JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. 8, No. 3, Desember 2025, pp. 275-284 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v8i3.10891 e-ISSN: 2656-1948

DETECTION AND CLASSIFICATION OF GRAM-STAINED BACTERIA IN MICROSCOPIC IMAGES USING YOLOV8 WITH CBAM

Karyna Budi Sanjaya[1*], Daniel Martomanggolo Wonohadidjojo[2]

[1,2]School of Information Technology, Universitas Ciputra Surabaya
Email: [*1]ksanjaya04@student.ciputra.ac.id, [2]daniel.m.w@ciputra.ac.id

(Received: 22 October 2025, Revised: 4 November 2025, Accepted: 24 November 2025)

Abstract

Bloodstream infection accounts for approximately 11 million deaths annually, and yet conventional blood culture methods require 40-48 hours to complete pathogen identification which delays definitive therapeutic decisions. Gram staining does provide preliminary bacterial classification within hours, but manual interpretation still remains a labor-intensive task and is prone to variability. This study develops an automated bacterial detection and classification system by integrating CBAM into the YOLOv8 architecture. The model was trained on Gram-stained microscopic images across four bacterial categories: Gram-positive cocci, Gram-negative cocci, Gram-positive bacilli, and Gram-negative bacilli. Dataset preprocessing involved quality selection, noise reduction, and targeted augmentation to address severe class imbalances. The inclusion of CBAM improved feature discrimination and localization performance, with an increase of 1.4% in mAP@0.5:0.95 (from 70.8% to 72.2%). The proposed model also reduced cross-class misclassifications, particularly among morphologically similar cocci. These findings demonstrate that integrating lightweight attention mechanisms can enhance bacterial detection reliability in microscopic imaging and support the development of automated systems for faster, more consistent preliminary bacterial identification.

Keywords: *YOLOv8, CBAM, Gram-stained microscopy, bacterial detection, deep learning, object detection.*

**Corresponding Author: Daniel Martomanggolo Wonohadidjojo*

1. INTRODUCTION

One of the critical causes of morbidity and mortality worldwide, bloodstream infection (BSI) accounts for approximately 50 million cases and 11 million deaths annually, with an incidence rate reaching 150 per 100,000 individuals [1], [2]. Appropriate treatment during the first 48 hours of admission is strongly associated with the survival rate among patients with positive blood cultures [3]. However, conventional identification of bacteria through blood culture requires 16 to 24 hours for pathogen identification and an additional 24 hours for antimicrobial susceptibility testing (AST), effectively delaying definitive therapeutic decisions [4].

One of the fundamental methods used in microbiology laboratories to accelerate preliminary bacterial identification is Gram staining. This technique classifies bacteria into Gram-positive and Gram-negative based on cell wall differences, information that not only influences bacterial response to antibiotics but also provides early guidance for more targeted empirical therapy and can reduce unnecessary use of broad-spectrum agents [5]. Several studies have validated the clinical significance of it. Yamamoto et al. [6] reported that Gram-stain evaluation can guide early antimicrobial therapy even before culture confirmation. Similarly, in acute cholangitis, Tian et al. [7] reported that infections with mixed Gram-negative and Gram-positive bacteria were associated with worse outcomes, highlighting the importance of early Gram-based assessment in determining appropriate empiric coverage.

However, interpretation of Gram-stained slides is not an easy task. It remains a manual, labor-intensive process that depends on the expertise of a trained microbiologist. Each Gram-stained slide is subject to variability and errors due to inconsistent staining quality, uneven focus, and background artifacts [8]. To address this limitation, deep learning offers a promising approach for automating microorganism detection from microscopic images.

In recent years, object detection algorithms have gained increasing attention within deep learning–

based medical image analysis, due to their ability to classify and localize visual patterns simultaneously. The You Only Look Once (YOLO) framework is one of the most promising, providing real-time detection with strong accuracy and efficiency. [9].

Among the various studies using YOLO, Chin et al. [10] applied YOLOv4 to detect growth stages of Escherichia coli in microscopic images. With a mAP of 98%, Chin et al. 's model certainly performed well within its scope. But the study only examined E coli. Which leaves open the question of whether YOLO can handle the morphological diversity found in real Gram-stained specimens [11]. Or the complexity of mixed bacterial forms that appear under varying staining conditions and image quality.

YOLOv8, one of the most popular versions of YOLO, offers architectural advantages over the earlier versions. It offers better feature pyramid networks, improved cross-stage connections, and more efficient backbone processing. In theory, these changes should help detect small, crowded objects. But theory needs testing. Not many studies are found that have systematically tested YOLOv8 on Gram-stained bacteria, where detection accounts for not only the shape variation, but also for the Gram-positive or Gram-negative distinction that directly informs antibiotic selection.

This is where attention mechanisms become relevant. Kincaid [12] integrated a Convolutional Block Attention Module (CBAM) into the YOLOv4 framework for colony detection and reported a 9% accuracy increase in accuracy in comparison to the baseline. CBAM enhances both spatial and channel features which could have helped differentiate purple Gram-positive cells from the pink Gram-negative ones. However, its application in YOLOv8 for bacterial classification has not yet been explored. This gap raises two critical questions: How well can YOLOv8 detect and classify bacteria in Gram-stained microscopic images, and does adding CBAM improve its performance?

Unlike earlier studies that applied standard YOLO models without architectural modifications [10, 19, 20], this research makes three distinct contributions: (1) systematic integration of CBAM attention modules into the YOLOv8 backbone specifically for Gram-stained bacterial detection, (2) development of a targeted preprocessing pipeline combining Laplacian-based quality filtering with Non-Local means denoising optimized for microscopic bacterial images, and (3) implementation of morphology-aware augmentation strategies that address severe class imbalances (standard deviation reduced from 16.49% to 4.81%) while preserving Gram-staining characteristics. These combined modifications represent a novel approach to bacterial detection that addresses both architectural and data-level challenges specific to clinical microscopy. It also has the potential to speed up the identification of bacteria in clinical settings, which would lessen the workload for laboratory staff who interpret the results and allow for quicker therapeutic decisions.

