Science & Technology Development Journal 2024, 27():1-12

Open Access Full Text Article

Research article

ACNE8M - An acnes detection and differential diagnosis system using AI technologies

Phuc Khang Nguyen[1], Tan Duy Le[2,*], Bao Anh Nguyen[3], Phuong Anh Nguyen[4]

Use your smartphone to scan this QR code and download this article

ABSTRACT

Acne is a prevalent skin condition that can lead to serious consequences in severe cases. Traditional treatment requires patients to visit a dermatologist. However, acne diagnosis performed by dermatologists often encounters issues, such as being manual and highly inaccurate. Therefore, there is a need for machinery to assist in the acne diagnosis phase. Numerous image analysis algorithms have been developed using images captured by mobile devices. Nonetheless, most of these algorithms primarily rely on outdated features such as color models or texture-based features, which may result in poor performance when dealing with the intricate nature of acne lesions. Consequently, AI models have been developed for the task of acne detection. However, due to the rarity of high-quality datasets for acne, some of these models have yet to achieve significant results. To overcome these limitations, this paper proposes the ACNE8M, an AI model developed based on the YOLOv8 pre-trained model, to accurately detect seven primary and secondary types of acne lesions, as well as differentiate five additional diagnoses. The model is trained on a well-prepared dataset containing 9,440 images with numerous acne lesions adequately labeled. The results show that the model achieves state-of-the-art performance with a mean Average Precision (mAP) score of 0.69 across the 12 types. The accuracy of detecting each type of acne is impressively high and balanced between the classes, despite the dataset's imbalance caused by the unequal number of images in each acne category. With this study, ACNE8M is expected to provide medical support in the acne diagnosis process and help patients understand their conditions for better treatment.

Key words: ACNE8M, acne detection, acne AI

[1]*School of Computer Science andEngineering, International University, Viet Nam National University Ho Chi Minh City, Ho Chi Minh City, Viet Nam*

[2]*Ho Chi Minh City International University, Viet Nam National University Ho Chi Minh City, Ho Chi Minh City, Viet Nam*

[3]*Gia Dinh People's Hospital, Ho Chi Minh City, Viet Nam*

[4]*Pham Ngoc Thach University of Medicine, Ho Chi Minh City, Viet Nam*

Correspondence

Tan Duy Le, Ho Chi Minh City International University, Viet Nam National University Ho Chi Minh City, Ho Chi Minh City, Viet Nam

Email: ldtan@hcmiu.edu.vn

History

- Received: 2024-04-02
- Accepted: 2024-07-09
- Published Online:

DOI :

Check for updates

INTRODUCTION

Acne vulgaris, commonly known as acne, is a widespread skin condition that results from damage to the sebaceous glands or when the process of inflammation clogs hair follicles beneath the skin. The most commonly affected areas are the face, shoulders, and back. In the absence of skin disorders, sebaceous glands produce sebum, which is discharged onto the skin surface through pores and openings in the follicles. Normally, as the body undergoes the natural process of shedding skin cells, specifically keratinocytes, these cells ascend to the skin's outer layer. When an area of the body is afflicted with acne, hair, sebum, and keratinocytes clump together inside the pore, preventing sebum from reaching the skin's surface. This blockage allows a mixture of oils from the sebaceous glands and skin cells to foster the growth of bacteria in the obstructed hair follicles, leading to inflammation characterized by swelling, redness, heat, and pain. The increased pressure within the blocked follicles eventually causes them to break down, releasing bacteria, skin cells, and sebum into the surrounding skin and forming lesions. Acne can lead to various types of lesions, but there are five primary types: comedones, papules, pustules, nodules, and cysts[1]. Within the category of comedones, there are two subtypes, including whiteheads and blackheads, making a total of six major categories. Figure 1 describes each of these six main types in detail, providing illustrations and real images.

Because periodic diagnoses are necessary for many patients, which requires a significant number of consultations that can be challenging due to the limited number of dermatologists, there is a significant need for assistance in the acne diagnosis process. Therefore, several image analysis algorithms have been developed. Despite extensive research in medical object detection, acne detection has received little attention, despite the disadvantages and consequences that acne patients can potentially suffer. Conventional image processing methods, including traditional handcrafted features such as color models or texture-based ones, have certain limitations. Given the complexity of skin lesions, these methods most significantly lack in detection performance and generalization capability. The Convolutional Neural Network (CNN) is the most well-known and commonly used among

Cite this article : Nguyen P K, Le T D, Nguyen B A, Nguyen P A. **ACNE8M - An acnes detection and differential diagnosis system using AI technologies**. *Sci. Tech. Dev. J.* 2024; 27():1-12.

Copyright

© VNUHCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

the proposed image analysis algorithms. Its performance in detecting biomedical objects, such as nuclei or fovea detection from fundus images, is proven.

In this study, an AI model called ACNE8M for acne detection will be developed by applying the pre-trained YOLOv8 model and fine-tuning it on a specialized acne dataset. ACNE8M is designed to accurately identify the five main types of acne lesions, as depicted in Figure 1, including whiteheads, blackheads, papules, pustules, and cysts. Beyond these primary lesions, ACNE8M can also recognize secondary lesions such as keloid and atrophic scars. Furthermore, it can distinguish acne lesions from similar conditions, including milium, flat wart, folliculitis, acne conglobata, and syringoma. This study briefly explains the steps to train the AI model and the technologies underpinning it. ACNE8M aims to achieve effective and balanced performance across the spectrum of 12 acne types and various acne classes, with a strong emphasis on achieving high accuracy, precision, and recall. Ultimately, this paper will offer a pragmatic solution designed to provide valuable support to dermatologists and individuals afflicted by acne, and contribute to advancements in acne treatment strategies.

