 shows illustration integrated SimAM module at the end of the YOLOv12 backbone, just before the feature map is passed to the neck (FPN/PAN). This location was strategically chosen to filter and refine the highest-level, most semantically rich feature maps before they are used for multi-scale detection.

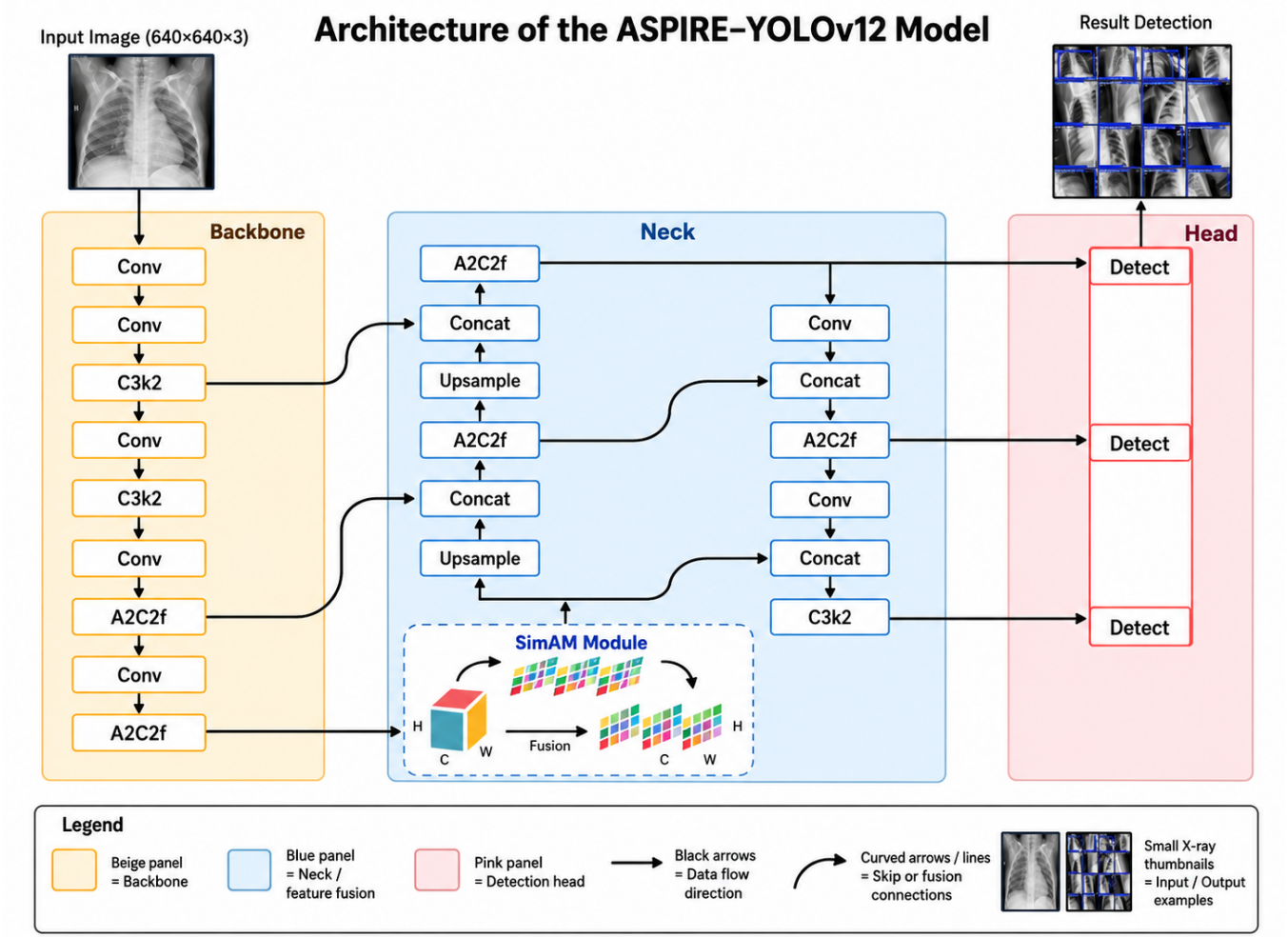

Fig. 4. Architecture Model ASPIRE-YOLOV12 (YOLOv12 + SimAM)

