INTERNATIONAL JOURNAL OF APPLIED METHODS IN ELECTRONICS AND COMPUTERS

www.ijamec.org

International Open Access

Volume 13
Issue 04
December, 2025

Research Article

Deep Learning-Based Skin Disease Detection: Comparative Performance Analysis of YOLOv8 and YOLOv11 Models

Bilal İNAN [a,*], ***İlker Ali ÖZKAN*** [b], ***Nurettin DOĞAN*** [b]

[a] *Sakarya University of Applied Sciences, School of Information Technologies, Department of Software, Application Development and Analysis, Sakarya, Turkiye.*
[b] *Selçuk University, Faculty of Technology, Department of Computer Engineering, Konya, Turkiye.*

ARTICLE INFO

Article history:
Received 18 November 2025
Accepted 29 December 2025

Keywords:
Deep Learning,
Object Detection,
Skin Disease Detection,
Skin Lesion Localization,
YOLO

ABSTRACT

Recent advances in the field of deep learning have revealed significant potential for automating the detection of skin diseases by overcoming the limitations of traditional diagnostic methods in terms of speed and accessibility. This study presents a comparative analysis aimed at evaluating the performance of the state-of-the-art object detection models YOLOv8 and YOLOv11 in detecting ten different skin disease classes using a publicly available dataset. The dataset consists of a total of 2,486 images, divided into training, validation, and test subsets, and both models were trained under identical conditions for 100 epochs on a Tesla T4 GPU. To assess model effectiveness, performance metrics including precision, recall, F1-score, and mean Average Precision (mAP@0.5) were analyzed. The results indicate that, compared to YOLOv8, which achieved 84% precision and 91% recall, YOLOv11 attained 85% precision and 92% recall, demonstrating a limited improvement in both precision and recall metrics. Class-wise analyses revealed high detection performance for conditions such as Kurap (fungal infection), whereas relatively lower performance was observed for classes such as Bisul and Urticaria. Furthermore, YOLOv11 achieves this improvement while reducing the number of parameters by approximately 22%, thereby offering an advantage in terms of computational efficiency. These findings demonstrate that YOLOv11 provides a computationally efficient approach and indicate its potential use as a decision-support tool for dermatologists in resource-constrained environments.

This is an open access article under the CC BY-SA 4.0 license.
(https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Recent research on skin disease analysis has predominantly focused on image classification–based machine learning and deep learning approaches. These methods aim to automate dermatological diagnosis by assigning diagnostic labels to skin images, thereby reducing reliance on traditional techniques such as biopsies, which are time-consuming and costly [1]. Convolutional Neural Networks (CNNs) and transfer learning–based architectures have demonstrated strong performance in extracting discriminative visual features such as color, texture, and shape from skin images [2]. Numerous classification studies have reported remarkably high classification accuracy values. For instance, deep learning–based models have achieved accuracy rates as high as 99.04% across multi-class skin disease datasets [3], while other studies successfully classified eight different skin diseases with accuracy levels ranging between 80% and 90% [2]. Hybrid classification pipelines combining computer vision and machine learning techniques have also reported accuracy values of approximately 95% for multiple disease categories [4]. Furthermore, state-of-the-art CNN architectures such as Xception, DenseNet, ResNet, and AlexNet have consistently demonstrated superior classification performance on benchmark datasets including ISIC, HAM10000, and PH2, often exceeding dermatologist-level diagnostic accuracy [5], [6], [7], [8],

* Corresponding author. E-mail address: *bilalinan@subu.edu.tr*
DOI: 10.58190/ijamec.2025.155

[9], [10], [11], [12]. Despite these impressive results, classification-based approaches operate at the image level and do not explicitly provide spatial information regarding lesion location[13]. As a result, they lack the ability to indicate where pathological regions are located within an image, which limits interpretability and reduces their direct applicability in clinical decision support systems[14]. In contrast, object detection–based frameworks address this limitation by jointly performing disease identification and lesion localization through bounding box predictions, thereby enabling transparent visual validation and supporting clinically meaningful dermatological analysis[15].

However, most existing object detection studies in dermatology have focused on binary lesion detection or have been constrained by the use of older detection architectures. Nevertheless, real-world clinical deployment requires not only high detection performance, as measured by metrics such as precision, recall, and mAP, but also computational efficiency and inference speed. Although these factors vary significantly across different detector architectures, they remain insufficiently explored in dermatological applications[16].

These studies collectively demonstrate the potential of deep learning methods for skin lesion analysis across classification, segmentation, and detection tasks, while simultaneously revealing dataset- and evaluation-related challenges that motivate further research. Unlike image classification approaches, object detection methods provide spatial localization of skin lesions, which is critical for clinical interpretability and real-world clinical applications[17], [18].

