Deep Learning-based Multi-Class Acne Detection on Facial Images Using YOLOv11

Nayla Nur Fadhillah[1]* Alam Rahmatulloh[1] Neng Ika Kurniati[1]

[1]*Informatics Department, Faculty of Engineering, Siliwangi University, West Java, Indonesia*
* Corresponding author's Email: 217006097@student.unsil.ac.id
(Received: January 3, 2025. Accepted: January 20, 2026. Published: January 20, 2026.)

Abstract

This study implements YOLOv11-based model capable of detecting six types of acne on facial images, namely blackheads, whiteheads, papules, pustules, nodules, and cysts. Unlike prior works focusing on single-class or severity classification, this approach performs multi-class detection. A dataset of 1.884 images sourced from Roboflow was expanded to 6.116 through augmentation and mosaic techniques. The YOLOv11m model, trained for 150 epochs with transfer learning, achieved an mAP@50 of 87,1%, mAP@50–95 of 64,9%, precision of 85%, recall of 82,9%, F1-score of 84%, and an inference speed of 26 FPS, enabling near real-time performance. Results indicate strong accuracy, particularly for cysts (99%), while whiteheads remained challenging (75%). Key limitations include high computational cost, difficulty in detecting small lesions, and limited generalization to unseen data. Future research should enhance small-lesion future extraction, employ knowledge distillation, and integrate dermatologist validation for clinical reliability.

Keywords: Acne, Facial Image, Object Detection, YOLOv11.

1. Introduction

Acne is one of the most common skin condition experienced by adolescents and young adults [1, 2]. It not only causes physical discomfort but also has significant psychological impacts, such as reduces self-esteem and increased risk of stress or depression [3]. Therefore, accurate early detection of acne is an important step in dermatological care. Traditionally, acne detection has been performed manually by dermatologist. However, this approach has several limitations, including dependency on the subjective expertise of individuals, time consumption, and relatively high costs [4, 5]. In recent years, there has been growing trend of skin health education through social media by dermatologist or beauty influencers, reflecting the public needs for accurate and efficient self-detection methods.

With rapid advancement of artificial intelligent (AI), particularly deep learning, there is significant potential to overcome these challenges by providing deep learning-based models capable of detecting acne quickly and accurately. Previous studies have explored various deep learning architectures for acne detection. For example, H. Zhang & Ma [6] utilized a combination of ResNet and YOLOv5 for acne severity classification. Huey Gan et al. [7] applied YOLOv5 to detect multiple skin lesions, such as acne, pigmentation, and fine lines, achieving a mean Average Precision (mAP) of 42.4%. Meanwhile, D. Zhang et al. [8] developed a YOLOv7-based model that reached an mAP of 83.7%, though it was still limited to single-class detection. On other hand, Faruq Aziz & Saputri [9] employed a more efficient approach using YOLOv9, obtaining an mAP of 81.4%. Their model demonstrated strong capabilities in detecting various types of skin lesions, including acne, atopic dermatitis, psoriasis, using images from open-source media datasets available on the Roboflow platform.

Despite the contributions, prior research still shows several limitations regarding the scope of acne detection. No existing study has specifically focused on detecting six distinct types of acne. Most prior works concentrated on single-class or acne severity classification. Research on multi-class classification of acne detection remains scarce, primarily due to high visual similarity between classes and imbalanced data distribution across acne categories. Moreover, the overall performance of previous model remains suboptimal. Although YOLOv5, YOLOv7, and YOLOv9 have demonstrated promising results in detecting acne and other skin lesions, their achieved mAP values generally remain below 85%, leaving considerable room for improvement.

YOLOv11, the latest version of the YOLO family, offers significant improvements in detection speed and accuracy. With features such as the c3k2 backbone module, Spatial Pyramid Pooling Fast (SPPF), and spatial attention block (C2SPA), YOLOv11 is designed to enhance precision in

detecting small objects. According to Sharma et al. [10], YOLOv11 achieved the fastest inference time of 13.5 ms, outperforming its predecessors. Similarly, Khanam et al. [11]demonstrated that YOLOv11 delivers more balanced performance across varying object scales compared to earlier YOLO models. This makes YOLOv11 particularly effective for detecting objects of diverse sizes, a major challenge in medical and dermatological image analysis.

The contributions of this study are threefold: (1) constructing the first harmonized dataset of six acne lesion categories, (2) implementing and optimizing a YOLOv11m framework with targeted augmentation startegies to handle class imbalance, and (3) providing a comprehensive evaluation including per-class analysis, confusion matrix interpretation and comparison with prior YOLO-based studies.

2. Research methods

This study employed an experimental research design to develop a deep learning model for multi-class acne detection. The methodological framework consisted of four sequential stages as illustrated in Fig. 1.

The first stage, data collection, involved acquiring facial acne images and annotating them into six categories (blackheads, whiteheads, papules, pustules, nodules, and cysts). The second stage, dataset augmentation, applied both classical and mosaic augmentation techniques to improve the representation of minority classes and enhance variability across samples.