The evaluation of the prediction model utilizes a confusion matrix to map the relationship between predicted outcomes and actual ground truth values. This matrix relies on four fundamental classification indicators, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), which determine the model's correctness. From these indicators, several key performance metrics are derived, specifically precision, recall, and F1-score [24]. Precision is defined as the proportion of true positives relative to all positive predictions (TP + FP), while recall measures the ratio of true positives to the total actual positive instances (TP + FN). The F1-score provides a balanced assessment by combining precision and recall. The mathematical formulations for accuracy, precision, recall, and F1-score are presented in Equations (1), (2), (3), and (4).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Mean Average Precision (mAP) is used as the principal metric for evaluating detection accuracy in object-detection tasks. It is calculated as the mean of the Average Precision (AP) values obtained for all N classes [25].

$$\text{mAP} = \frac{1}{N}\sum_{c=1}^{N} \text{AP}_c \quad (5)$$

In Equation (5), AP denotes the Average Precision for class c, and N denotes the total number of classes. In object detection, mAP is commonly reported at two evaluation settings: mAP@0.5, which uses an Intersection over Union (IoU) threshold of 0.5, and mAP@0.5:0.95, which averages the AP values over IoU thresholds from 0.5 to 0.95 in increments of 0.05.

III. Result and Discussion

The experimental phase utilized the Google Colaboratory environment, equipped with an NVIDIA Tesla T4 GPU for computational acceleration. Training was conducted over 20 epochs with a batch size of 16. For optimization, the AdamW algorithm was employed, configured with an initial learning rate of 0.001111 and a momentum factor of 0.9. The entire implementation leveraged the PyTorch-based Ultralytics framework, with further specifications detailed in Table 2 below.

TABLE II. Experimental Configuration

Tools/Environment/Framework	Type/Version
OS	Windows version 11
Platform Code	Google Colaboratory
GPU	NVIDIA T4/15 GB memory
Deep Learning Framework	PyTorch version 2.4.0, CUDA 12.1

To validate the contribution of each proposed component, namely Contrast Limited Adaptive Histogram Equalization (CLAHE) in the pre-processing stage and Simple Attention Module (SimAM) in the feature extraction stage, an ablation study was conducted. This experiment aimed to identify the individual and combined effects of these two components on improving detection performance. All tests were conducted while maintaining the basic architecture of YOLOv12 on two model scales (nano and small) and identical hyperparameter configurations.

TABLE III. PERFORMANCE COMPARISON OF ASPIRE-YOLOV12 AND BASELINE MODELS

Model	Precision ↑	Recal ↑	F1-Score ↑	mAP50 ↑	Map 50-95↑	Flops (G) ↓	Params (M) ↓
Yolov12n	70 %	70 %	70 %	65 %	64 %	**6.5**	**2.6**
Yolov12s	73 %	73 %	73 %	68 %	67 %	21.4	9.3
ASPIRE-YOLOv12	**81 %**	**82 %**	**82 %**	**77 %**	**76 %**	10.2	3.3

Table III compares the performance of the evaluated models. The ASPIRE-YOLOv12 model shows better performance, achieving an mAP50 of 77% and an mAP50-95 of 76%. Overall, the ablation study results confirm that the integration of these two components directly contributes to improved accuracy in multi-lung disease detection on the YOLOv12 architecture without compromising model efficiency. Complete evaluation indicates that ASPIRE-YOLOv12 outperforms the baseline YOLOv12n and YOLOv12s models. This improvement is associated with the integration of CLAHE as a preprocessing method and SimAM as an attention module, which are designed to address low contrast and noise in chest X-ray images.