2. RESEARCH METHOD

A quantitative experimental approach is followed for this study. The workflow of stages includes data collection, data selection, preprocessing, augmentation, model training and model evaluation. Each stage was structured carefully to ensure that the results align with the study objectives.

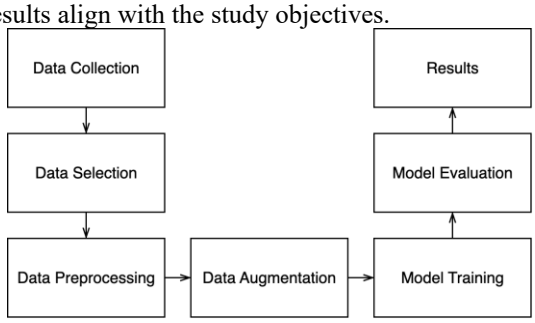

Figure 1. Workflow of Bacteria Detection Model Development

2.1 Data Collection

This study uses X. Wang's Clinical Bacteria Dataset [13]. The database is publicly accessible, found on Zenodo, and does not require institutional ethical approval. The collection includes 6,005 Gram-stained microscopic images with 11,824 annotated four bacterial instances. The four bacterial categories annotated are Gram-positive cocci (G+ cocci), Gram-negative cocci (G– cocci), Gram-positive bacilli (G+ bacilli). Expert microbiologists drew bounding boxes around instances of each bacterial cell type. Wang provided the data pre-split into train, validation and test subsets. Figure 2 provides a few representative examples of the annotated microscopic images from the dataset.

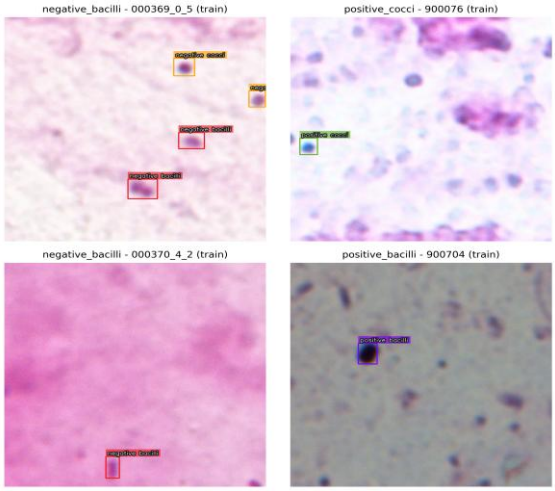

Figure 2. Annotated Images of the Clinical Bacteria Dataset by X. Wang

The dataset used in this study is the Clinical Bacteria Dataset released by Wang et al. (2024) [13] on Zenodo (DOI: 10.5281/zenodo.10526360). It is publicly available under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. All

images were de-identified and distributed for open research use, as stated in the original publication. This study used only that public dataset and did not involve new human or animal data collection.

2.2 Data Selection

The selection stage aims to remove unusable images, starting with manual screening. The images are deemed unusable and will be removed based on their focus quality and artifact interference. The focus quality used the Laplacian variance to check. The images that scored below 100 are deemed too blurry for reliable annotation and were discarded. For the artifact interference, if images containing staining debris, precipitate or uneven dye coverage obscuring more than 30% of the field, the images were discarded. The images were carefully screened and discarded to make sure the detection model does not learn on corrupted data.

This process removed 37% of the original dataset, leaving 5,517 images with adequate clarity and staining quality. The dataset was then reorganized into train, validation and test subsets with a ratio of 75%, 12.5% and 12.5% respectively. The training subset had to be large enough for the model to learn diverse bacterial morphologies, while the validation and test subsets had to remain independent and just enough for a reliable performance evaluation. Table 1 shows the distribution of the dataset across these splits as well as before and after augmentation.

Table 1. Dataset Distribution Across Processing Stages

Dataset Split	Initial Dataset	After Selection	After Augmentation
Training	4,203	3,721	7,222
Validation	1,601	689	689
Testing	990	695	695
Total	**6,794**	**5,517**	**8,606**

2.3 Data Preprocessing

Continuing, all images in the dataset underwent preprocessing to reduce noise and enhance edge clarity. The process began with Non-Local Means denoising (the color variant, fastN1MeansDenoisingColored), to smooth the background grain without destroying fine cellular details. This was followed by sharpening using an unsharp mask. Parameter tuning was performed on a validation subset of 100 images through grid-search optimization. The final parameter chosen being a denoising strength of 8 and sharpening values of 3 for both intensity and Gaussian σ. All three subsets went through the same preprocessing before the augmentation stage, to maintain fair conditions for later performance evaluation.

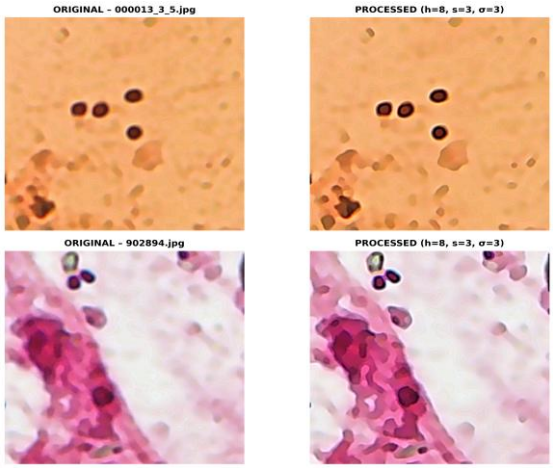

Figure 3. Sample Images Before and After Applying Sharpness and Denoising

2.4 Data Augmentation

After an initial examination of the training subset was completed, it was discovered that there was a substantial class imbalance. G– bacilli accounted for just over half the instances (50.6%), G– cocci for roughly a quarter (28.2%), while G+ cocci and G+ bacilli lagged at 12.5% and 8.6% respectively. To fix the imbalance and to improve the model's generalization, augmentation was applied to the train subset. Validation and test data were left alone as modifying the evaluation subsets would compromise the integrity of the evaluation. Figure 4 shows the original instance distribution in the train subset, where the imbalance can be clearly seen.