BACKGROUND AND RELATED RESEARCH

Background

Numer atedious image analysis algorithms have been developed in the field of acne diagnosis. However, they primarily depend on traditional, handcrafted features, such as color models or texture-based approaches. In contemporary settings, the RGB (Red-Green-Blue) and HSV (Hue Saturation Values) color models emerge as the most popular choices. These two color models were implemented in an acne detection algorithm proposed by Kittigul[3]. The standard approach for color-model-based algorithms involves leveraging the values of each component within the color model to identify acne objects. Specifically, algorithms based on RGB rely on the values of the R, G, and B color components for detection. Conversely, HSV utilizes the H, S, and V values. A common weakness in these methods is the variability in component values within the color model. This instability means that minor differences in these values can significantly impact the algorithms' performance, possibly causing false predictions or misclassification of acne objects to the point of them being undetectable. Consequently, texture-based algorithms, building on the features of color-model-based methods, emerged as an alternative but also exhibited certain drawbacks.

Related Works

A method proposed in 2022 by Faizal Makhrus et al.[4] employed the Gaussian Mixture Model (GMM) to detect acne objects. Despite incorporating texture features alongside color components to enhance the algorithm's analysis of acne information, this method remains susceptible to misdetection. The suboptimal efficiency of the model is evidenced by an accuracy of only 67% when employing Gabor features. This analysis unveils the inherent limitations of early algorithms when confronted with the complexities of skin lesions. Consequently, the pivotal areas for improvement in these methodologies relate to addressing poor detection performance and enhancing generalization capabilities. The integration of computer vision concepts, particularly the utilization of Convolutional Neural Networks (CNNs), has markedly advanced skin image analysis. This progress is evident in its success in detecting various biomedical objects, exemplified by achievements in nuclei and fovea detection from fundus images. CNNs, a powerful subset of deep learning algorithms, have revolutionized visual data analysis by emulating the hierarchical structure of the human visual cortex. As a form of a deep learning algorithm, it is adept at autonomously and adaptively learning spatial hierarchies of features from varying levels, ranging from low to high patterns. In addition, Abas et al.[5] developed an approach using entropy-based filtering and thresholding to identify the region of interest, subsequently utilizing discrete wavelet frames to extract acne features. This methodology demonstrated the ability to classify six distinct types of acne lesions and scars, achieving a classification accuracy of 85.5%. However, this result only reveals moderate efficacy, potentially owing to the manual aspect of the feature extraction phase, which could lead to inaccuracies.

Despite these advancements, there is a noticeable lack of utilization of more sophisticated computer techniques, like deep learning, in the methods mentioned for analyzing images. This gap represents a missed opportunity, especially since these contemporary computer methods have proven their strength in identifying crucial details more efficiently than traditional, labor-intensive methods. This oversight misses a chance to further improve accuracy and effectiveness in acne diagnosis. A groundbreaking study by Chuan-Yu Chang and Heng-Yi Liao[6] attempted to bridge this gap by employing a special kind of computer model (SVM classifier) to differentiate between spots, acne, and normal skin. Their approach achieved a remarkable accuracy of 99.4% in distinguishing spots from

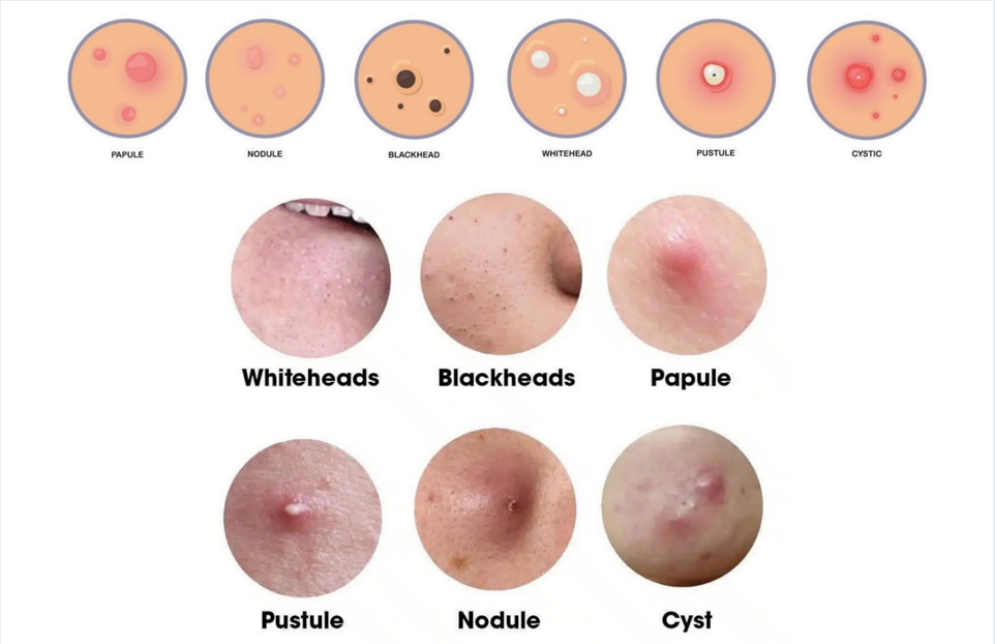

Figure 1: Types of acne lesions[2]

acne. However, the sensitivity rate, at 80.91%, indicates there is room for improvement, especially in reducing the likelihood of false detections.

METHODOLOGY

A. Dataset

The dataset[7] utilized for this model is sourced from Roboflow and authenticated by our dermatologists. Roboflow is a free and open-source platform containing over 200,000 image datasets across various fields of study. It also provides a suite of tools for dataset customization, including splitting with appropriate ratios for training, testing, and validation; applying pre-processing and augmentation; and labeling images within the dataset. After exploring numerous datasets on Roboflow, six that met the criteria of this AI model were selected, resulting in a combined total of 9,440 pre-processed, augmented, and correctly labeled images.