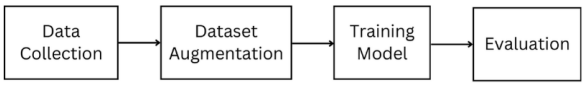

Figure 1. Methodological framework

The third stage, training model, utilized the YOLOv11m architecture with transfer learning and fine-tuning on the augmented dataset under optimized hyperparameter settings. Finally, the fourth stage, evaluation, assessed the trained model using standard object detection metrics including precision, recall, F1-score, and mean Average Precision (mAP), as well as confusion matrix analysis to capture inter-class misclassification patterns. This structured methodology ensured that the proposed model was developed and validated systematically for reliable multi-class acne detection

2.1 Dataset collection

The dataset used in this study was compiled from three publicly available and annotated acne detection datasets hosted on Roboflow platform: ACNEDET V1 [12] (1.705 images), DetectDataSkin-2 [13] (3.462 images), and additional Acne Dataset [14] (381 images). These datasets included multiple lesions categories, some of which overlapped across sources but were inconsistently labeled. For instance, DetectDataSkin-2 contained labels such as pores and acne scars, while ACNEDET V1 included categories like dark spot in addition to acne-related classes.

To build a consistent dataset, all images from the three public sources were combined and their annotations were carefully rechecked and adjusted by the author to ensure uniform labeling . The review process involved checking and adjusting bounding box annotations to ensure label consistency across sources. Non-acne-related labels (e.g., pores, dark spot, acne scars) were excluded, while synonymous acne-related terms were merged. After this harmonization process, six final acne lesion classes were established: blackheads, whiteheads, papules, pustules, nodules, and cysts.

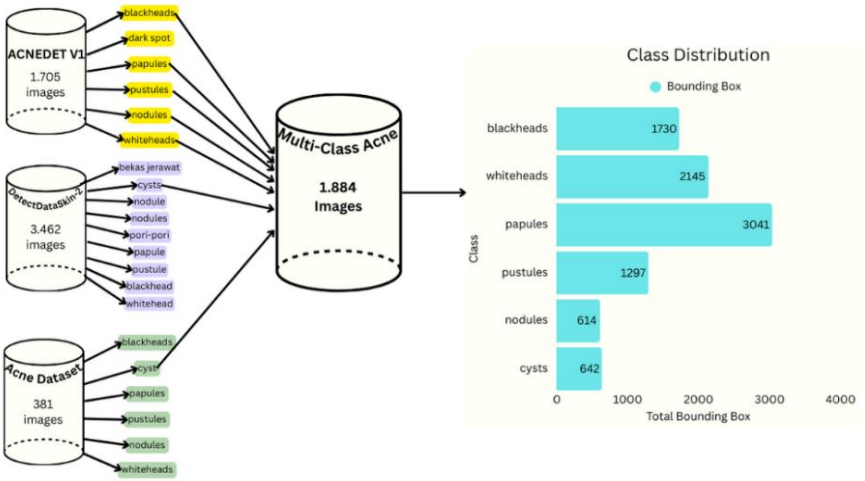

Figure. 2 Dataset harmonization and class distribution across six acne categories before augmentation

 doi: 10.22266/inassexpress.2026.002

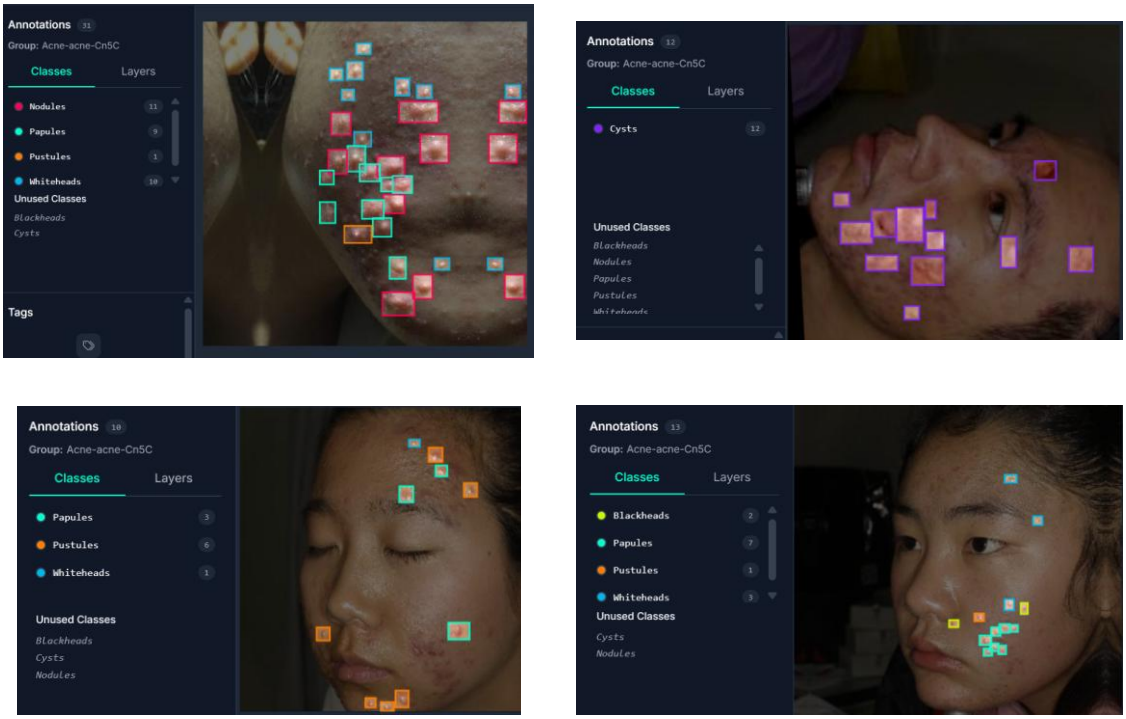

Figure. 3 Representative annotated images of acne lesions across six categories

All annotations were rechecked and, when necessary, corrected manually by the author based on visual inspection. Although this re-annotation process was not conducted by a medical expert, it followed a structured and consistent procedure to ensure logical coherence among the six lesion types. The resulting integrated dataset is referred to as the Multi-class Acne Dataset [15] , consisting of 1,884 images and 9,764 bounding boxes before augmentation (Fig. 2).