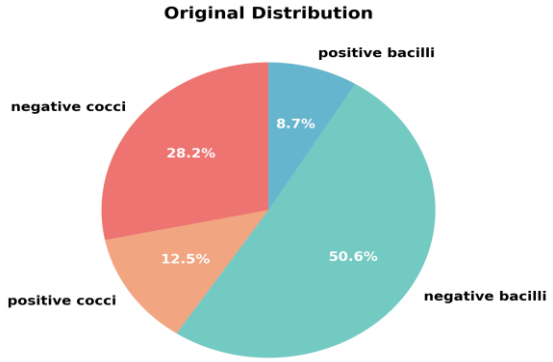

Figure 4. Instance Distribution Across Classes in the Training Subset Before Augmentation

The augmentation technique used in this study involved geometric and photometric transformation implemented in OpenCV. These transformations included random rotations of ±15°, ±30°, 90°, 180°, and 270°, horizontal and vertical flips, and brightness or contrast adjustments. The goal was to simulate realistic shifts in microscope orientation and lighting. While also making sure that the bacterial morphology or Gram-staining characteristics remain intact. New bounding box coordinates were also calculated to maintain label consistency for rotations and vertical flips.

The augmentation stage nearly doubled the training set, expanding it from 3,593 images to 7,222. Bacterial instances increased by 75.9% from 6,468 to 12,252. The extra samples focused on the underrepresented classes (G+ cocci and G+ bacilli) which brought their counts much closer to the dominant classes. Class distribution became more balanced. The standard deviation of the instance distribution dropped from 16.49% to 4.81%. Figure 5 shows the instance distribution across the train subset after the augmentation process.

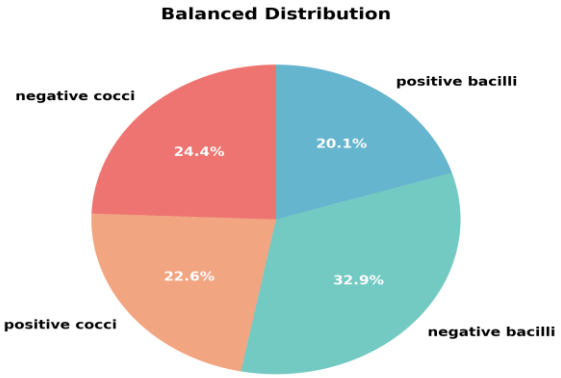

Figure 5. Instance Distribution Across Classes in the Training Subset After Augmentation

Before continuing, a qualitative inspection of the train subset was performed to confirm that the augmented images retained the morphological and Gram staining features of the bacteria. Validation and test subsets remained untouched and unmodified to maintain their role as objective measures of how the model performs on original, real-world data.

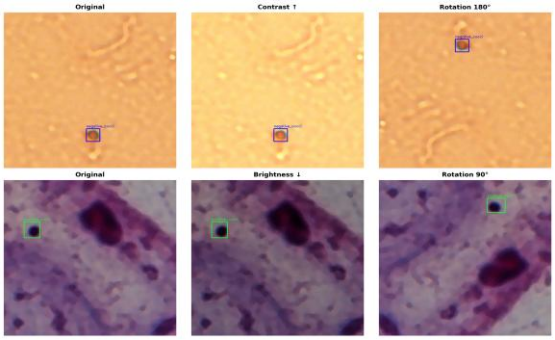

Figure 6. Sample Images of the Implemented Augmentation Strategy Applied to the Training Subset

2.5 Model Training

This study used YOLOv8 as its detection framework. It is recognized for balancing speed with accuracy in real-time tasks. Unlike from its previous versions, which rely on predefined anchor boxes, YOLOv8 predicts bounding boxes directly from object centers and dimensions. This anchor-free approach considerably simplifies the detection pipeline, making training easier and improving generalization. The architecture consists of three components: a backbone for feature extraction, a neck for multi-scale feature fusion, and a head for producing bounding box coordinates, confidence scores, and class probabilities. Figure 7 illustrates the YOLOv8 architecture.

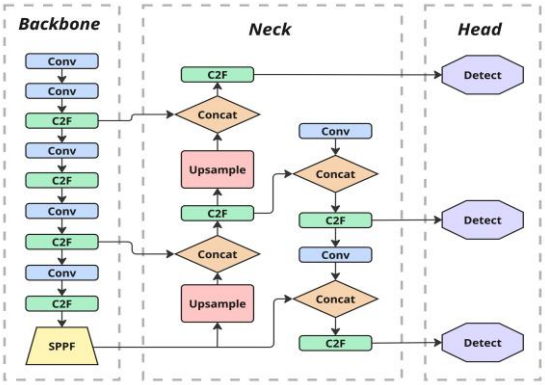

Figure 7. Structure of YOLOv8

CBAM is integrated into the foundation of the standard YOLOv8 architecture in the second model. By continuously applying channel and spatial attention, CBAM improves feature representation by enabling the network to suppress irrelevant features and concentrate on informative regions. This mechanism is potentially helpful for microscopic bacterial images, where the objects are small, and often faint, or overlapping. By guiding attention toward critical morphological and Gram staining features, CBAM is expected to improve detection accuracy without appreciably increasing computational complexity.

Both models were trained with the same configurations detailed in Table 2. The training ran for 150 epochs at 640 x 640-pixel resolution to preserve the bacterial morphology. Hardware acceleration was provided by an Apple M4 Pro using Metal Performance Shaders, with mixed precision enabled for memory efficiency. Standard augmentation techniques such as mosaic, mixup, and copy-paste were disabled to prevent unrealistic bacterial arrangements that do not reflect actual microscopic conditions. Both models took approximately 8-10 hours to complete training.

Table 2. Model Training Configuration for both YOLOv8 and YOLOv8 integrated CBAM Model

Parameter	Value
Training Epochs	150
Image Size	640 x 640 px
Batch Size	8
Optimizer	AdamW
Workers	4
Random Seed	42
Mixed Precision (AMP)	Enabled
Hardware	Apple M4 Pro (MPS)
Augmentation	Disabled

2.6 Model Evaluation

The evaluation of the models will be done using standard object detection metrics: precision, recall, and mean average precision (mAP). Precision measures how many of the detected microorganisms were correct and recall measures how many of the real ones the model was able to detect. Mean average

precision is calculated at two levels: mAP@0.5, which uses a single IoU threshold of 0.5, and mAP@0.5:0.95, which averages performance across IoU thresholds ranging from 0.5 to 0.95. Together these metrics provide a fuller picture of detection accuracy under both lenient and strict localization standards.