The performance improvement is consistently observed across all key evaluation metrics, highlighting the model's robustness in multi-disease detection tasks. Specifically, ASPIRE-YOLOv12 achieves a precision value of 81%, indicating a high rate of correct positive identifications with minimal false positives, which is crucial in clinical settings to avoid unnecessary treatments. The recall stands at 82%, demonstrating the model's effectiveness in capturing most true positive instances, thereby reducing the risk of missed diagnoses for conditions like tuberculosis or coronavirus. Complementing these, the F1-score of 82% reflects a balanced harmony between precision and recall, making the model reliable for real-world applications where both false alarms and oversights carry significant consequences. The mAP results provide further evidence of the model's detection capability. An mAP50 of 0.77 indicates strong performance at an IoU threshold of 0.5, whereas an mAP50-95 of 0.76 represents the average precision across IoU thresholds ranging from 0.5 to 0.95.

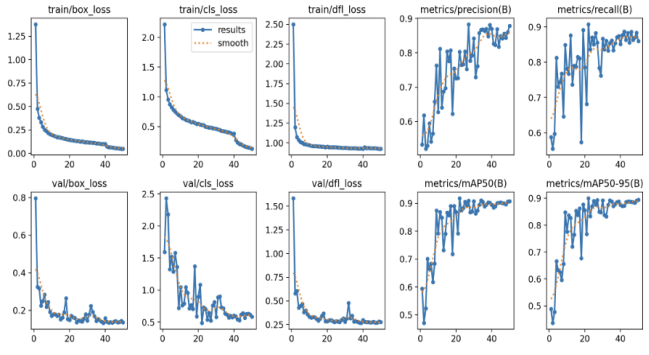

Fig. 5. Training and Validation Loss Curves and Performance Metrics of ASPIRE-YOLOv12

As shown in Figure 5, the training and validation loss curves (train/box_loss, train/cls_loss, train/dfl_loss, val/box_loss, val/cls_loss, val/dfl_loss) show a steady decline throughout the training process until convergence is reached at epoch 40. This downward trend is accompanied by a consistent increase in the precision, recall, mAP@0.5 and mAP@0.5:0.95 metrics, indicating that the integration of CLAHE in the preprocessing stage and SimAM in the backbone successfully improved the quality of feature representation and reduced the risk of overfitting.

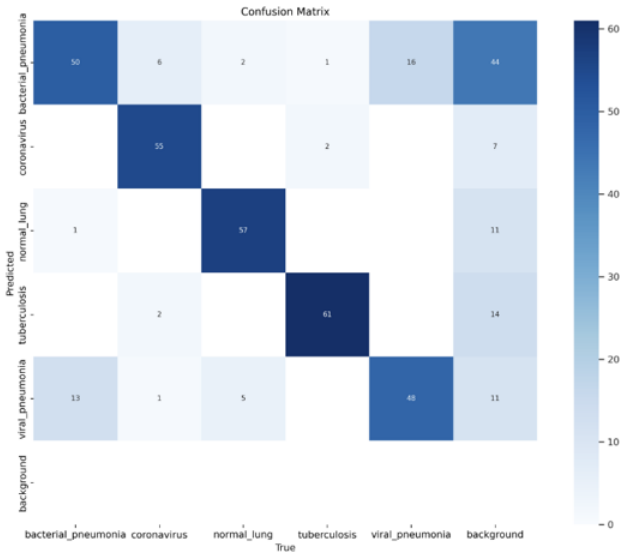

Fig. 6. Confusion Matrix Results of ASPIRE-YOLOv12 on Multi-Disease Pulmonary Detection