To provide a clearer view of the dataset characteristics, several representative annotated images are shown in Fig. 3. These examples illustrate how acne lesions of different categories were labeled with bounding boxes across varying skin tones and lesion sizes. The figure demonstrates the visual challenges of distinguished between acne classes with subtle morphological differences.

2.2 Dataset augmentation

In the initial stage, a classical augmentation strategy was applied to balance bounding box distribution across all six acne categories. Augmentation techniques included random rotation, horizontal and vertical flipping, as well as adjustments in brightness and contrast. This produced an approximately balanced dataset, with ~4,000 bounding boxes per class. However, when the model was trained on this balanced dataset, the overall detection performance showed limited improvement, with mAP@50 plateauing at ~55%. Notably, the most significant gains were observed only for the Cyst class, while other categories exhibited marginal or no improvements.

Based on these observations, only the augmented Cyst images from the first stage were retained and merged back into the base dataset, while the remaining augmented images were discarded. This selective integration preserved improvements in the Cyst class while preventing unnecessary redundancy for other categories.

In the final stage, mosaic augmentation was applied to the entire dataset (base dataset + augmented Cyst images). Mosaic augmentation, which combines four images into one, provided diverse spatial contexts and scale variations, thereby improving the robustness of the model without distorting acne morphology. This process expanded the dataset to a total of 6,116 images, which was subsequently used for the final training.

Fig. 4 illustrates the final distribution of bounding boxes across the six acne categories after the application of mosaic augmentation. Although the class distribution remains imbalanced, with papules and whiteheads dominating the dataset, this augmentation strategy provided richer spatial contexts and scale variations, which ultimately improved the robustness and generalization ability of the YOLOv11m model during training.

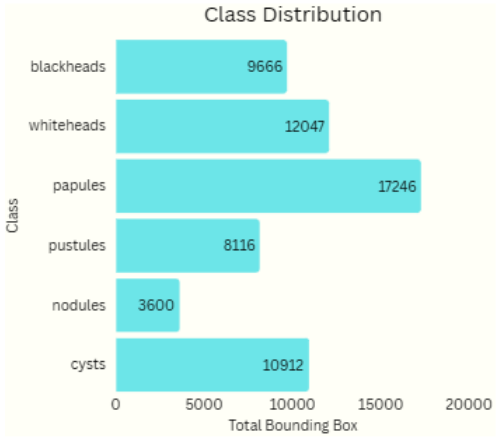

Figure. 4 Final bounding boxes distributions

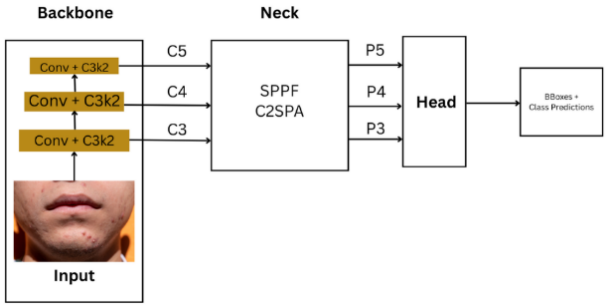

Figure 5. Simplified architecture of YOLOv11

2.3 Model architecture

This study implemented the YOLOv11m object detection framework for acne lesion detection. YOLOv11m is a medium-scale variant in the YOLOv11 family that balances accuracy and computational efficiency, making it well-suited for medical imaging tasks where both precision and practicality are required. The YOLOv11architecture integrates several improvements over its predecessors, including the C3k2 backbone, Spatial Pyramid Pooling Fast (SPPF), and the C2SPA spatial attention block, which together enhance the detection of small-scale objects and improve feature representation across multiple scales [16].

Fig. 5 illustrates that the model retains the canonical three-part structure of the YOLO family, consisting of a backbone, a neck, and a detection head. The backbone maintains convolutional layers for progressive feature extraction but introduces the C3k2 block, which replaces the C2f block used in earlier versions, offering improved computational efficiency by employing two small-kernel convolutions instead of a single large one [11].

The neck integrates the SPPF module to aggregate features across receptive fields and the C2SPA attention mechanism to emphasize salient spatial regions, thereby improving detection of small or partially occluded [17]. Finally, the head follows the conventional YOLO design with predictions at three scales (P3, P4, P5), ensuring robust detection of objects ranging from small to large.

2.4 Training procedure

The training process of YOLOv11m was conducted on Google Colab with CUDA T4 GPU, using Pytorch 2.6 and Python 3.11. The model was trained for 150 epochs with an input image resolution of 800x800 pixels to ensure small acne lesion could still be detected. A batch size of 16 was selected as trade-off between computational efficiency and detection accuracy.