Figure 2 illustrates the dataset preparation process. All images are resized to 800 x 800 pixels, with several augmentations applied: mosaic, horizontal and vertical flips, and rotation augmentation within a range of -25 to 25 degrees. Roboflow provides six options for resize augmentations: stretch, fill, fit within, fit with reflected edges, fit with white edges, and fit with black edges. For this dataset, the stretch resizing technique is chosen to ensure a proportional adjustment that maintains the integrity of the original image data. Mosaic augmentation involves creating a composite training sample from multiple images; in this dataset, it results in a training image created from four individual pre-processed images. The acne detection dataset is divided in a 90:10 ratio for training and testing. Specifically, 8,700 images in this dataset are designated for training, while the remaining 740 images are further divided into 580 for validation and 160 for inference. This method produces a dataset containing images similar to those depicted in Figure 3 below. The images are then labeled in the YOLOv8 formatting style, utilizing specialized tools provided by Roboflow.

Following these processing steps, the images will resemble the example shown in Figure 3 - a composite image comprised of four individual images. It is important to highlight that among these images, there may be slight variations in angles and flips due to the implementation of horizontal and vertical flipping, along with rotation augmentation, ranging from -25 to 25 degrees.

B. Methods

The YOLO (You Only Look Once) framework is renowned for its real-time object detection algorithms, offering high speed, accuracy, and state-of-the-art performance. Within the YOLO framework, the models have been developed by multiple authors. One of the contributors to the YOLO

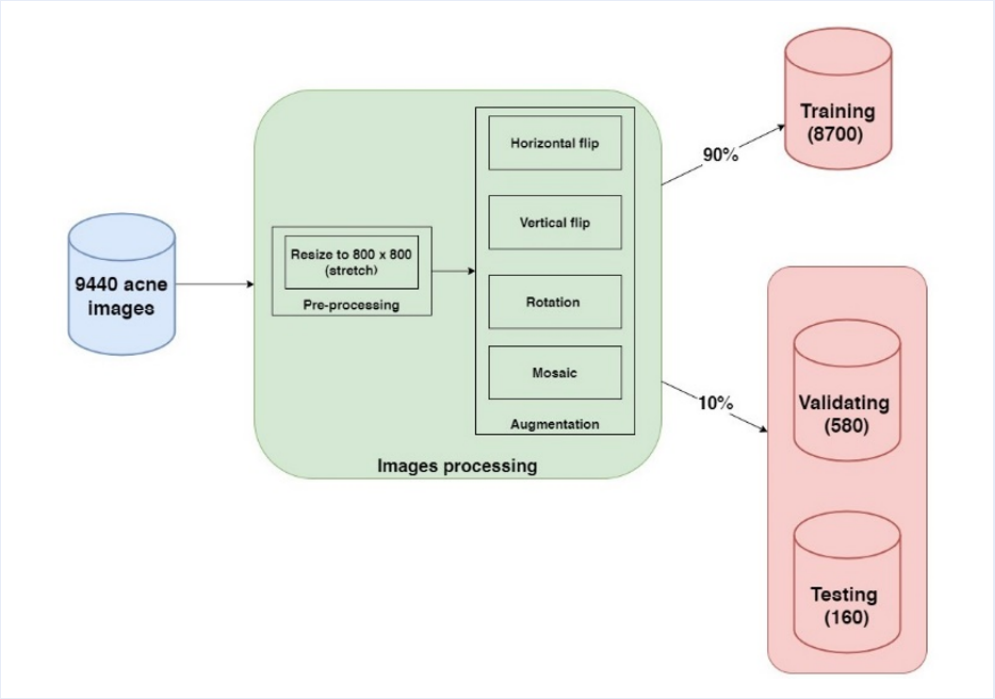

Figure 2: Dataset preparation process

Figure 3: Sample images in the dataset

models' development is Ultralytics, which has developed three versions in the YOLO model family; these are YOLOv3, YOLOv5, and YOLOv8, with YOLOv8 being the latest version. YOLOv8 is built on the PyTorch framework and features an adjusted backbone called YOLOv8CSPDarknet, adopted from the YOLOv5 model. Compared to its most recent predecessor, YOLOv7, YOLOv8 has shown better performance in tomato detection, achieving an accuracy of 93.4%. For drone detection, YOLOv8 also surpassed YOLOv7 with an accuracy of 50.16% compared to 48.16% [8], and in pothole detection, YOLOv8 outperformed YOLOv7 with an accuracy of 78.6% in terms of mAP [9]. Besides object detection, YOLOv8 can handle various computer vision tasks, including object classification, segmentation, and tracking.

Given the promising potential of YOLOv8, this model was selected for acne detection training. The YOLOv8 model offers five architectural versions: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. The architecture of YOLOv8m, part of the YOLOv8 series, comprises 218 layers with over 25M pre-trained pa-

rameters and has achieved an accuracy of 53.9% on the COCO 2017 dataset. The COCO (Common Objects in Context) dataset is notable in the field of object detection, containing 91 object categories with a total of 2.5M labeled object instances across 328K images. This performance and dataset scale have led to the selection of the YOLOv8m architecture for this study.

The ACNE8M training is conducted on a Google Colab Tesla T4 GPU, with configurations and steps detailed in Figure 4.