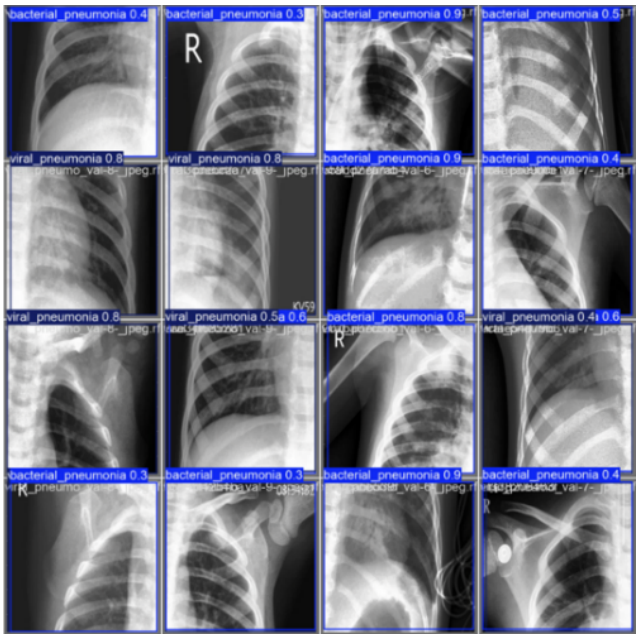

Fig. 7. Visualization of ASPIRE-YOLOv12 detection results on X-ray test images

Furthermore, Figure 6 displays confusion matrix results prediction on data tests. The ASPIRE-YOLOv12 model is capable of classifying five categories of lung disease with high accuracy, where the bacterial pneumonia class obtained approximately 50 true positives with a low error rate, while the coronavirus and tuberculosis classes achieved 55 and 61 true positives, respectively. The dominant diagonal distribution in the confusion matrix indicates the model's ability to distinguish radiographic patterns between diseases with a low misclassification rate.

The visualization of detection results in Figure 7 shows that ASPIRE-YOLOv12 can detect various lesion areas with high confidence scores, ranging from 0.7 to 0.9, especially in images with low contrast. This proves the effectiveness of the combination of CLAHE based contrast enhancement and SimAM based spatial attention in clarifying anatomical boundaries and strengthening multi-disease detection in medical X-ray images.

IV. CONCLUSION

This study successfully proposed and evaluated ASPIRE-YOLOv12, a hybrid architecture for multi-disease detection in lung X-ray images. This model combines two main components, namely Contrast Limited Adaptive Histogram Equalization (CLAHE) as pre-processing to improve image contrast, and the integration of Simple Attention Module (SimAM) as a parameter-free attention mechanism on the YOLOv12 backbone to strengthen the focus on important features. The results of the experiment show that the proposed model provides a significant performance improvement compared to the baseline YOLOv12 model. The ASPIRE-YOLOv12 model achieves a precision of 81%, recall of 82%, F1-score of 82%, mAP50 of 77%, and mAP50-95 of 76%, which indicates an increase of about 9 points in mAP50 compared to the baseline. This improvement confirms that the synergy between image quality enhancement through CLAHE and feature representation strengthening through SimAM effectively improves detection accuracy in the medical domain with complex visual conditions. Moving forward, future research can focus on optimizing inference speed and reducing model complexity to enable real-time implementation in clinical settings. Thus, ASPIRE-YOLOv12 has the potential to become the foundation for the development of efficient, interpretable, and reliable automated diagnostic support systems in modern radiology practice.

ACKNOWLEDGE

We gratefully acknowledge the valuable academic guidance, support, and encouragement provided by the Artificial Intelligence Siliwangi Research Group and Telkom University. The assistance received from the group and department played an important role in the development, implementation, and successful completion of this project.

REFERENCES

[1] Z. Cai, K. Zhou, and Z. Liao, "A Systematic Review of YOLO-Based Object Detection in Medical Imaging: Advances, Challenges, and Future Directions," 2025, *Tech Science Press*. doi: 10.32604/cmc.2025.067994.

[2] C. M. Jones *et al.*, "Chest radiographs and machine learning – Past, present and future," Aug. 01, 2021, *John Wiley and Sons Inc*. doi: 10.1111/1754-9485.13274.

[3] R. Bista *et al.*, "Advancing Tuberculosis Detection in Chest X-rays: A YOLOv7-Based Approach," *Information (Switzerland)*, vol. 14, no. 12, Dec. 2023, doi: 10.3390/info14120655.

[4] C. T. Yen and C. Y. Tsao, "Lightweight convolutional neural network for chest X-ray images classification," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-80826-z.

[5] X. Wang, H. Li, X. Yue, and L. Meng, "A comprehensive survey on object detection YOLO," 2023. [Online]. Available: http://ceur-ws.org

[6] Y. Tian, Q. Ye, and D. Doermann, "YOLOv12: Attention-Centric Real-Time Object Detectors," *ArXiv*, 2025, doi: 10.0.