Although numerous deep learning-based classification studies exist for skin disease analysis, comparative evaluations of recent YOLO-based object detection models remain limited. In particular, the performance differences between YOLOv8 and the newly introduced YOLOv11 have not been systematically investigated for multi-class skin disease detection[19].

As a result, recent studies have focused on deep learning methods for skin diseases. These methods have been shown to be advantageous in many aspects, such as speed, cost, efficiency, and time, compared to traditional pathological methods [20].

The primary objective of this study is to evaluate the performance of YOLOv8 and YOLOv11 in detecting ten different skin disease classes using a standardized dataset. We hypothesize that YOLOv11 will achieve comparable or superior detection performance while requiring fewer computational resources, thereby making it more suitable for deployment in clinical settings. To test this hypothesis, a publicly available dataset consisting of 2,486 images, partitioned into training, validation, and test subsets, is employed, and both models are trained under identical conditions. To determine model effectiveness, performance metrics including precision, recall, F1-score, and average precision (AP/mAP@0.5) are analyzed. Additionally, class-wise performance is examined to identify the strengths and weaknesses of the models in detecting specific skin conditions.

The main contributions of this study are as follows: (i) a comparative evaluation of YOLOv8 and YOLOv11 for multi-class skin disease detection, (ii) a standardized evaluation using object detection metrics, and (iii) an analysis of detection performance and model complexity. The remainder of this paper is organized as follows: Section II reviews related work, Section III describes the materials and methods, Section IV presents the experimental results, and Section V concludes the study.

2. LITERATURE REVIEW

A significant portion of recent studies on skin disease analysis has focused on image classification approaches based on convolutional neural networks and transfer learning techniques. Studies conducted on reference datasets such as ISIC, HAM10000, and PH2 in the literature have shown quite high success rates [7], [8], [9], [10], [11], [12]. However, the most fundamental limitation of these classification-based methods is that they do not provide localization information regarding the lesion's position on the image. Unfortunately, this situation limits the direct applicability and transparency of these systems in clinical decision support processes [17].

Object detection-based approaches developed to overcome these limitations aim not only to categorize skin diseases but also to determine the location of pathological areas using bounding boxes. Such applications have gained significant momentum in the field of medical imaging in recent years. Single-stage models, particularly Single Shot Detector (SSD) and YOLO, are frequently preferred in medical imaging due to their real-time detection capabilities and competitive detection performance in terms of precision, recall, and mAP.

Reviewing the dermatology literature, it is evident that Hsiao et. al. [21] achieved promising results in conditions such as psoriasis and atopic dermatitis using the SSD architecture. Using a similar approach, Ding et. al. [22] achieved high classification performance in clinical images by using deep convolutional neural networks for the automatic identification of benign pigmented skin lesions.

YOLO-based architectures have recently become the focus of academic interest due to their ability to strike a good balance between processing speed and detection performance. Talukdar et al. [2] achieved successful results in dermoscopic images using YOLOv8, while Elshahawy et al. [23] combined YOLOv5 and ResNet50 in a hybrid structure, bringing together the speed of YOLO and the deep feature extraction power of ResNet for early

melanoma diagnosis. In edge computing-focused scenarios, Datta et al. [24] demonstrated the potential of lightweight models for mobile dermatology applications through their work on YOLOv7-tiny. AlSadhan et al. [25] provided a comprehensive perspective on skin cancer detection by comparing the performance of various YOLO versions.

The effectiveness of YOLOv8 in dermatological applications becomes even more pronounced when combined with next-generation segmentation models. Gül et al. [17] developed a fully automated system by integrating this architecture with the Segment Anything Model (SAM) and demonstrated that YOLOv8 is an effective preprocessing tool for lesion localization. Furthermore, Ma et al. [26] and Li and Shen [27] reported important findings highlighting the superiority of pre-trained models and automatic feature extraction in medical imaging tasks.

Despite all these advances, a comprehensive comparison of YOLO architectures in the detection of multi-class skin diseases is still limited. Most existing studies generally focus on binary classification or a small number of disease categories [18], [27]. However, scenarios involving ten or more diseases better reflect real-world challenges such as data imbalance, intra-class variation, and inter-class similarity, which necessitates a more rigorous testing of the models' generalization ability.

The existing literature highlights a clear trend toward leveraging object detection models for the diagnosis of skin diseases, with YOLO variants standing out due to their favorable balance between inference speed and detection performance. However, comparative studies between YOLOv8 and YOLOv11, particularly in the field of dermatology, remain notably scarce. This gap is critical, as newer iterations of YOLO may offer substantial advantages in computational efficiency and detection performance, both of which are essential for clinical adoption.