Beyond these metrics, confusion matrices will be examined to pinpoint recurring misclassifications across bacterial categories. Finally, an ablation study contrasts the unmodified YOLOv8 with the CBAM-integrated variant, isolating the specific impact of adding attention mechanisms to the architecture.

3. RESULTS AND DISCUSSION

3.1 Model Evaluation

Training dynamics for both models were tracked using box loss and classification loss on the training and validation subsets. Figures 8-11 illustrate how these losses evolved for both baseline YOLOv8 model and YOLOv8-integrated CBAM model, showing how each architecture learned over the course of training.

First to analyze YOLOv8 training trajectory. From figures 8 and 9, it is shown there is a consistent downward trend in both box and classification losses. The curves suggest stable training, where the validation losses show small variations, most seen in the classification loss. These patterns suggest the model effectively minimized both types of error throughout the training phase.

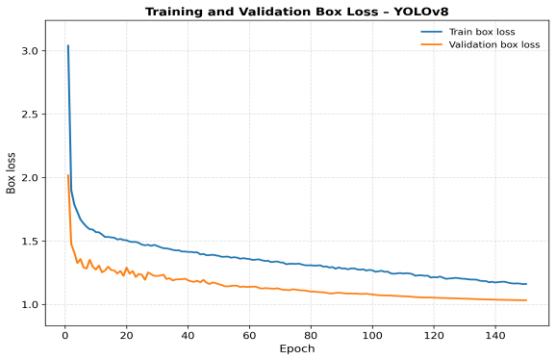

Figure 8. Training and Validation Box Loss Graph of the YOLOv8 model

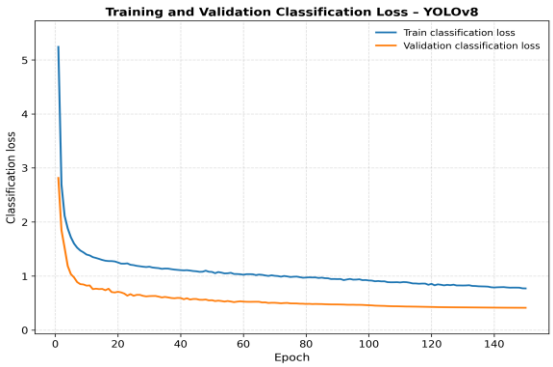

Figure 9. Training and Validation Classification Loss Graph of the YOLOv8 model

Analyzing the YOLOv8 integrated CBAM model, figures 10 and 11 shows both box and classification losses also decreased steadily across epochs. When compared to the baseline YOLOv8 model, the loss curves appear slightly smoother and less erratic. These patterns suggest that the model trained with greater stability.

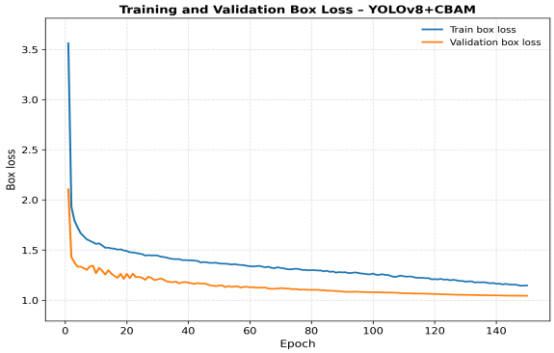

Figure 10. Training and Validation Box Loss Graph of the YOLOv8 integrated CBAM model

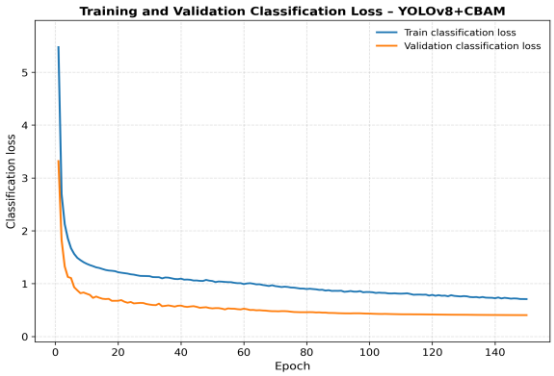

Figure 11. Training and Validation Classification Loss Graph of the YOLOv8 integrated CBAM model

Next the validation accuracy metrics are examined to gauge detection performance over the training process. Figure 12 displays metrics: precision, recall, mAP@0.5 and mAP@0.5:0.95, for baseline YOLOv8 model in a 2x2 layout. All four metrics rose steeply within the first 10 epochs before plateauing later in training. Precision settled above 0.9, Recall and mAP@0.5 climbed near 1.0 with little fluctuations, and mAP@0.5:0.95 hovered around 0.7 near toward the end. The curves suggest consistent learning with slight erratic behavior.

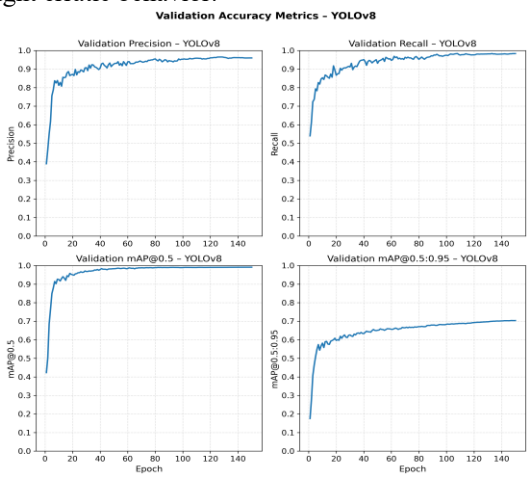

Figure 12. Precision, Recall, mAP@0.5, and mAP@0.5:0.95 Validation of the baseline YOLOv8 model

Figure 13 displays metrics: precision, recall, mAP@0.5 and mAP@0.5:0.95, for the YOLOv8 integrated CBAM model in a 2x2 layout. Similar to the baseline model, all four metrics rose steeply within the first 10 metrics and plateaued later in training. Precision settled above 0.9, Recall and mAP@0.5 climbed near 1.0 with little fluctuations, and mAP@0.5:0.95 hovered around 0.7 near toward the end, slightly higher compared to the baseline. The curves suggest consistent learning with slightly erratic behavior, a little less compared to baseline.