To minimize the risk of overfitting, an early stopping mechanism with a patience value of 20 epochs was applied, allowing the training process to halt if no further improvements in validation performance were observed. The optimization process utilized Stochastic Gradients Descent (SGD), which provides a balance between convergence speed and generalization across large-scale datasets. The learning rate was initialized at 0.001 and adjusted progressively trough a cosine annealing scheduler, a strategy to supports more stable convergence. Additionally, the first three epochs were allocated as warm-up phases, where the learning rate started at a lower value to avoid gradient instability in the initial training stage.

2.5 Evaluation

Evaluating multi-class object detection models requires special attention to the inherent challenges in distinguished visually similar classes and handling class imbalance. According to Kothala & Guntur [18] and Shrawne et al. [19], multi-class object detection becomes particularly difficult when the target objects are small, overlapping, and share high visual similiarity. This condition is highly relevant in acne detection, where lesions such as papules, pustules, and whitehead often exhibit overlapping morphological features, making them prone to misclassification.

Therefore, evaluation in this study was not limit to global performance but also emphasize per-class analysis. Matrics employed included Precision, Recall, F1-Score, and mean Average Precision (mAP). Precision quantifies the accuracy of positive predictions, defined as the ratio of True Positives (TP) to the sum of TP and False Positives (FP) [20], as shown in Eq. (1).

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

 doi: 10.22266/inassexpress.2026.002

Recall measures the model's ability to capture all relevant objects, expressed as the ratio of TP to the sum of TP and False Negative (FN) [21], as given in Eq. (2).

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

Because precision and recall typically exhibit a trade-off, the F1-score balances them using the harmonic mean, as formulated in Eq. (3) [22].

$$F1 - Score = \frac{2\ x\ Precision\ x\ Recall}{Precision+Recall} \quad (3)$$

The primary evaluation metric in this study was mAP, which averages the precision across all classes at different Intersection over Union (IoU) threshold. Both mAP@50 and mAP50-95 were reported following the general formulation in Eq. (4) [23].

$$mAP = \frac{\sum_{i=1}^{c} AP_i}{c} \quad (4)$$

Finally, to further understand misclassification patterns, a confusion matrix was used to capture prediction tendencies between visually similar classes, such as papules being misclassified as pustules, or blackheads as whiteheads. Prior research [24] has shown effectiveness of confusion matrices in evaluating YOLO-based model for fine-grained classification tasks.

3. Results and discussion

3.1 Training and validation loss

The YOLOv11m model was trained for 150 epochs with an average training time of approximately 5 minutes per epoch. Early stopping was not triggered, indicating that the model continued to improve performance until the final stage of training, albeit at a slower rate in later epochs.

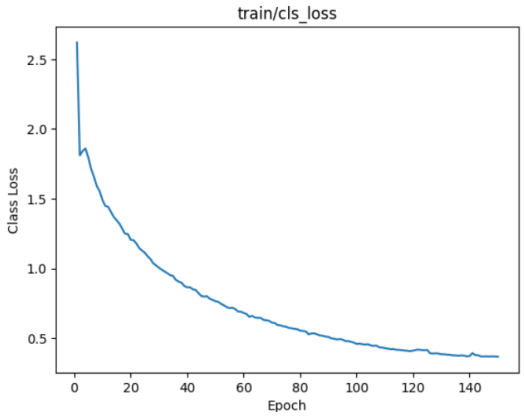

Figure. 6 Training class loss paragraph

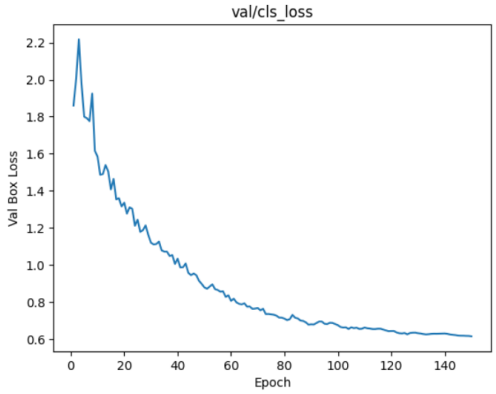

Figure 7. Validation class loss graph

As illustrated in Fig. 6 the training class loss decreased sharply during the first 50 epochs, after which it stabilized with minor fluctuations until epoch 150. Specifically, the training class loss dropped from 2.6199 in the initial epoch to 0.3677 in the final epoch, reflecting effective convergence of the model.

Similarly, Fig. 7 presents the validation class loss curve, which followed a comparable trend. Validation class loss decreased from 1.859 to 0.6155, with a relatively small gap compared to the training curve. This suggests that the model achieved stable generalization capability without significant overfitting.

The consistent reduction across both training and validation losses indicates that the chosen training configuration—namely input resolution of 800 × 800 pixels, cosine learning rate scheduling, minority-class augmentation, and mosaic augmentation—contributed positively to the ability of the model to learn discriminative visual features of multi-class acne lesions effectively.

3.2 Overall detection performance

The overall detection performance of the YOLOv11m model during training is summarized in Fig. 8, which presents the evolution of precision, recall, and mAP metrics across epochs.