3. Materials and Methods

The methodology employed in this study is designed to systematically evaluate the performance of YOLOv8 and YOLOv11 for skin disease detection. The technical details regarding the dataset, model architectures, experimental setup, and evaluation metrics are presented in this section. The overall methodological pipeline of the proposed system is illustrated in Figure 1.

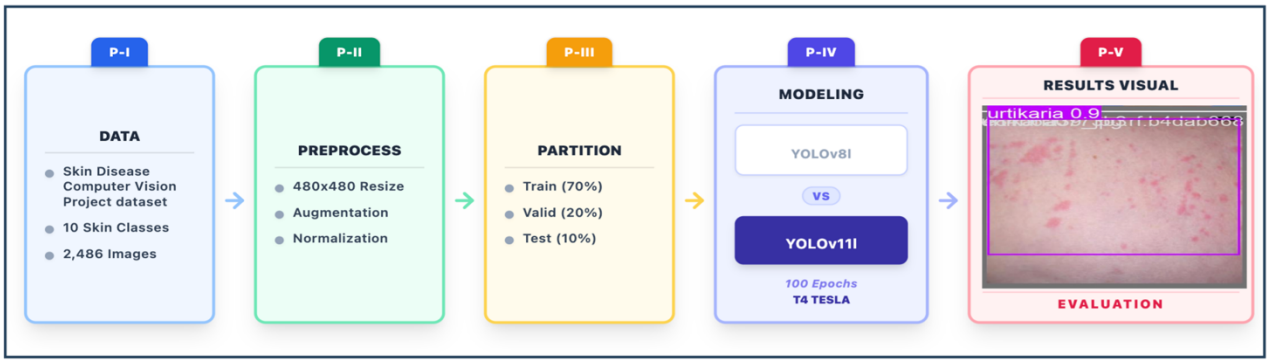

Figure 1. Overall Methodology Pipeline of the proposed system

Figure 2. Representative sample images from each skin disease class included in the dataset

3.1. Dataset Preparation

This study utilized images of real skin diseases obtained from the Skin Disease Computer Vision Project dataset available on the Roboflow platform [28]. The dataset was accessed in December 2024 and is publicly available for research purposes. No personal or identifiable patient information is included in the dataset, ensuring compliance with ethical research and data usage standards.

The dataset consists of a total of 2,486 images distributed across 10 labeled skin disease classes**,** namely: “biang_keringat (heat rash)”, “bisul (boil)”, “Herpes (cold sore)”, “jerawat (acne/pimple)”, “kanker_melanoma (melanoma cancer)”, “kurap (fungus)”, “kutil (wart)”, “psoriasis (psoriasis)”, “seborrheic_keratoses (seborrheic keratosis)”, and “urtikaria (hives)”. Representative samples from the dataset are illustrated in Figure 2.

The dataset was provided with a predefined and fixed split by the Roboflow platform, consisting of approximately 70% training (1,748 images), 20% validation (495 images), and 10% testing (243 images). Since the split was predefined at the dataset level, the same images were consistently assigned to each subset, thereby reducing the risk of data leakage between training, validation, and testing phases.

The class distribution within the dataset is not perfectly balanced, with certain skin disease categories containing a higher number of samples than others. This imbalance reflects real-world clinical data characteristics and introduces additional challenges for multi-class skin disease detection. To mitigate potential bias and improve model generalization, standard data augmentation techniques integrated into the YOLO framework were employed during training, and performance evaluation was conducted using class-aware object detection metrics.

Annotation files generated on the Roboflow platform were used in YOLO format. Each image was annotated using bounding boxes, with most images containing a single lesion, while some images include multiple lesion instances. The corresponding labels were stored in text files containing the class index and normalized bounding box parameters, including center coordinates (x, y), width, and height.

3.2. YOLO Architecture

YOLO (You Only Look Once) is a real-time object detection algorithm that uses convolutional neural networks (CNNs) to detect objects in a single pass [29]. By dividing images into grid cells, it enables faster detection compared to two-stage methods and is particularly useful for real-time applications. YOLO has evolved through multiple versions, each improving speed, detection accuracy, and the ability to detect objects of various sizes [29]. The YOLO object detection architecture is presented in Figure 3.

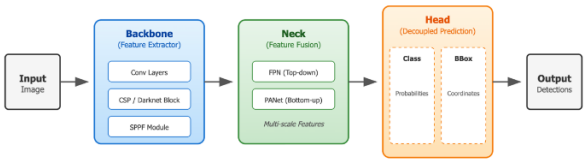

Figure 3. YOLO Object Detection Architecture[30]

The YOLO algorithm predicts bounding boxes to localize objects within an image and assigns an objectness confidence score to each prediction, indicating the likelihood that an object is present within the predicted region. Each bounding box is further associated with class probability scores representing the model’s confidence for each target class. A schematic illustration of the bounding box parameters is shown in Figure 4 [30].