In this training, some important hyperparameters are finetuned to fulfill the requirements of the acne detection task. From Figure 4, some custom training configurations for the ACNE8m model include:

- **task = detect**: command argument defines the specific task the model should perform, which is detection.
- **mode = train**: Training mode
- **model = yolov8m.pt**: Argument specifies pretrained YOLOv8 model of choice. This training uses the YOLOv8m architecture, so the pretrained model YOLOv8m is chosen.
- **batch = 16**: Number of images per batch
- **imgsz = 800:** Size of input images
- **save = True**: Save train checkpoints and predict results
- **save period = 10**: Save train checkpoints every ten epochs in case of corruption.
- **pretrained = True**: Load weights from a pretrained model. Because this model is finetuned based on the pre-trained YOLOv8m, this option should be True.
- **optimizer = auto**: Optimizer to be used. The optimizer helps dynamically finetune the model throughout the training process, aiming to minimize a predefined loss function. Available optimizers: SGD, Adam, Adamax, AdamW, NAdam, RAdam, RMSProp. SGD (Stochastic Gradient Decent) was chosen for this training.
- **momentum = 0.937**: SGD momentum. Because SGD requires a large number of iterations (training epochs) to reach the optimal minimum, the computation time is significantly slow. As a result, momentum is implemented to facilitate the convergence of the loss function.
- **lr0 = 0.01, lrf = 0.01**: Initial and final learning rate. The learning rate is a hyperparameter that dictates the speed at which an algorithm adjusts or learns the parameter estimate. In other words, the learning rate governs the adjustments of neural network weights in response to the loss gradient. Achieving model accuracy requires a careful balance between the learning rate and momentum. A higher momentum corresponds to a lower learning rate. In this case, a learning rate of 0.01 was selected to balance loss convergence and training time.

If a configuration is not specified in the list, it will be set to default values as defined by Ultralytics. The training settings are implemented as described above, after considering best practices, available computational resources, and the balance between training efficiency and model performance. First, best practices in the field of machine learning provide foundational guidelines. For instance, a batch size of 16 is often recommended because it offers a good balance between the stability of the gradient descent process and computational efficiency. Smaller batch sizes can lead to noisy gradients, while larger batch sizes require more memory and can slow down the training process. The learning rate of 0.01 was chosen based on empirical evidence and extensive experimentation. A learning rate that is too high can cause the model to converge too quickly to a suboptimal solution, while a learning rate that is too low can make the training process unnecessarily slow. A learning rate of 0.01 is widely recognized as a good starting point, providing a balanced approach to achieve both reasonable convergence speed and model accuracy.

Furthermore, these specific values were fine-tuned, considering the computational resources at our disposal. The training was conducted on a Google Colab T4 GPU, as mentioned earlier, which provided a set amount of GPU memory, processing power, and time constraints. These factors were critical in determining the batch size and learning rate to ensure that the model could be trained efficiently within the given resource limitations without compromising performance.

Figure 5 supports Figure 3 by providing a comprehensive overview of the system's workflow, detailing each step required for its operation. Initially, acne images are sourced from the Roboflow dataset platform, as referenced earlier. These images undergo thorough examination by dermatologists to ensure their relevance and accuracy. The initial phase of image preparation involves preprocessing, as outlined in Section 3A. During preprocessing, any images with over 50% inaccuracies in their labels are identified and excluded from the dataset, a crucial step known as dataset cleanup. Following the preprocessing, the YOLOv8 pre-trained weight is adopted and fine-tuned specifically for this dataset. Additionally, a YAML file is implemented to define the data pipeline configurations