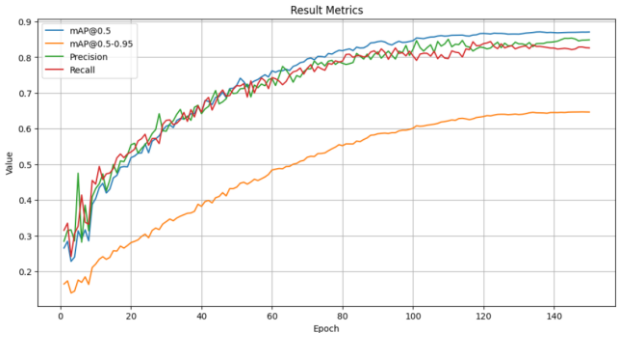

Figure. 8 Result matrics

Table 1. Per-class performance

Class	*Precision*	*Recall*	*mAP50*	*mAP50-95*
All	0.85	0.829	0.871	0.649
Cysts	0.995	0.995	0.995	0.969
blackheads	0.806	0.809	0.86	0.541
nodules	0.79	0.806	0.864	0.677
papules	0.833	0.862	0.893	0.638
pustules	0.853	0.802	0.859	0.607
whiteheads	0.823	0.701	0.755	0.465

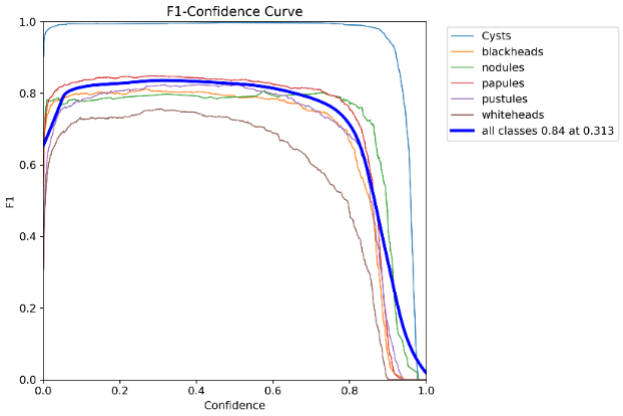

Figure. 9 F-1 confidence score

The model demonstrated steady improvements, converging toward stable values by the final epochs. At convergence, the model achieved an average precision of 0.85, recall of 0.829, mAP@50 of 0.871, and mAP@50–95 of 0.649. The relatively high mAP@50 (0.871) demonstrates the model's ability to generate bounding boxes with sufficient localization accuracy at the IoU threshold of 0.5.

A high precision value indicates that the majority of predictions corresponded to true acne lesions, with only a small proportion classified as false positives. Meanwhile, a recall of 0.829 reflects the model's ability to detect most acne lesions within the dataset, although some false negatives persisted, particularly in cases of small or visually ambiguous lesions.

Further insights are provided in Fig. 9, which illustrates the F1–confidence curve. This visualization highlights the trade-off between precision and recall across varying confidence thresholds. The model achieved its optimal mean F1-score of 0.84 at a confidence threshold of 0.313, representing the best balance between minimizing false positives and maximizing true detections across all classes

In terms of computational efficiency, the model achieved an inference speed of 26 FPS at 800 × 800 resolution on a CUDA T4 GPU. While this speed falls slightly below the common benchmark for real-time processing (≥30 FPS), it can still be categorized as near real-time, making the model practically feasible for semi-interactive clinical applications or offline batch analysis scenarios.

3.3 Per-class performance and confusion matrix

To further examine the detection capability of the model, evaluation metrics were computed for each acne class individually. Table 1 summarizes the per-class precision, recall, mAP@50, and mAP@50–95 scores.

The Cysts class achieved the highest performance, with nearly perfect precision (0.995), recall (0.995), and mAP@50 (0.995). This superior outcome can be attributed to the distinct morphological features of cystic lesions and the effectiveness of targeted data augmentation, which allowed the model to distinguish cysts with minimal confusion.

For blackheads and nodules, the performance was moderate but consistent. Blackheads obtained precision of 0.806 and recall of 0.809, with mAP@50 of 0.860, but experienced a significant drop at stricter localization (mAP@50–95 = 0.541). This suggests difficulty in producing highly precise bounding boxes, likely due to the subtle texture and small size of blackheads. Nodules, in contrast, performed slightly better at higher IoU thresholds (mAP@50–95 = 0.677) despite being underrepresented in the dataset, indicating that their larger and more distinctive appearance supported reliable classification.

Papules and pustules both demonstrated strong performance, achieving mAP@50 of 0.893 and 0.859, respectively. Although visually similar, particularly in inflamed skin regions, the model was still able to distinguish them with reasonable consistency. Papules achieved the highest recall among major classes (0.862), whereas pustules showed competitive precision (0.853), indicating a relatively low rate of false positives.

On the other hand, whiteheads proved to be the most challenging class. Despite being the second most abundant lesion type in terms of bounding box annotations, whiteheads yielded only recall of 0.701

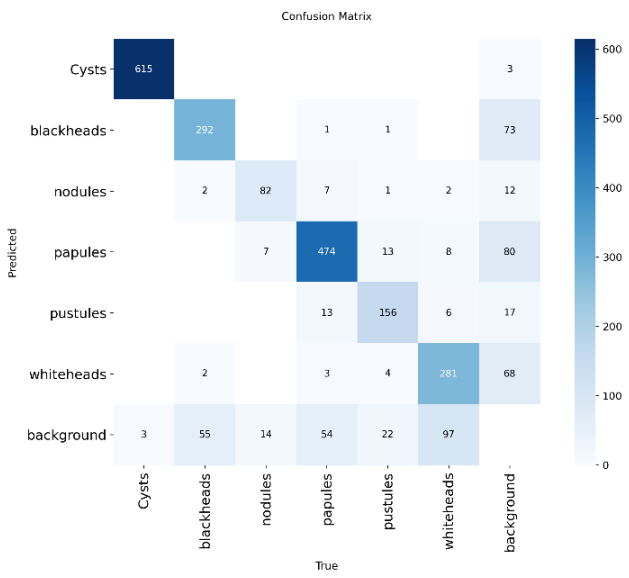

Figure. 10 Confusion matrix

Figure. 11 In-sample visualization

and mAP@50–95 of 0.465. Their small size and resemblance to normal skin textures made them difficult to detect, often resulting in missed detections or misclassification into other small-lesion categories.