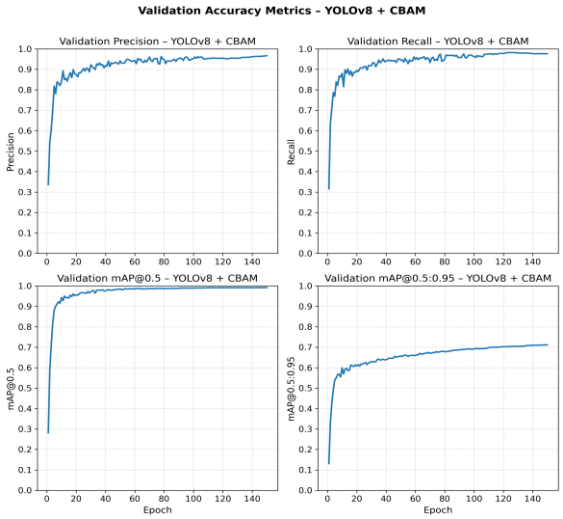

Figure 13. Precision, Recall, mAP@0.5, and mAP@0.5:0.95 Validation of the YOLOv8 integrated CBAM model

Both models appear similar, showing steady gains in precision, recall, and mAP metrics. Each metrics stabilized once convergence was reached. YOLOv8 integrated CBAM model's progression appeared somewhat smoother and more consistent across epochs while still achieving final accuracy levels like the baseline.

Next, both models were tested on the test subset. Tables 3 and 4 break down detection performance for baseline YOLOv8 model and YOLOv8 integrated CBAM model across the four bacterial categories. Metrics reported include precision, recall, mAP@0.5 and mAP@0.5:0.95.

First analyzing the baseline YOLOv8 model's performance. Overall precision at 97.5%, recall at 97.2%, mAP@0.5 at 99.2% and mAP@0.5:0.95 at 70.8%. Among the four categories, G+ cocci and G– bacilli have a slightly weaker mAP@0.5:0.95 performance. This suggests that subtle variations in detection difficulty are tied to the bacterial morphology.

Next to analyze the YOLOv8 integrated CBAM model. Overall precision at 96.4%, recall at 97.9%, mAP@0.5 at 99.2% and mAP@0.5:0.95 at 72.2%. A clear improvement in comparison to the baseline YOLOv8 can be seen in the bacterial class G– cocci and G+ cocci, where mAP@0.5:0.95 increased from 75.2% to 76.6% and from 67.3% to 70.4% respectively. This pattern suggests that CBAM helps the model distinguish classes better that share visual similarities in shape or staining.

Table 3. Evaluation Metric of YOLOv8 Model on the Test Subset

Bacteria Class	Precision	Recall	mAP@0.5	mAP@0.5:0.95
All	0.975	0.972	0.992	0.708
G– cocci	0.938	0.981	0.988	0.752
G+ cocci	0.994	0.950	0.992	0.673
G– bacilli	0.970	0.984	0.993	0.697
G+ bacilli	0.999	0.971	0.995	0.711

Table 4. Evaluation Metric of YOLOv8+CBAM Model on the Test Subset

Bacteria Class	Precision	Recall	mAP@0.5	mAP@0.5:0.95
All	0.964	0.979	0.992	0.722
G– cocci	0.938	0.982	0.991	0.766
G+ cocci	0.994	0.960	0.994	0.704
G– bacilli	0.970	0.984	0.992	0.706
G+ bacilli	0.954	0.989	0.991	0.711

To continue, the confusion matrix produced from testing results from the test subset was analyzed. This offers insight into how the model behaves. Figure 14 displays the results for the baseline YOLOv8 model, where most predictions are seen to be clustered along the diagonal. The model correctly identified 333 Gram– cocci, 152 Gram+ cocci, 620 Gram-negative bacilli, and 101 Gram-positive bacilli samples. There were minor misclassifications primarily between G+ cocci and G– cocci, this reflects the close morphological similarity of coccal structures in Gram-stained images. Some bacilli samples ended up being classified as background, but the number was small and had little effect on the model's aggregate performance.

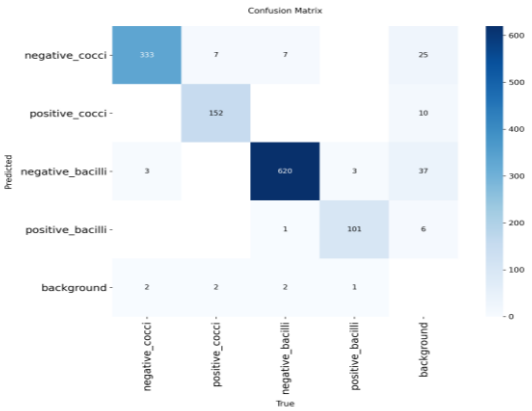

Figure 14. Confusion Matrix of YOLOv8 Model on the Test Subset

Figure 15 displays results for the YOLOv8 integrated CBAM model, like the baseline YOLOv8 model, most predictions are clustered along the diagonal. In comparison to the baseline, the misclassifications between G+ cocci and G– cocci fell from 10 to 3, with G+ bacilli predictions showing greater stability. The diagonal was also stronger

overall, which suggests clearer class separation and a more reliable prediction.