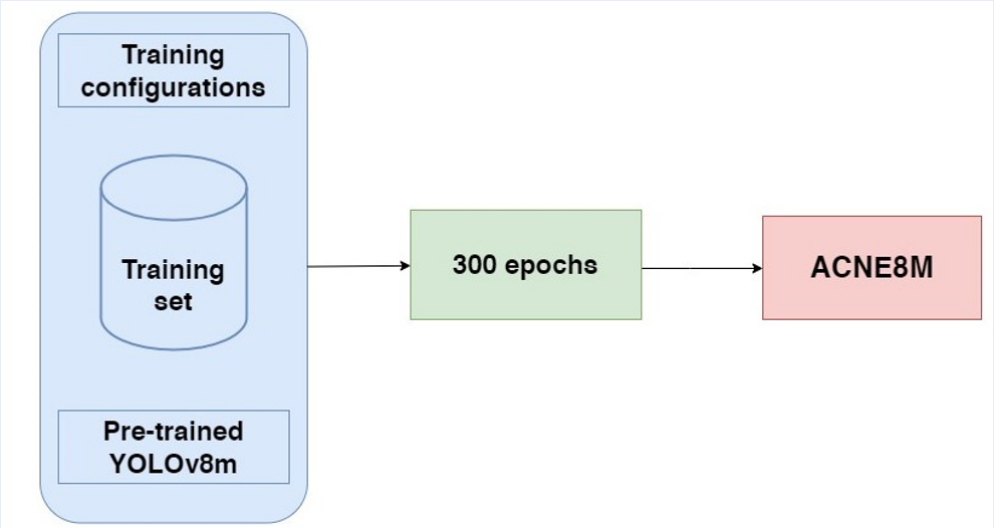

Figure 4: Training configurations and steps

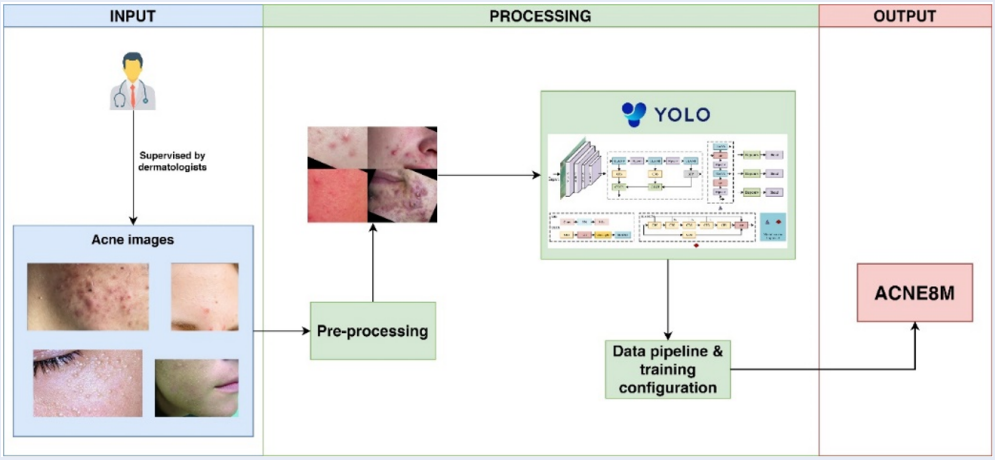

Figure 5: ACNE8M system design

and specify the new classes. This configuration, along with the prepared training weights and the training settings presented, was utilized to train the ACNE8M model. The training phase extends over 300 epochs, culminating in the readiness of ACNE8M to accurately identify and categorize the twelve distinct acne lesions and related differential diagnoses highlighted in our study. Post-training, dermatologists test the model on both validation and test sets in the datasets prepared, ensuring its predictions align with professional diagnostic standards.

RESULTS

A. Training results

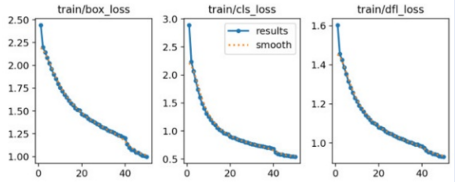

Figure 6: Curves of training losses

From Figure 6, it can be seen that the curve shows a stable decrease trend with minor fluctuation. Based

on the knowledge explained about the behavior of the learning curve, the shape of these curves indicates that there can be minor overfitting in the model but not significant. Hence, the validation results are more reliable. The figure consists of 3 curves indicating the box loss, cls loss, and dfl loss, respectively.

Box loss: The box loss quantifies the discrepancies in the coordinates of the bounding box, indicating the predictions made by the model compared to the ground truth coordinates of the bounding boxes that encapsulate the target object.

Cls loss (class loss): The class loss assesses the disparity in classifying the object classes associated with each bounding box. In other words, it represents the distinction between the predicted object class and the actual class of the object as per the ground truth.

DFL loss (Distribution Focal Loss): DFL loss serves as a metric designed to tackle challenges related to class imbalance. It amplifies the influence of challenging samples by diminishing the weight assigned to easier samples. In doing so, it effectively mitigates the class imbalance issue.

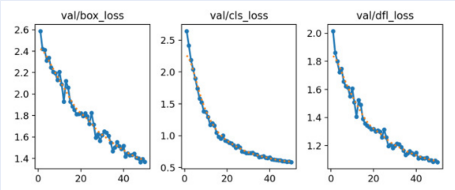

Figure 7: Curves of validation losses

B. Validation results

With box loss, cls loss, and dfl loss explained, the validation curve is evaluated in a similar way to the training curve. Compared to the training curve, all three validation curves exhibit a similar overall pattern, illustrating a downward trend. This suggests that the model effectively generalizes with the dataset. Nonetheless, slight fluctuations in each of the three validation curves hint at the potential for mild overfitting. It's worth noting that these fluctuations are mild and are unlikely to impact the overall performance of the model substantially.

Using the YOLO-standardized evaluation metrics, the performance of ACNE8M can be assessed correctly. In addition to the metrics, additional scoring methods such as normalized confusion matrix, F1 score, or PR scores will also be introduced to clarify the performance of ACNE8M better.

Figure 8 shows that the performance of the model is very decent, with very high precision (above 97%) for each acne-type object when validating with the validation set. In this figure, the labels "acne_scars," "sebo-crystan-conglo," "papular," and "purulent" correspond to atrophic scars, acne conglobata, papules, and pustules, respectively. With an IoU threshold of 0.5 (easy detection), the model produces outstanding results with an average precision of 0.984 among the 12 categories of acne lesions and differentials diagnoses introduced. Across multiple levels of difficulty detecting acne objects, the mAP50-95 scoring of ACNE8M shows consistent performance. This result can be observed through the mAP score of the 12 object categories (0.69), the best score (0.727) in the cystic type, and the lowest score (0.651) in the blackhead type. Recall that precision serves as a metric indicating the reliability of predictions. In other words, it evaluates the probability of a prediction being genuinely accurate.

Apart from precision and the mAP score, the performance of ACNE8M is also assessed based on its recall ability. Recall, referred to as sensitivity, pertains to the model's capability to identify all positive samples, representing the miss rate. A higher recall score implies that an AI model is less likely to overlook a positive sample. In the case of ACNE8M, evaluating all 12 classes of acne, the recall scoring is high (above 0.9), with the best one being the folliculitis type and the lowest being the papular type, with the scoring of 0.99 and 0.94, respectively. This indicates a robust ability to capture positive samples across various acne classes, underscoring the effectiveness of ACNE8M in recognizing diverse instances of acne. Besides the scoring methods mentioned, the performance of this model can be clarified further using the confusion matrices illustrated in Figure 9.