The confusion matrix in Fig. 10 provides further insight into class-specific errors. Consistent with the quantitative metrics, Cysts exhibited near-perfect predictions, with 615 out of 618 samples correctly identified and only three misclassified as background. For blackheads, misclassifications occurred mainly as background (55 cases) or whiteheads (19 cases), reflecting their subtle visual similarity. Nodules were mostly detected correctly (82 out of 103), with some confusion with papules and background.

The most frequent cross-class errors occurred between papules and pustules, where inflammatory lesions were sometimes misclassified into each other's categories. Similarly, whiteheads showed widespread errors: only 281 out of 394 correctly detected, with a large proportion misclassified as background (97 cases) and some confused with papules or pustules. Notably, background regions were occasionally mistaken for acne lesions (e.g., 80 false positives as pustules, 73 as blackheads, and 68 as whiteheads), highlighting the challenge of suppressing spurious detections in visually complex skin textures.

Taken together, these findings indicate that while the model performs competitively across most classes, it remains particularly sensitive to lesion size and visual subtlety. Large, visually distinctive lesions (e.g., cysts, nodules) are recognized with high accuracy, whereas small and less distinct lesions (e.g., whiteheads, blackheads) remain the primary source of detection errors.

3.4 Bounding box visualization

To qualitatively assess the detection capability of the proposed YOLOv11m model, bounding box visualizations were generated for both in-sample (validation set) and out-sample (external dataset) images. These visualizations provide complementary evidence to the quantitative metrics, illustrating how the model responds to real facial acne distributions.

Fig. 11 presents a grid-based collage (4 × 4 layout) of validation images with predicted bounding boxes overlaid. Each panel is referenced according to its grid position (R = row, C = column). The results demonstrate that the model successfully detected all six acne classes, with confidence scores consistent with the quantitative evaluation.

In the first row (R1C1–R1C4), dense clusters of papules and pustules are observed, with overlapping bounding boxes corresponding to inflamed lesion aggregations rather than redundant detections. In the second row, several well-localized nodules are identified (confidence ≥ 0.9), while whiteheads and blackheads frequently overlap in adjacent facial regions (R2C3–R2C4). The third row (R3C1–R3C4) illustrates mixed detections, where multiple lesion types (whiteheads, blackheads, pustules) co-occur, particularly in R3C1. In the fourth row, nodules are consistently detected in R4C2–R4C3, and a cyst is confidently identified in R4C4 (confidence = 1.0), despite the rarity of cysts in the images.

To further examine generalization, the model was evaluated on external images from publicly available dermatology datasets (DermNet, ACNE04). Fig. 12 illustrates five representative cases, each showing the original image (top row) and the detection output (bottom row).

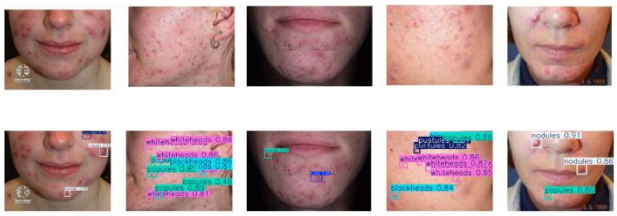

Figure. 12 Out-sample visualization

In the first case, the model correctly detected nodules (confidence 0.86–0.89) and pustules (0.79), although some lesions were missed. The second image featured numerous small lesions on the cheek, where the model detected whiteheads (0.81–0.86), papules (0.46–0.89), and blackheads (0.81–0.86). Overlapping bounding boxes were common in this case due to the high density of lesions.

The third case showed a correctly identified cyst (0.85) alongside papules with lower confidence (0.42). Some lesions were missed, likely due to watermark artifacts and reduced image quality. The fourth case depicted a dense acne distribution on the forehead, where whiteheads, papules, pustules, and blackheads were detected with confidence ≥ 0.82. The clustering of bounding boxes aligned with the actual lesion density, confirming that the model's multiple detections in tight regions often represent true lesion clusters. In the fifth case, the model produced stable detections with high confidence for nodules (0.86–0.91) and papules (0.86), consistent with the strong performance of medium-to-large lesion classes.

Collectively, these visualizations highlight two key findings. First, the model demonstrates strong generalization for large and distinctive acne lesions across unseen datasets. Second, while performance for small lesions remains limited, the detection outputs still align with clinical expectations, particularly in dense lesion areas where over-detection may in fact reflect real clustering rather than noise.

3.5 Comparison with previous study

To contextualize the performance of the proposed model, the results of this study were compared with several prior works that applied YOLO-based architectures for acne or skin lesion detection. A summary is presented in Table 2.