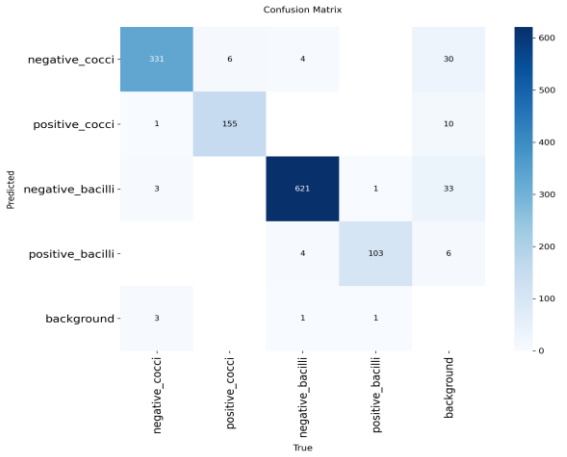

Figure 15. Confusion Matrix of YOLOv8 + CBAM model on the Test Subset

The confusion matrices in Figures 14 and 15 show that most classification mistakes happened between G+ cocci and G– cocci. This pattern appears because the two classes look very similar under different staining conditions. In the baseline YOLOv8 model, several G+ cocci were predicted as G– cocci, likely because of color overlap caused by uneven staining. After adding CBAM, these mistakes dropped from ten cases to three. This shows that the attention module helped the model notice small color and texture details inside coccal cells. The stronger diagonal pattern in Figure 15 confirms that CBAM improved the model's ability to separate classes that look almost the same.

Quantitatively, the CBAM integration reduced the overall misclassification rate between cocci classes by 70% (from 10 to 3 instances). This is clinically significant because G+ and G– cocci require different antibiotic treatments, and misclassification could lead to inappropriate empirical therapy. The reduction in false negatives for G+ bacilli (improved recall from 97.1% to 98.9%) is equally important, as missing these infections in immunocompromised patients can be life-threatening."

3.2 Ablation Study

To further evaluate the effect of integration of the CBAM into the YOLOv8 architecture an ablation study is conducted. Table 5 summarizes the evaluation metrics of both models on the test subset to get a clearer view.

YOLOv8 integrated with the CBAM model showed a slight drop in precision of 1.1%, while recall improved by 0.7%. Both models maintained identical mAP@0.5 score of 99.2% which indicates no change in detection performance at the more lenient IoU threshold. Most notably to mention, is the improvement of mAP@0.5:0.95 which increased by 1.4%. This data means that there is a better localization precision across ranges of IoU thresholds. Although the trade-off in raw precision remained negligible.

Table 5. Evaluation Metrics of the YOLOv8 and YOLOv8 integrated CBAM Models on the Test Subset

Bacteria Class	Precision	Recall	mAP@0.5	mAP@0.5:0.95
YOLOv8	0.975	0.972	0.992	0.708
YOLOv8 +CBAM	0.964	0.979	0.992	0.722

The contribution that CBAM brings likely comes from how it handles attention at the channel level and spatially. CBAM selectively amplifies useful features while also suppressing irrelevant background noise. This pushes the network to focus on distinguishing bacterial characteristics. This resulted in a sharper feature selectivity and tighter localization. The mechanism also appeared to smooth optimization and improve generalization beyond the training samples.

While the ablation study in Table 5 focuses on the internal evaluation between the baseline YOLOv8 and the CBAM-integrated variant, it is also important to contextualize these results with previous YOLO-based studies on bacterial or microscopic image detection. Table 6 summarizes key quantitative findings from recent peer-reviewed works that applied different YOLO versions to similar biomedical imaging tasks. Because most of these studies only report mAP@0.5 and omit mAP@0.5:0.95, the comparison primarily uses mAP@0.5 and other available metrics.

Table 6. Performance comparison of YOLO models on bacterial and microscopic datasets.

Study	Model	Dataset	mAP@0.5
Wang et. al., 2024 [20]	YOLOv5	Gram-stained bacteria (Clinical Bacteria Dataset)	73%
Romphosri et. al., 2024 [22]	YOLOv8 s	Bacterial colonies (E. coli, S. aureus)	97%
Chin et al., 2024 [10]	YOLOv4	E. coli microscopy (growth stages)	98%
Wu et al., 2023 [23]	SDE-YOLO	Blood cell detection (BCCD dataset)	96%
This study	YOLOv8 + CBAM	Gram-stained bacteria (Clinical Bacteria Dataset)	99.2%

The performance comparison reveals several insights. First, when applied to the same Clinical Bacteria Dataset, this study's YOLOv8+CBAM (99.2% mAP@0.5) significantly outperforms Wang et al.'s [20] YOLOv5 baseline (73%), representing a 26.2% improvement. This substantial gap better handles the variable sizes of bacterial cells, (2) CBAM's channel and spatial attention that enhances discrimination of subtle color differences between Gram-positive and Gram-negative bacteria, and (3) the targeted preprocessing pipeline that reduces noise while preserving morphological features.

Second, while Romphosri et al. [22] and Chin et al. [10] achieved high performance (97-98%), their tasks involved more morphologically distinct targets, colony-level detection and single-species

classification, which are inherently less challenging than multi-class Gram-stained bacterial detection with overlapping morphologies. Third, the inclusion of mAP@0.5:0.95 (72.2%) in this study provides a more rigorous evaluation than previous works, as most prior studies only reported mAP@0.5, which uses lenient localization thresholds.

3.3 Discussion

The result of this study indicates that the YOLOv8 integrated CBAM model performed better than the baseline YOLOv8. The 1.4% rise in mAP@0.5:0.95 might not seem like much at first glance. However, in clinical settings where detection systems must function reliably and consistently across a range of IoU levels, it is a noticeable advancement and a tangible step forward. Most of the improvement was observed in coccal bacteria, where both Gram– and Gram+ cocci showed improved detection rates. When considering their morphology, this seems logical. When staining varies even slightly, it becomes difficult for the model to distinguish between them.

The sequential attention structure of CBAM appears to be a good fit for this task. The model benefited from channel attention by prioritizing color information, the purple-to-pink distinction that defines Gram classification. The location of the bacterial cells in the frame was then accentuated by spatial attention, which also muffled background noise. When cells were closely grouped or staining artifacts caused visual interference, this sequential refining was especially helpful. This improved localization in densely packed bacterial clusters that conventional architectures might inadvertently mix.

The mAP@0.5 staying at 99.2% for both models shows that the baseline had already performed near-optimally at the lenient IoU thresholds level. Precision dropped by 1.1% which means that the model occasionally flagged false positives. Even so, the recall improved by 0.7%, which matters more in diagnostics. Missing a bacterial infection carries a higher risk than overcalling one. The pattern is shown more clearly in the confusion matrix: misclassifications between morphologically similar cocci decreased from ten instances to three. Not only did the model learn the obvious distinguishing features, but it also learned the subtler features.