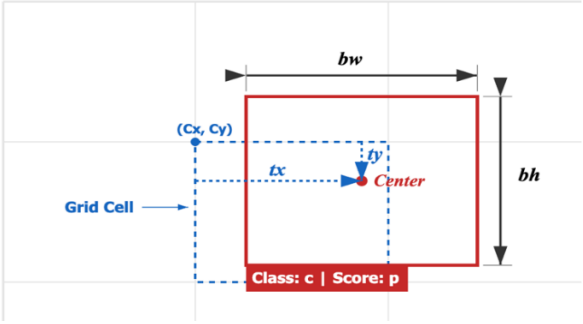

Figure 4. Bounding box prediction [31]

As illustrated in Figure 4, a bounding box is defined by its center coordinates (tx, ty), width (bw), and height (bh). The objectness confidence score reflects the probability of object presence within the predicted region, while class probabilities represent the likelihood of each class conditioned on the presence of an object.

YOLO processes images by dividing them into grids. Each of these divided grids attempts to determine whether an object is present within its own boundaries. When an object is detected, its center point is examined to check whether the object is within its own grid boundary. If this step is passed, a box is drawn around the object to confirm the object detection.

After the object detection process is completed, it is determined which of the labeled classes this object belongs to. Since many grids indicate that the object belongs to its own class, more than one box may be drawn. These boxes can sometimes be unnecessary boxes. This reveals YOLO's accuracy rate. The security score measures the accuracy rate.

3.2.1. YOLOv8

Developed by Ultralytics, YOLOv8 represents one of the most widely used versions in the YOLO object detection model family, offering speed and accuracy developed for practical applications [32]. It features a fully convolutional neural network with a backbone and head component that incorporates self-attention mechanisms for

multi-scale object detection. YOLOv8 supports advanced computer vision techniques such as instance segmentation and increases processing speed [33] . There are five variants of the model: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, catering to different tasks such as detection, segmentation, and classification [18]. The YOLOv8 model outperforms the YOLOv7 model in terms of accuracy and speed. YOLOv8 employs a modern, anchor-free architecture with a CSP-based backbone and a decoupled detection head, rather than the Darknet-53 backbone used in earlier YOLO versions.

3.2.2. YOLOv11

YOLOv11 is the most recent version of the YOLO family released by Ultralytics and incorporates architectural refinements derived from previous versions to improve detection efficiency. According to comparative evaluations reported in recent surveys, YOLOv11 demonstrates competitive performance in terms of accuracy–speed trade-off when compared with earlier YOLO variants, as illustrated in Figure 5 [34].

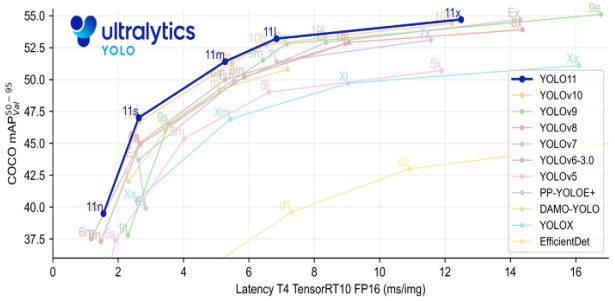

Figure 5. Accuracy comparison of YOLO versions [34]

As shown in Figure 5, when comparing YOLO versions, YOLOv11, the latest version of YOLO, has been found to have a higher accuracy rate than other versions. This demonstrates that the model effectively localizes and detects objects with high precision and recall. It employs a backbone–neck architecture, where the backbone extracts hierarchical feature representations from the input image, and the neck aggregates multi-scale features to support accurate object localization and classification. There are five variants of the model: YOLOv11n, YOLOv11s, YOLOv11m, YOLOv11l, and YOLOv11x, which cater to different tasks such as detection, segmentation, and classification. In this study, the YOLOv8l and YOLOv11l model variants were evaluated. According to the Ultralytics model specifications, YOLOv11l achieves comparable detection performance while using approximately 22% fewer parameters than YOLOv8l. It runs seamlessly on multiple platforms and provides maximum flexibility. It is an algorithm designed to tackle various computer vision challenges such as object detection, classification, instance segmentation, image classification, and pose estimation.

Table 1. Model Complexity Comparison of YOLOv8l and YOLOv11l [34]

Model	Parameters (M)	GFLOPs
YOLOv8l	43.7	165.2
YOLOv11l	34.1	128.4

As shown in Table 1, YOLOv11l reduces model complexity in terms of both parameter count and computational cost (GFLOPs) while maintaining comparable detection accuracy.