The proposed YOLOv11m model trained on a custom dataset of 1,884 raw images (expanded to 6,116 through augmentation) achieved a mAP@50 of 87.1%, which represents the highest performance among the studies compared. Compared to D. Zhang et al., 2024 [8], who reported 83.7% mAP@50 using an Improved-YOLOv7 trained on 1,457 images from the ACNE04 dataset, this study's YOLOv11m still outperformed despite a dataset of similar scale. This indicates the advantages of leveraging the more recent YOLOv11 architecture together with targeted data augmentation strategies.

Similarly, against Le et al., 2024 [26] and Faruq Aziz & Saputri, 2024 [9],who used significantly larger datasets (≈5,000 and 3,154 images, respectively) with YOLOv8 and YOLOv9, the present model still achieved superior accuracy (87.1% vs. 78% and 81.4%). This demonstrates that carefully designed augmentation strategies can compensate for limited raw dataset size, ensuring competitive performance even compared to models trained on larger corpora.

Earlier works such as Sangha & Rizvi, 2021 [25], and Huey Gan et al., 2024 [7] reported substantially lower accuracy (37.97% and 42.4%, respectively). The performance gap is attributable to both the limited dataset sizes (as few as 403 images) and the reliance on earlier YOLO generations (YOLOv5/YOLOv5s), which are less optimized for multi-class detection of small, visually similar objects.

Table 2. Result comparison with previous study

Author	Dataset	Total Images	Model	Class	mAP50
Huey Gan et al., 2024 [7]	*Custom Dataset*	- 9 classes	YOLOv5s	Multi-class (acne/skin lesion)	42.4%
Sangha & Rizvi, 2021 [25]	ACNE04	403	YOLOv5	Single-class	37.97%
D. Zhang et al., 2024 [8]	ACNE04	1.457	Improved-YOLOv7	Single-class	83.7%
Le et al., 2024 [26]	DermNet-NZ	±5.000 (5 classes)	YOLOv8	Multi-class (acne/skin lesion)	78%
Faruq Aziz & Saputri, 2024 [9]	*Custom Dataset*	3,154 (6 classes)	YOLOv9	Multi-class (acne/skin lesion)	81.4%
Our research	***Custom Dataset***	**1.884 (6 classes)**	**YOLOv11m**	**Multi-class (acne)**	**87.1%**

Overall, the superior performance achieved in this study can be attributed to two main factors. First, the adoption of the YOLOv11m architecture, which integrates improvements in feature extraction and multi-scale detection. Second, the systematic augmentation strategy, which not only increased dataset size threefold but also enhanced minority class representation, particularly for cysts, leading to more stable predictions. Taken together, these results establish the proposed model as a state-of-the-art benchmark for multi-class acne detection, surpassing the accuracy of previous studies.

4. Conclusion and recommendation

This study successfully implemented YOLOv11m for detecting six acne types (blackheads, whiteheads, papules, pustules, nodules, and cysts) using a dataset of 1,884 images expanded to 6,116 through augmentation. The model achieved mAP@50 of 87.1%, mAP@50–95 of 64.9%, precision of 85%, recall of 82.9%, and inference speed of 26 FPS, which is considered near real-time (30 FPS). Per-class evaluation revealed excellent performance on cysts (99%), while smaller lesions such as whiteheads remained more difficult to detect (75%). These results confirm that the quality of augmentation and data representation is more decisive for model accuracy than strictly balancing class distributions.

Future improvements should target enhanced recognition of small lesions through attention mechanisms or multi-scale feature fusion, as well as optimization for deployment on resource-limited devices via lightweight models or knowledge distillation. Clinical validation with dermatologists is also essential to ensure practical applicability. In summary, the integration of targeted augmentation with YOLOv11m delivers competitive performance in multi-class acne detection, highlighting its potential for clinical use and as a foundation for further research on efficient and accurate dermatological AI systems.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, Nayla Nur Fadhillah, Alam Rahmatulloh, and Neng Ika Kurniati; methodology, Nayla Nur Fadhillah; software, Nayla Nur Fadhillah; validation, Nayla Nur Fadhillah, Alam Rahmatulloh, and Neng Ika Kurniati; formal analysis, Nayla Nur Fadhillah; resources, Nayla Nur Fadhillah; data curation, Nayla Nur Fadhillah; writing—original draft preparation, Nayla Nur Fadhillah; writing—review and editing, Alam Rahmatulloh and Neng Ika Kurniati; visualization, Nayla Nur Fadhillah; supervision, Alam Rahmatulloh.

References

[1] Quattrini, C. Boër, T. Leidi, and R. Paydar, "A deep learning-based facial acne classification system," *Clin. Cosmet. Investig. Dermatol.*, vol. 15, pp. 851–857, 2022, doi: 10.2147/CCID.S360450.

[2] D. T. Aryani and W. Riyaningrum, "Hubungan acne vulgaris (AV) dengan kepercayaan diri pada mahasiswa Universitas Muhammadiyah Purwokerto angkatan 2021," *Jurnal Kesehatan Tambusai*, vol. 3, no. 3, 2022.

[3] R. F. Autrilia, D. Retno, and H. Ninin, "Eksplorasi dampak psikologis pada remaja yang memiliki masalah penampilan dengan jerawat," *J. Psikologi Udayana*, vol. 9, no. 2, pp. 194–205, 2022, doi: 10.24843/JPU/2022.v09.i02.p09.