From the confusion matrices illustrated in Figure 9, it can be seen that the predictions made by the model are highly reliable, even though there are still rare occurrences of false predictions within the 12 categories of acne. This can be verified from the normalized confusion matrix, which shows that the accuracy of predictions in each of the 12 types is above 0.95. In other words, most of the predictions the model makes have a 95% chance of correctly detecting and classifying. Based on the non-normalized confusion matrix, the accuracy of this model can be calculated using the formula [10]:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

Where: TP = True Positive, TN = True Negative

Class	Images	Instances	Box(P	R	mAP50	mAP50-95)
all	580	5556	0.972	0.967	0.984	0.69
acne_scars	580	287	0.979	0.951	0.98	0.67
blackhead	580	489	0.981	0.945	0.978	0.651
cystic	580	126	0.961	0.984	0.983	0.727
flat_wart	580	282	0.946	0.989	0.989	0.703
folliculitis	580	103	0.986	0.99	0.995	0.67
keloid	580	178	0.994	0.989	0.993	0.689
milium	580	234	0.968	0.97	0.985	0.721
papular	580	1208	0.986	0.968	0.988	0.704
purulent	580	769	0.963	0.94	0.97	0.675
sebo-crystan-conglo	580	447	0.971	0.958	0.982	0.705
syringoma	580	94	0.948	0.973	0.991	0.695
whitehead	580	1339	0.98	0.942	0.973	0.673

Figure 8: Validation of ACNE8M across 12 types of acne

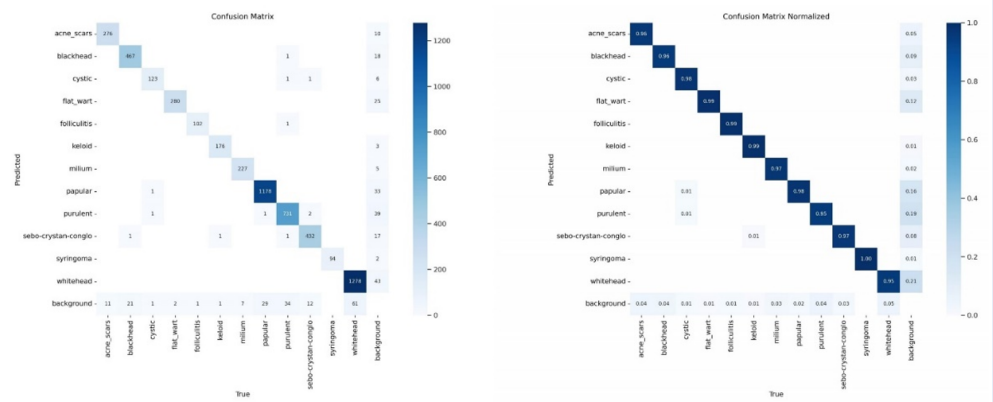

Figure 9: Confusion matrices (non-normalized and normalized)

FN = False Negative, FP = False positive

Using the formula, the accuracy of this model is calculated to be approximately 0.976, which is an impressive number. It is important to emphasize that the confusion matrices exclusively validate the accuracy of predictions generated by ACNE8M. This encompasses not only the correct categorization of objects but also the precise localization, providing insights into the confidence of the model in predicting detected acne objects within the input image.

DISCUSSION

For a result to be deemed correct, the majority of the objects had to be classified accurately. This stringent validation process was applied across the entire validation and test sets. Subsequently, the model was evaluated using key metrics such as accuracy, precision, and recall. These metrics were then compared against other methods on the same dataset to ensure a comprehensive performance assessment. In Figure 10, real-time observations reveal that ACNE8M consistently generates highly confident results, with confidence scores predominantly ranging above 0.7. The noteworthy aspect is the elevated confidence scores and the detection of multiple instances of acne objects. This results in a minimal number of undetected objects, showcasing the model's effectiveness in accurately identifying and categorizing relevant features.

Deep learning algorithms for skin image analysis have been developed and are improving occasionally. AI systems are showing their potential as medical assistants in the diagnosis and treatment processes. A good algorithm produces accurate results, benefiting the patients with the information retrieved for better treatment. In this research, ACNE8M was developed with the aim to detect and classify 12 different types of acne with high accuracy. There have been many studies on acne detection, but since acne lesions can get very complex, the proposed algorithms could have achieved better results. Table 1 summarizes and compares the performance and capabilities of ACNE8M

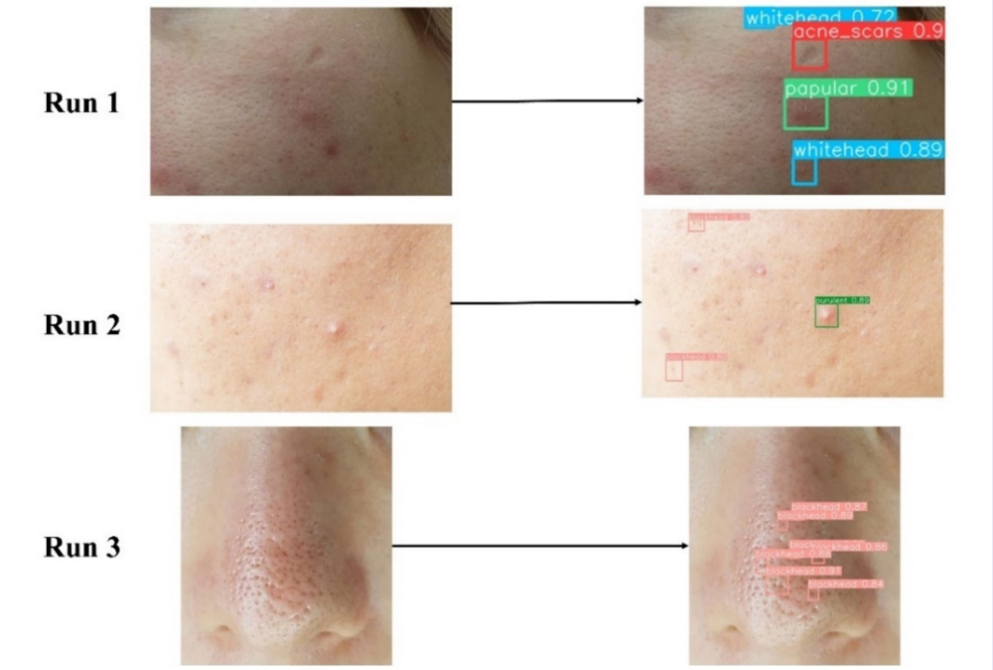

Figure 10: Real-time performance of ACNE8M

and models from some reliable studies.