3.2.3. Model Configuration and Experimental Training Setup

Training was conducted in Google Colaboratory using a Tesla T4 GPU. The labeled data was first transferred to the drive environment via Roboflow using code, and then integrated into the working environment from the drive environment. The labeled data was fed into YOLO as 480x480 input. Both YOLOv8 and YOLOv11 models were trained using the Ultralytics YOLO framework with explicitly defined training configurations rather than unspecified default settings. The optimizer was Stochastic Gradient Descent (SGD) with an initial learning rate of 0.01. The batch size was automatically determined based on GPU memory. All experiments were conducted for 100 epochs with an input image size of 480×480 on a Tesla T4 GPU. The epoch count was set to 100. During training, the default data augmentation mechanisms integrated into the Ultralytics YOLO framework, such as mosaic augmentation and random flipping, were utilized to enhance model generalization. The main reason for this was to obtain faster training results. With these parameter settings, each training session lasted one and a half to two hours. The training was performed on the same dataset using the YOLOv8l and YOLOv11l model variants for two different models with the same parameters.

Table 2. Training and Evaluation Hyperparameters

Parameter	Value
Model variants	YOLOv8l, YOLOv11l
Input image size	480 × 480
Optimizer	SGD
Initial learning rate	0.01
Number of epochs	100
IoU threshold (evaluation)	0.5
Confidence threshold	0.25
NMS IoU threshold	0.45
Hardware	Tesla T4 GPU
Framework	Ultralytics YOLO (PyTorch)

3.2.4. Evaluation Protocol and Metrics

This section will discuss the most commonly used performance metrics in YOLO. The most commonly used performance metrics in YOLO-based object detection include Precision, Recall, F1-score, and mean Average Precision (mAP@0.5)[35]. Performance metrics are

metrics that show how well a model detects and identifies objects in an image. Additionally, they help understand how false positives and false negatives are handled. These metrics are crucial in the model's performance improvement, performance control, and performance evaluation stages [36]. In object detection tasks, true negative (TN) values are not explicitly defined. Therefore, the performance of YOLO-based models is evaluated using standard detection metrics such as precision, recall, F1-score, and average precision (AP/mAP). For evaluation, an IoU threshold of 0.5 was used to compute mAP@0.5. The confidence threshold was set to 0.25, and Non-Maximum Suppression (NMS) was applied using an IoU threshold of 0.45, following the default Ultralytics YOLO configuration.

The PR value, also referred to as the positive prediction value, is the probability that the actual positive boxes match the predicted bounding boxes. The calculation for the PR value is shown in equation 1.

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

Re is obtained by dividing the true positive value by the sum of the true positive value and the false negative value. It is a value that reveals whether basic real objects are correctly perceived or not. The formula used to calculate the Re value is shown in equation 2.

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

The F1-score is defined as the harmonic mean of precision and recall. It shows a blended average instead of a simple average. The reason for this is that extreme cases are disregarded. The formula for calculating the F1 score value is shown in equation 3.

$$F_1 = 2 \; x \; \frac{Pr \; x \; Re}{Pr + Re} \quad (3)$$

A precision curve is a measurement that provides an evaluation across the model by summarizing both precision and accuracy. Since both measurements are balanced on the values, it accurately reflects the measurement result on an unbalanced data set. One of the performance metrics below this curve is known as the AP curve. The purpose of AP is to monitor object performance. AP is computed as the area under the precision–recall (PR) curve for each class. For each class, Average Precision (AP) is computed as the area under the precision–recall curve at an IoU threshold of 0.5. Finally, mAP@0.5 is obtained by averaging the AP values of all classes at an Intersection over Union (IoU) threshold of 0.5. The mAP metric calculation is shown in Equation 4.

$$\text{mAP} = \frac{1}{\text{N}} \sum_{\text{i}=1}^{\text{N}} \text{AP}_\text{i} \quad (4)$$

In Equation 4, AP_i denotes the Average Precision of class i, computed as the area under the precision–recall curve at an Intersection over Union (IoU) threshold of 0.5, and N represents the total number of classes in the dataset. The final mAP@0.5 score is obtained by averaging the AP values across all classes.

4. Results

This section presents the empirical findings obtained from the comparative evaluation of the YOLOv8 and YOLOv11 models for skin disease detection. Performance metrics, training dynamics, and class-specific analyses are systematically reported to highlight the strengths and limitations of each model. The results demonstrate the effectiveness of both architectures while also revealing key differences in detection performance and computational efficiency.

In this study, the YOLO deep learning model, which has become quite popular in recent years, was compared between its YOLOv8 and YOLOv11 versions using the L package. The results of the findings will be shared in this section. Both versions were trained under the same conditions in the Google Colab environment (in terms of GPU resources). Figure 6 shows the detection results of the model on the test data.