[4] R. L. Hasanah and M. Hasan, "Deteksi lesi acne vulgaris pada citra jerawat wajah menggunakan metode K-means clustering," *Indonesian J. Softw. Eng.*, vol. 8, no. 1, pp. 46–51, 2022. [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ijse46

[5] G. V. Agustin, M. Ayub, and S. L. Liliawati, "Deteksi dan klasifikasi tingkat keparahan jerawat: perbandingan metode you only look once," *J. Teknik Informatika dan Sistem Informasi*, vol. 10, 2024, doi: 10.28932/jutisi.v10i3.9414.

[6] H. Zhang and T. Ma, "Acne detection by ensemble neural networks," *Sensors*, vol. 22, no. 18, Sep. 2022, doi: 10.3390/s22186828.

[7] Y. H. Gan, S. Y. Ooi, Y. H. Pang, Y. H. Tay, and Q. F. Yeo, "Facial skin analysis in Malaysians using YOLOv5: A deep learning perspective," *J. Informatics Web Eng.*, vol. 3, no. 2, 2024, doi: 10.33093/jiwe.2024.3.2.1.

[8] D. Zhang, C. Jin, Z. Zhang, X. Cao, and C. Xue, "Automatic acne detection model based on improved YOLOv7," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3520641.

[9] F. Aziz and D. U. E. Saputri, "Efficient skin lesion detection using YOLOv9 network," *J. Med. Informatics Technol.*, pp. 11–15, Mar. 2024, doi: 10.37034/medinftech.v2i1.30.

[10] A. Sharma, V. Kumar, and L. Longchamps, "Comparative performance of YOLOv8,

YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species," *Smart Agric. Technol.*, vol. 9, p. 100648, Dec. 2024, doi: 10.1016/j.atech.2024.100648.

[11] R. Khanam and M. Hussain, "YOLOv11: An overview of the key architectural enhancements," Oct. 2024. [Online]. Available: http://arxiv.org/abs/2410.17725

[12] AcneDet, "ACNEDet v1 dataset," Roboflow Universe, accessed Apr. 10, 2025. [Online]. Available: https://universe.roboflow.com/acnedet/acnedet-v1

[13] Skindetect, "DetectDataSkin-2 dataset," Roboflow Universe, accessed Apr. 10, 2025. [Online]. Available: https://universe.roboflow.com/skindetect-xueeo/detectdataskin-2

[14] Rungroj, "Acne dataset," Roboflow Universe, accessed Apr. 10, 2025. [Online]. Available: https://universe.roboflow.com/rungroj/acne-xnbio

[15] Nayla, "Multi-class acne dataset," Roboflow Universe, accessed Dec. 23, 2025. [Online]. Available: https://universe.roboflow.com/trial-sth5h/multi-class-acne-yiwfq

[16] M. L. Ali and Z. Zhang, "The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection," *Computers*, vol. 13, no. 12, Dec. 2024, doi: 10.3390/computers13120336.

[17] Y.-H. Lee and H.-J. Kim, "Comparative analysis of YOLO series (from V1 to V11) and their application in computer vision," 2024.

[18] L. P. Kothala and S. R. Guntur, "GEL-TTA Net: A global ensemble learning network for localization of small-scale and mixed intracranial hemorrhages through test time augmentations," *Multimed. Tools Appl.*, vol. 84, no. 14, pp. 13005–13036, Jun. 2024, doi: 10.1007/s11042-024-19393-4.

[19] S. C. Shrawne et al., "Multiclass fruit detection using improved YOLOv3 algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 9, 2024, doi: 10.14569/IJACSA.2024.01509100.

[20] M. Kim, J. Jeong, and S. Kim, "Ecap-yolo: Efficient channel attention pyramid yolo for small object detection in aerial image," *Remote Sens.*, vol. 13, no. 23, Dec. 2021, doi: 10.3390/rs13234851.

[21] J. Wang, J. Yu, and Z. He, "DECA: A novel multi-scale efficient channel attention module for object detection in real-life fire images," *Appl. Intell.*, vol. 52, no. 2, pp. 1362–1375, Jan. 2022, doi: 10.1007/s10489-021-02496-y.

[22] A. D. Khairkar et al., "Predictive YOLO V7 model of dental implant for radiographic images," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 18s, pp. 656–661, 2024.

[23] Ahmed et al., "Enhancing wrist fracture detection with YOLO: Analysis of state-of-the-art single-stage detection models," *Biomed. Signal Process. Control*, vol. 93, Jul. 2024, doi: 10.1016/j.bspc.2024.106144.

[24] M. Bakirci et al., "Multi-class vehicle detection and classification with YOLO11 on UAV-captured aerial imagery," in *Proc. 2024 IEEE 7th Int. Conf. Actual Problems of Unmanned Aerial Vehicles Development (APUAVD)*, Oct. 2024, pp. 191–196, doi: 10.1109/APUAVD64488.2024.10765862.

[25] A. Sangha and M. Rizvi, "Detection of acne by deep learning object detection," Dec. 11, 2021, doi: 10.1101/2021.12.05.21267310.

[26] H. H. Le et al., "Initial approach to the identification degree of skin damage and classification of acne by YOLOv8," in *Proc. 2024 9th Int. Conf. Intelligent Information Technology*, New York, NY, USA: ACM, Feb. 2024, pp. 19–25, doi: 10.1145/3654522.3654526.