Training dynamics further support this interpretation. The CBAM model's loss curves were slightly smoother, and the validation metrics, steadier. Attention mechanisms may help optimization avoid local minima that trap standard architectures. This aligns with Kincaid's findings [12], who reported a 9% accuracy increase by integrating CBAM to YOLOv4 for colony detection. The smaller improvement likely reflected a harder task with Gram-stained microscopy having finer discrimination than colony morphology.

The extent to which these conclusions can be applied generally is, however, limited by several factors. First off, the dataset only covers four bacterial categories. Clinical samples present much bigger diversity such as mixed infections, rare species and atypical morphologies. Gram staining also varies across labs due to their reagent quality, technique and equipment. Model performance in a controlled experiment doesn't always translate to clinical workflows where imaging conditions fluctuate. Finally, this study also lacks inter-rater reliability data comparing the model to multiple microbiologists. This would clarify whether its errors align with or diverge from human disagreement patterns.

This study suggests a pathway toward semi-automated bacterial screening. The model can pre-classify samples, flag ambiguous cases for experts to review while automatically confirming clear positives or negatives. This would not replace microbiologists but could very well redistribute their time toward the more pressing or complex cases. Integration would require validation against hospital laboratory workflows, regulatory approval, and safeguards preventing over-reliance on automated outputs when edge cases appear.

What remains unclear is how CBAM might react to harder scenarios, from heavily overlapping cells, different techniques or equipment of staining, or other bacterial species outside of the ones in the dataset used. Those conditions define the real diagnostic challenges, and testing there would clarify the model's practical ceiling.

3.4 Clinical Implementation Considerations

The deployment of this system in clinical laboratories would require consideration of several practical factors. First, the model demonstrates sufficient speed for real-time application, processing images at approximately 45 frames per second on standard GPU hardware, making it compatible with automated microscopy workflows. Second, the system should be implemented as a decision-support tool rather than autonomous diagnostic system, where predictions flagged with confidence scores below 0.85 are automatically referred to expert microbiologists for manual review. Third, regulatory compliance with medical device standards (such as FDA Class II or CE marking) would be necessary before clinical adoption. Finally, continuous model monitoring and periodic retraining with new data from diverse laboratories would be essential to maintain performance as staining protocols and equipment evolve across different clinical settings.

4. CONCLUSION

This study explored the integration of CBAM into YOLOv8 for detecting and classifying bacteria in microscopic images. The results showed that YOLOv8 improved the accuracy of bacterial identification when integrated with CBAM. The improved model outperformed the baseline YOLOv8 model by 1.4% in the mAP@0.5:0.95 metric, achieving 99.2% mAP@0.5 and 72.2% mAP@0.5:0.95. Because of the

high morphological overlap between G+ and G-cocci, it was found that coccal bacteria presented genuine identification challenges. The model was able to learn slightly more distinguishing features through the attention mechanism.

During dataset preparation, the data quality also shaped the overall performance. The quality was enhanced by removing 37% of images that were faintly stained or poorly focused. However, this step reduced the model's learning set, which limited the range of conditions it could learn from. The augmentation also properly balanced the instance distribution by nearly doubling the training subset. Although this crafted a more favorable learning environment, the model ended up training only on carefully selected, high-quality samples, meaning it did not experience the noisy and inconsistent conditions typical of real diagnostic laboratories.

Both models achieved the same mAP@0.5 of 99.2%, demonstrating that the baseline YOLOv8 was already capable of meeting lenient detection thresholds. CBAM's advantage became more apparent in the stricter mAP@0.5:0.95 metric, where accurate localization is crucial for downstream analysis. This was also reflected in the training curves where the smoother convergence of the YOLOv8 integrated CBAM model suggests that attention layers prevented optimization from getting trapped in suboptimal regions.

There are several limitations that need to be considered before these findings are implemented. The dataset only included four bacterial categories. When compared to clinical samples that frequently contain ten or more bacterial species (usually in polymicrobial combinations), it is not much, and more bacterial categories need to be further trained on. The images in the dataset come from a single institution in which they follow a single staining procedure, but different laboratories use different microscopes, reagents, and Gram-staining techniques, all of which affect image characteristics. Additionally, the model's performance on slides with a lot of debris, overlapping cells, weak staining, or slides with a lot of debris remains uncertain.

The next step is to move beyond controlled experimental conditions. The scalability of the system should be tested using a minimum of ten different types of bacteria. Integrating the model with automated microscopy techniques could potentially offer clinical value because it would allow for real-time detection while scanning slides. Predicting antibiotic resistance patterns in addition to bacterial identity would have further clinical utility. It is necessary to include validation from various tools and staining procedures to show how effectively the model truly generalizes. Finally, adding a feature where the Artificial Intelligent gives an explanation as to which image regions influenced each decision it took, could help microbiologists and laboratory technicians verify the model's reasoning before finalizing reports.

The feasibility of this approach has been demonstrated by this study. Under controlled circumstances, YOLOv8 integrated with CBAM can precisely detect and classify Gram-stained microorganisms. Whether it holds up in unsupervised clinical use is another question. This question must be addressed in real-world diagnostic settings, where equipment, image quality, and bacterial presentations vary far beyond textbook examples.

Acknowledgment

We deeply appreciate the support that was provided by the School of Information Technology of Ciputra University Surabaya.

5. REFERENCES

[1] K. E. Rudd *et al.*, "Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study," *The Lancet*, vol. 395, no. 10219, pp. 200–211, Jan. 2020, doi: 10.1016/S0140-6736(19)32989-7.

[2] M. Verway *et al.*, "Prevalence and Mortality Associated with Bloodstream Organisms: a Population-Wide Retrospective Cohort Study," *J Clin Microbiol*, vol. 60, no. 4, pp. e02429-21, Apr. 2022, doi: 10.1128/jcm.02429-21.