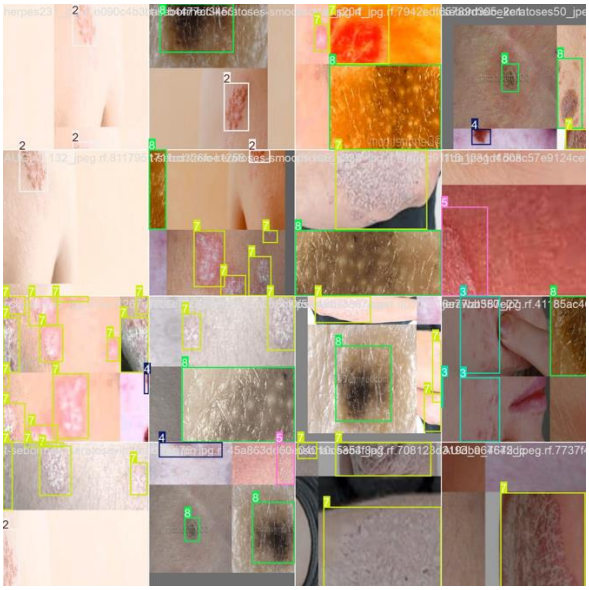

Figure 6. Model detections on the test data

Figure 6 displays the model's detections on the test data. The model draws bounding boxes in different colors according to the classes of the skin diseases it detects. It also writes the class score on them.

The YOLOv11 model demonstrated strong performance across all evaluated skin disease classes, achieving a mean Average Precision (mAP@0.5) of 85%, along with 92% recall and an F1-score of 87%. As shown in Table 3, the model exhibited particularly robust detection capabilities for conditions such as Kurap (fungal infection), attaining nearly perfect detection performance, with all evaluation metrics reaching 98%. This high level of performance can be attributed to the model's advanced feature extraction mechanisms, which effectively capture the distinctive visual patterns associated with fungal infections.

Table 3. YOLOv11 Skin Disease Detection Performance Metrics (mAP@0.5)

Class	Precision	Recall	F1	AP
Biang_keringat	0.89	0.89	0.89	0.89
Bisul	0.76	0.82	0.79	0.76
Herpes	0.71	0.94	0.81	0.71
Jerawat	0.89	0.92	0.90	0.89
Kanker_Melanoma	0.92	0.98	0.95	0.92
Kurap	0.98	0.98	0.98	0.98
Kutil	0.82	0.98	0.89	0.82
Psoriasis	0.90	0.98	0.94	0.90
Seborrheic_k	0.84	0.91	0.87	0.84
urtikaria	0.75	0.78	0.78	0.78
Mean	0.85	0.92	0.87	0.85

Table 3 presents the performance results of the YOLOv11 model for skin disease detection, summarized under four evaluation metrics: precision, recall, F1-score, and AP. The study revealed that precision and recall values ranged between 0.85 and 0.92. The F1 score was found to be 0.87. Overall, the model did not perform perfectly in the "Bisul" and 'urtikaria' classes, but in all other classes, especially the "kurap" class, the pr, rc, f1, and ap values were 0.98, indicating that the model achieved near-optimal detection performance for this class. The overall averages indicate that the model performs successfully. The same results are presented in Table 4, which compares the model with YOLO's v8 version.

Table 4. Comparison of YOLOv8 and YOLOv11 in Terms of Detection Performance Metrics (mAP@0.5)

Model	Precision	Recall	F1	mAP
YOLOv8	0.84	0.91	0.88	0.84
YOLOv11	0.85	0.92	0.87	0.85

Table 4 shows the performance of YOLO versions v8 and v11 on the dataset used in the study. While the pr value of the YOLOv11 version is 85%, the YOLOv8 version has a success rate of 84%. This reveals that the YOLOv11 model has a higher rate of correctly predicting positive values. It has been found that the Rc value of 92% in YOLOv11 is 91% in YOLOv8. This shows that the YOLOv11 model better predicts true positives. A comparison of the mAP@0.5 values reveals that YOLOv11 achieves 85%, while YOLOv8 achieves 84%. Based on these results, it can be said that, in general, the YOLOv11 version produces better results than the YOLOv8 version on the dataset used in the study. While both models show similar performance in terms of F1 score (88% and 87%), YOLOv11 stands out in terms of parameter efficiency and computational cost. This model is one step ahead in terms of detection performance, particularly with respect to recall and mAP@0.5, on a version basis. Based on these findings, it can be said that the YOLOv11 model yields better results on a version basis. From this perspective, it can be said that it would be more advantageous to prefer the YOLOv11 model in future studies.