Table 1 shows that ACNE8M completely outperformed the four earlier methods thanks to being trained on a richer dataset containing more than 9000 images of acne objects and categorized into 12 different types. Among the model architectures, ACNE8M was developed using the YOLO framework, specifically the YOLOv8 pre-trained model, the latest version in the YOLO model family. Although ACNE8M is capable of detecting multiple categories of acne, it does not sacrifice precision. It achieved the highest average precision score of 0.69 compared to the three previous studies, which was measured across varied thresholds of IoU ranging from 0.50 to 0.95. For the recall, ACNE8M achieved a recall score of 0.967, which, compared with the precision, is very close if precision is considered at the IoU threshold of 0.5, which is the standard for a positive prediction to be considered true. Finally, despite the many types this model is capable of, it can also handle the imbalanced classes in the dataset due to the scarcity of high-quality images among the classes of acne. Despite the big difference in the number of photos in each class, the accuracy of each class is pretty close to each other, with only about a 5% difference at most (Figure 8).

Regarding dataset difference, although Kyungseo Min et al.[5] proposed an acne detection model, it was trained on the ACNE04 dataset, which was made for acne severity grading rather than detecting and classifying acne objects. Therefore, it could only detect general acne objects without indicating what the types of the detected acne objects were. The dataset that Kuladech et al.[7] used was processed and prepared in a unique way that was not easy to reach. For Quan Thanh Huynh et al.[13], the dataset is more affluent compared to the other two, and the images in this dataset can be quickly taken using smartphones; however, because 1572 images is a pretty small dataset size to be trained, and the model was trained for 13000 epochs, overfitting was possible, and there were potential signs showing unstable performance. For our method, we have a rich dataset of more than 9000 images that are correctly labeled, along with preprocessing and augmentations. Still, we only need 300 epochs for training, resulting in reliable, stable, and better results than the other 3.

In addition to Table 1, because of dataset differences, the evaluation of ACNE8M can be questionable. As a result, an extensive study was conducted to test some widely used algorithms in the field of object detection using the proposed acne dataset. The results are presented in Table 2.

In Table 2, the evaluation of ACNE8M compared to two other well-known models, Faster-RCNN and

Table 1: Comparison of the performance between ACNE8M and previous studies

Authors	No. of acne types	Total images in the dataset	Model	mAP
Kuladech et al. [11]	4	871	Faster R-CNN, R-FCN	Faster R-CNN: 0.233 R-FCN: 0.283
Kyungseo Min et al. [12]	1	1457	ACNet	0.205
Quan Thanh Huynh et al. [13]	4	1572	Faster R-CNN	0.54
Faizal Makhrus et al. [4]	1	60 (ACNE04)	Gaussian Mixture Model	0.52
Our method (ACNE8M)	**7 + 5**	**9440**	**YOLOv8**	**0.69**

Table 2: Comparison of the performance between ACNE8M and other models on the proposed acne dataset

Criteria	Faster-RCNN	RetinaNet	YOLOv8
No.Params	41.7M	36.5M	**25.5M**
Accuracy	0.769	0.481	**0.976**
mAP50	0.789	0.485	**0.984**
mAP50-95	0.328	0.112	**0.69**
Recall	0.434	0.261	**0.967**
F1	0.374	0.157	**0.805**

RetinaNet, is described. Faster-RCNN is a region-based CNN with a long history of popularity in object detection, known for its accuracy on PASCAL VOC 2007, 2012, and MS COCO datasets. RetinaNet, on the other hand, is a one-stage dense object detection algorithm trained on focal loss, designed to match the speed of one-stage detectors and bypass other two-stage methods. Both Faster-RCNN and RetinaNet are easy to train due to the availability of boilerplate codes and supporting frameworks, which is why they were chosen for testing alongside YOLO algorithms. Despite having fewer parameters than the other two methods, our approach outperforms them across various widely used evaluation metrics in object detection. This demonstrates the efficiency and effectiveness of our method, achieving superior performance without the need for a more complex model.

Despite such an impressive performance, this study encountered some limitations. While not severe or significant, overcoming the limitations can improve the results. Although the images trained for the model are smartphone images, they need to be highly focused on the face or the acne-affected areas rather than random pictures. In addition, as mentioned earlier, the dataset used in this study has a significant inequality within the 12 classes of acne. Therefore, for the categories currently starving of training, validating, and testing images, the result can be better if more images of these classes can be found and adequately labeled to create a balanced number of images among each class for the best results. Despite exhibiting a low level of overfitting, as confirmed by the training curve in Figure 6, the mAP score for ACNE8M stands at 0.69, deemed a fair performance. This mAP score is calculated based on IoU thresholds ranging from 0.5 to 0.95. In an ideal scenario, an object detection model should effectively balance precision and recall, detecting a substantial number of positive samples, predominantly true positives rather than false positives. It is crucial to recall that a prediction is considered a true positive if the predicted bounding boxes overlap by more than half the area of the ground truth bounding boxes; otherwise, it is categorized as a false positive. Notably, such errors do not lead to entirely incorrect predictions, such as misclassification or placing bounding boxes at entirely wrong coordinates; instead, the error is confined to a difference smaller than one between the predicted and ground truth bounding boxes. Examining various IoU thresholds, ACNE8M faces challenges in achieving precise detection. This becomes apparent in the real-time test run, as illustrated in Figure 10, with

the default IoU threshold set at 0.7, showcasing discernible differences between the results. Run 1 detected fewer acne objects, while Run 2 and 3 were able to identify more. This result potentially shows that the number of acne objects detected in run 1 is the smallest compared to the other two. However, most of the predictions in Run 1 are likely to be true positives, while for the other two, there can be a slight error in predicted bounding box coordinates. Importantly, all three predictions are correct, and the primary distinction lies in how closely the predicted bounding boxes align with the expected ground truth. In practice, this may not be severe because the primary goal of the object detection model is to show the object on the result at least and correctly classify the class of the object. However, improving this aspect can boost precision and contribute to an overall enhancement in the performance of the model.