[3] C. P. Fischer, E. Kastoft, B. R. S. Olesen, and B. Myrup, "Delayed Treatment of Bloodstream Infection at Admission is Associated With Initial Low Early Warning Score and Increased Mortality," *Critical Care Explorations*, vol. 5, no. 9, p. e0959, Aug. 2023, doi: 10.1097/CCE.0000000000000959.

[4] M. Demir and G. Hazırolan, "Rapid Bacterial Identification from Positive Blood Cultures by MALDI-TOF MS Following Short-Term Incubation on Solid Media," *Infect Dis Clin Microbiol*, vol. 6, no. 2, pp. 141–146, Jun. 2024, doi: 10.36519/idcm.2024.319.

[5] H. Ito, Y. Tomura, J. Oshida, S. Fukui, T. Kodama, and D. Kobayashi, "The role of gram stain in reducing broad-spectrum antibiotic use: A systematic literature review and meta-analysis," *Infectious Diseases Now*, vol. 53, no. 6, p. 104764, Sep. 2023, doi: 10.1016/j.idnow.2023.104764.

[6] K. Yamamoto *et al.*, "Accuracy of classification of urinary Gram-stain findings by a computer-aided diagnosis app compared with microbiology specialists," *Journal of Medical Microbiology*, vol. 74, no. 4, Apr. 2025, doi: 10.1099/jmm.0.002008.

[7] S. Tian *et al.*, "Clinical characteristics of Gram-negative and Gram-positive bacterial infection in acute cholangitis: a retrospective observational study," *BMC Infect Dis*, vol. 22, no. 1, p. 269, Dec. 2022, doi: 10.1186/s12879-021-06964-1.

[8] J. McMahon *et al.*, "A novel framework for the automated characterization of Gram-stained blood culture slides using a large-scale vision

transformer," *J Clin Microbiol*, vol. 63, no. 3, pp. e01514-24, Mar. 2025, doi: 10.1128/jcm.01514-24.

[9] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," *GitHub repository*, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[10] S. Y. Chin *et al*., "Bacterial image analysis using multi-task deep learning approaches for clinical microscopy," *PeerJ Computer Science*, vol. 10, p. e2180, Aug. 2024, doi: 10.7717/peerj-cs.2180.

[11] N. A. Megantara and E. Utami, "Object Detection Using YOLOv8 : A Systematic Review," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 3, pp. 1186–1193, May 2025, doi: 10.32520/stmsi.v14i3.5081.

[12] L. Kincaid, "Microbial colony species recognition using an enhanced YOLOv4 algorithm with CBAM and k-means++ optimization," *Transactions on Computational and Scientific Methods*, vol 4, no.10, Oct. 2024. Available: https://pspress.org/index.php/tcsm/article/view/134.

[13] X. Wang, "Clinical Bacteria DataSet." Zenodo, Jan. 18, 2024. doi: 10.5281/ZENODO.10526360.

[14] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Computer Vision – ECCV 2018*, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11211. , Cham: Springer International Publishing, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.

[15] G. Yao, S. Zhu, L. Zhang, and M. Qi, "HP-YOLOv8: High-Precision Small Object Detection Algorithm for Remote Sensing Images," *Sensors*, vol. 24, no. 15, p. 4858, Jul. 2024, doi: 10.3390/s24154858.

[16] J. Yan *et al*., "Enhanced object detection in pediatric bronchoscopy images using YOLO-based algorithms with CBAM attention mechanism," *Heliyon*, vol. 10, no. 12, p. e32678, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32678.

[17] C. P. Fischer, E. Kastoft, B. R. S. Olesen, and B. Myrup, "Delayed Treatment of Bloodstream Infection at Admission is Associated With Initial Low Early Warning Score and Increased Mortality," *Critical Care Explorations*, vol. 5, no. 9, p. e0959, Aug. 2023, doi: 10.1097/CCE.0000000000000959.

[18] T. Jiang, C. Li, M. Yang, and Z. Wang, "An Improved YOLOv5s Algorithm for Object Detection with an Attention Mechanism," *Electronics*, vol. 11, no. 16, p. 2494, Aug. 2022, doi: 10.3390/electronics11162494.

[19] U. Kashino, K. Taira, and K. Hirata, "Detecting Bacteria from Gram Stained Smears Images by the Family of YOLOs," in *Proceedings of the 2024 7th International Conference on Digital Medicine and Image Processing*, Osaka Japan: ACM, Nov. 2024, pp. 6–10. doi: 10.1145/3705927.3705929.

[20] X. Wang *et al*., "A Clinical Bacterial Dataset for Deep Learning in Microbiological Rapid On-Site Evaluation," *Sci Data*, vol. 11, no. 1, p. 608, Jun. 2024, doi: 10.1038/s41597-024-03370-5.

[21] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif Intell Rev*, vol. 57, no. 4, p. 99, Mar. 2024, doi: 10.1007/s10462-024-10721-6.

[22] S. Romphosri, D. Pissuwan, N. Wattanavichean, P. Buabthong, and T. Waritanant, "Rapid alignment-free bacteria identification via optical scattering with LEDs and YOLOv8," *Sci Rep*, vol. 14, no. 1, p. 20498, Sep. 2024, doi: 10.1038/s41598-024-71238-0.

[23] Y. Wu, D. Gao, Y. Fang, X. Xu, H. Gao, and Z. Ju, "SDE-YOLO: A Novel Method for Blood Cell Detection," *Biomimetics*, vol. 8, no. 5, p. 404, Sept. 2023, doi: 10.3390/biomimetics8050404.

[24] M. S. B. Nusantara and D. M. Wonohadidjojo, "CLASSIFICATION OF BONE FRACTURES IN THE WRIST AND HAND USING DENSENET AND XCEPTION," *JIKO*, vol. 8, no. 1, pp. 21–30, Apr. 2025, doi: 10.33387/jiko.v8i1.9201.

[25] D. D. Karyanto, D. D. Karyanto, J. Indra, A. R. Pratama, and T. Rohana, "DETECTION OF THE SIZE OF PLASTIC MINERAL WATER BOTTLE WASTE USING THE YOLOV5 METHOD," *JIKO*, vol. 7, no. 2, pp. 123–130, Aug. 2024, doi: 10.33387/jiko.v7i2.8535.