The training dynamics of YOLOv8 exhibited consistent improvements across all monitored metrics, indicating stable convergence and an effective learning process. As illustrated in Figure 7, the box loss decreased from 1.163 to 0.449 over 100 epochs, demonstrating progressively more accurate localization of skin lesions. This reduction reflects the model's enhanced ability to refine bounding box coordinate predictions, which is critical for precisely delineating disease boundaries in dermatological images. Similarly, the classification loss (cls_loss) followed a comparable trend, decreasing from 2.556 to 0.267, indicating a marked improvement in discriminative capability among different skin disease classes.

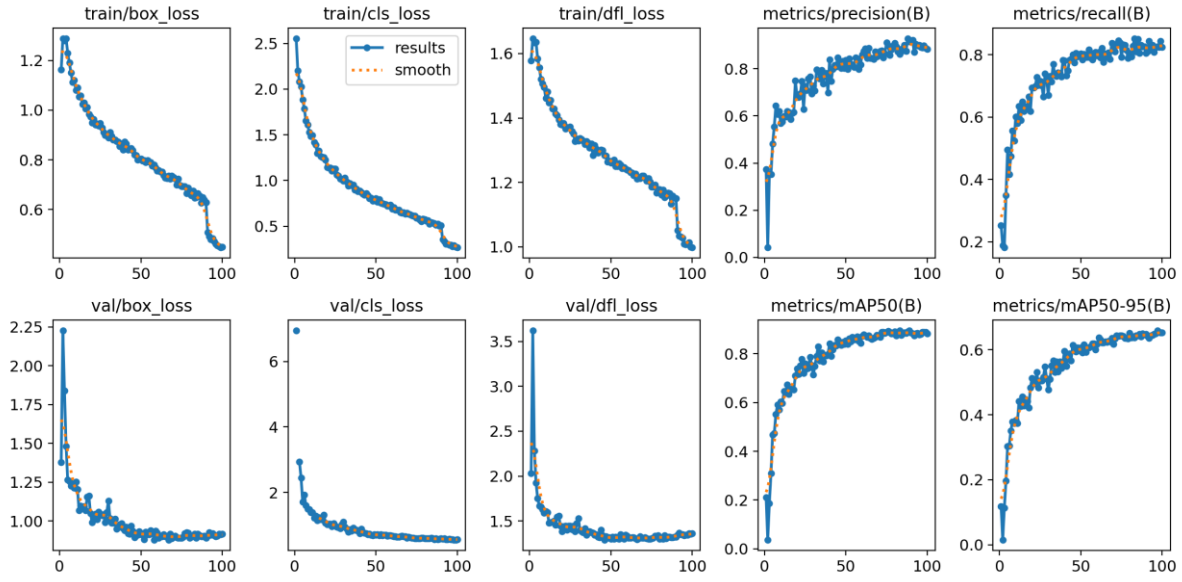

Figure 7. Training and validation loss and performance metrics over epochs for a YOLOv8 model

The performance metrics exhibited parallel improvements throughout the training process. Precision increased from 0.375 to 0.884, reflecting a reduction in false positive detections. This improvement is clinically significant, as the potential consequences of misdiagnosing malignant conditions directly affect the specificity of the diagnostic system. The recall value increased from 0.253 to 0.825, indicating an enhanced ability of the model to identify true skin disease cases. The mAP50 metric, which evaluates detection accuracy at a 50% intersection-over-union threshold, rose from 0.210 to 0.884, thereby confirming a comprehensive improvement in both localization and class prediction performance.

The training dynamics of YOLOv11 demonstrated consistent improvements across all evaluation metrics, exhibiting a stable trend particularly in loss reduction and performance enhancement. As shown in Figure 8, the box loss decreased from 0.805 to 0.480 over the 100-epoch training period, indicating progressively more accurate localization of skin lesions. This approximately 40% reduction in localization error suggests that the model architecture learned to delineate disease boundaries more precisely. The classification loss (cls_loss) declined from 0.789 to 0.288, reflecting strengthened discrimination among skin disease classes. Additionally, the distribution focal loss (dfl_loss) decreased from 1.211 to 0.994, demonstrating a meaningful improvement in the reliability of the model's predictions. Validation losses further supported these trends, with parallel decreases observed in val_box_loss, val_cls_loss, and val_dfl_loss, thereby confirming the model's ability to generalize to unseen data.