CONCLUSION AND FUTURE WORKS

In this study, an AI model called ACNE8M was developed to detect acne lesions. ACNE8M was implemented based on the YOLOv8 architecture and is capable of recognizing seven specific types of acne lesions. These include five primary lesions: papules, pustules, nodules, cysts, and comedones (categorized into whiteheads and blackheads), as well as two secondary types: atrophic scars and keloids. Additionally, ACNE8M is equipped to assist in the differential diagnosis of acne, distinguishing it from conditions with similar appearances, such as milium, flat warts, folliculitis, acne conglobata, and syringoma, thereby facilitating a more comprehensive understanding and treatment approach to acne nodules, cysts, comedones (whiteheads and blackheads), atrophic scars, and keloid. The scarcity of high-quality acne datasets presented challenges during training, particularly in addressing class imbalances within acne categories and ensuring appropriate image processing to generate sufficient training data. Managing these challenges presented a significant risk of overfitting. However, the problem was effectively mitigated, thereby preserving the capability of ACNE8M to attain state-of-the-art results, achieving a mAP score of 0.69 and an accuracy of 0.976 across the 12 distinct acne classes. With such an outstanding performance, ACNE8M is expected to be a helpful assistant, not only to dermatologists but also to patients. To treat patients better, dermatologists or acne experts can benefit from ACNE8M in acne diagnosis. As for the patients, ACNE8M can help them track their disease status, which can be crucial in post-treatment steps so that the skin can stay healthy in the long term. Because ACNE8M achieved a fair mAP score, indicating possible minor errors in detecting performance, there is a bright future for ACNE8M to be improved. On a larger scale, the size of this dataset - which is 9440 images in total, is considered not large enough. Because of this factor, only finetuning the hyperparameters or adequately preparing the dataset is not enough to enhance the performance of ACNE8M. The best method to improve ACNE8M is to supply more data covering a comprehensive range of acne types. Combining this with appropriate training configurations will potentially enhance the overall performance of ACNE8M. Besides this improvement plan, ACNE8M is projected to be integrated into cross-platform applications for commercial use, especially web, and mobile, specifically developed to aid in acne treatment.

ABBREVIATIONS

xxx

ACKNOWLEDGMENTS

xxx

AUTHOR'S CONTRIBUTIONS

xxx

FUNDING

xxx

AVAILABILITY OF DATA AND MATERIALS

Data and materials used and/or analyzed during the current study are available from the corresponding author on reasonable request.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

1. Acne [Internet]. U.S. Department of Health and Human Services; 2023;Available from: https://www.niams.nih.gov/health-topics/acne.
2. Cystic acne treatment & removal in Singapore: Obvious acne clearance [Internet]. 2024;Available from: https://slclinic.com.sg/acne-treatment-singapore/.

3. Kittigul N, Uyyanonvara B. Automatic acne detection system for medical treatment progress report. 2016 7th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES). 2016 Mar;Available from: https://doi.org/10.1109/ICTEmSys.2016.7467119.
4. Maimanah AN, Wahyono, Makhrus F. Acne classification with Gaussian mixture model based on texture features. International Journal of Advanced Computer Science and Applications. 2022;13(8);Available from: https://doi.org/10.14569/IJACSA.2022.0130844.
5. Abas FS, Kaffenberger B, Bikowski J, Gurcan MN. Acne image analysis: Lesion localization and classification. Medical Imaging 2016: Computer-Aided Diagnosis. 2016 Mar 24;Available from: https://doi.org/10.1117/12.2216444.
6. Chang C-Y, Liao H-Y. Automatic facial spots and acnes detection system. Journal of Cosmetics, Dermatological Sciences and Applications. 2013;03(01):28-35;Available from: https://doi.org/10.4236/jcdsa.2013.31A006.
7. Acnes detection dataset [Internet]. 2022;Available from: https://universe.roboflow.com/testdg2/fold_02.
8. Agarwal K, M S AS, Bakshi S, M V, J J, S D. Performance analysis of YOLOV7 and Yolov8 models for drone detection. 2023 International Conference on Network, Multimedia and Information Technology (NMITCON). 2023 Sept 1;Available from: https://doi.org/10.1109/NMITCON58196.2023.10276343.
9. Haimer Z, Mateur K, Farhan Y, Madi AA. Pothole detection: A performance comparison between Yolov7 and yolov8. 2023 9th International Conference on Optimization and Applications (ICOA). 2023 Oct 5;Available from: https://doi.org/10.1109/ICOA58279.2023.10308849.
10. Accuracy vs. precision vs. recall in machine learning: What's the difference? [Internet];Available from: https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall.
11. Rashataprucksa K, Chuangchaichatchavarn C, Triukose S, Nitinawarat S, Pongprutthipan M, Piromsopa K. Acne detection with deep neural networks. 2020 2nd International Conference on Image Processing and Machine Vision. 2020 Aug 5;Available from: https://doi.org/10.1145/3421558.3421566.
12. Min K, Lee G-H, Lee S-W. ACNet: Mask-aware attention with dynamic context enhancement for robust acne detection. 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2021 Oct 17;Available from: https://doi.org/10.1109/SMC52423.2021.9659243.
13. Huynh QT, Nguyen PH, Le HX, Ngo LT, Trinh N-T, Tran MT-T, et al. Automatic acne object detection and acne severity grading using smartphone images and artificial intelligence. Diagnostics. 2022 Aug 3;12(8):1879;PMID: 36010229. Available from: https://doi.org/10.3390/diagnostics12081879.