[7] R. Khanam and M. Hussain, "A Review of YOLOv12: Attention-Based Enhancements vs. Previous Versions," Apr. 2025, [Online]. Available: http://arxiv.org/abs/2504.11995

[8] Liu Xiwei and Nguyen Thao Do Chi, "Medical Images Enhancement by Integrating CLAHE with Wavelet Transform and Non-Local Means Denoising," *Academic Journal of Computing & Information Science*, vol. 7, no. 1, 2024, doi: 10.25236/ajcis.2024.070108.

[9] A. A. Murat and M. S. Kiran, "A comprehensive review on YOLO versions for object detection," Oct. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.jestch.2025.102161.

[10] B. Zhao, L. Chang, and Z. Liu, "Fast-YOLO Network Model for X-Ray Image Detection of Pneumonia," *Electronics (Switzerland)*, vol. 14, no. 5, Mar. 2025, doi: 10.3390/electronics14050903.

[11] M. Parveen Rahamathulla, W. R. Sam Emmanuel, A. Bindhu, and M. Mustaq Ahmed, "YOLOv8's advancements in tuberculosis identification from chest images," *Front. Big Data*, vol. 7, 2024, doi: 10.3389/fdata.2024.1401981.

[12] Z. Mustafa and H. Nsour, "Using Computer Vision Techniques to Automatically Detect Abnormalities in Chest X-rays," *Diagnostics*, vol. 13, no. 18, Sep. 2023, doi: 10.3390/diagnostics13182979.

[13] S. Hao, X. Li, W. Peng, Z. Fan, Z. Ji, and I. Ganchev, "YOLO-CXR: A Novel Detection Network for Locating Multiple Small Lesions in Chest X-Ray Images," *IEEE Access*, vol. 12, pp. 156003–156019, 2024, doi: 10.1109/ACCESS.2024.3482102.

[14] X. Zhang, L. Liu, X. Yang, L. Liu, and W. Peng, "NSEC-YOLO: Real-time lesion detection on chest X-ray with adaptive noise suppression and global perception aggregation," *J. Radiat. Res. Appl. Sci.*, vol. 18, no. 1, p. 101281, Mar. 2025, doi: 10.1016/j.jrras.2024.101281.

[15] K. Halloum and H. Ez-Zahraouy, "Enhancing Medical Image Classification through Transfer Learning and CLAHE Optimization," *Curr. Med. Imaging Rev.*, vol. 21, Jan. 2025, doi: 10.2174/0115734056342623241119061744.

[16] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 2023, [Online]. Available: http://arxiv.org/abs/1706.03762

[17] M. H. Guo *et al.*, "Attention mechanisms in computer vision: A survey," Sep. 01, 2022, *Tsinghua University*. doi: 10.1007/s41095-022-0271-y.

[18] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks," *Proceedings of the 38th International Conference on Machine Learning*, 2021.

[19] S. M. Pizer *et al.*, "Adaptive histogram equalization and its variations," *Comput. Vis. Graph. Image Process.*, vol. 39, no. 3, pp. 355–368, 1987, doi: https://doi.org/10.1016/S0734-189X(87)80186-X.

[20] "Detection of Pulmonary Diseases." Accessed: Nov. 28, 2025. [Online]. Available: https://app.roboflow.com/5-bersaudara/detection-of-pulmonary-diseases-bdhuj/2

[21] J. Shang, J. Wang, S. Liu, C. Wang, and B. Zheng, "Small Target Detection Algorithm for UAV Aerial Photography Based on Improved YOLOv5s," *Electronics (Switzerland)*, vol. 12, no. 11, Jun. 2023, doi: 10.3390/electronics12112434.

[22] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," May 2019, [Online]. Available: http://arxiv.org/abs/1709.01507

[23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.06521

[24] Y. Xie, B. Zhu, Y. Jiang, B. Zhao, and H. Yu, "Diagnosis of pneumonia from chest X-ray images using YOLO deep learning," *Front. Neurorobot.*, vol. 19, 2025, doi: 10.3389/fnbot.2025.1576438.

[25] S. Rajaraman, Z. Liang, Z. Xue, and S. K. Antani, "Ensembled YOLO for multiorgan detection in chest x-rays," in *HHS Public Access*, SPIE-Intl Soc Optical Eng, Feb. 2025, p. 24. doi: 10.1117/12.3047210.