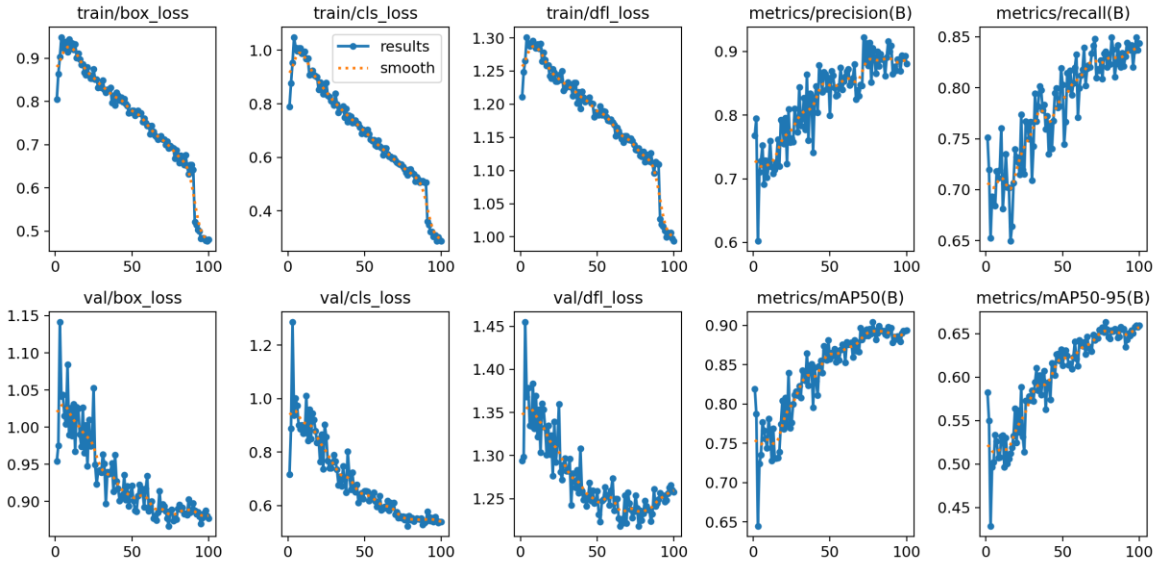

Figure 8. Training and validation loss and performance metrics over epochs for a YOLOv11 model

The performance metrics exhibited parallel improvements throughout training. The precision value increased from 0.768 to 0.881, indicating an approximately 15% improvement in the model's ability to suppress false positives. The recall metric rose from 0.751 to 0.844 (a 12% increase), demonstrating enhanced capability in detecting true disease cases. The mAP50 metric increased from 0.819 to 0.894, thereby confirming consistent progress in both localization and detection performance. The plateau observed in the metrics around the 85th epoch indicates that the model had largely extracted the learnable representations from the training data by this stage and had reached stable convergence.

5. Discussion and Conclusion

Object detection models in the diagnosis of skin diseases provide transparent evaluation in clinical processes thanks to their ability to determine the type of lesion and precisely mark its location [14], [21]. When reviewing the literature data, it is seen that the success of the models varies between 75% and 98% based on mAP, depending on architectural preferences and the number of targeted diseases.

One of the highest success rates in the literature has been reported by Elshahawy et al. [23]. The researchers achieved a mAP value of 98.3% with a hybrid architecture combining YOLOv5 and ResNet50 architectures. While these two studies demonstrate that supporting object detection networks with deep feature extractors enhances performance, such complex structures can remain cumbersome in mobile dermatology applications or scenarios requiring instant results due to high computational costs. Gül et al. [17] further validated this approach by integrating YOLOv8 with the Segment Anything Model (SAM) for automated skin cancer segmentation, demonstrating that hybrid architectures can achieve superior localization accuracy when properly optimized.

Focusing on broader datasets, Aishwarya et al. [19] achieved 88.0% mAP with YOLOv3 in a 9-class scenario.

Talukdar et al. [2] reached 88.2% mAP levels with YOLOv8 in an 8-class study. Hsiao et al. [21] employed a single-step multiframe detector for skin lesion identification, demonstrating effective real-time detection capabilities in clinical environments. Additionally, Ding et al. [22] successfully achieved automatic identification of benign pigmented skin lesions from clinical images using deep convolutional neural networks, highlighting the robustness of CNN-based approaches across diverse dermatological applications. The YOLOv11 model analyzed for performance in this study demonstrated an 85% mAP success rate across a broader spectrum of 10 different disease categories, presenting a graph consistent with the current literature. Although the increase in the number of classes is a factor that complicates the model's decision-making process, it was observed that the YOLOv11 architecture managed this complexity in a stable manner.

In terms of computational efficiency, the advantages offered by YOLOv11 are more clearly understood. Datta et al. [24] reported competitive detection performance using YOLOv7-tiny for edge device-focused applications, which is comparable to the computational efficiency offered by YOLOv11. The fact that the YOLOv11 model offers competitive detection performance despite using approximately 22% fewer parameters than its predecessor, YOLOv8, demonstrates the model's hardware efficiency. This indicates that the model has practical applicability potential for use in medical devices and smartphones with limited hardware.

The broader context of deep learning in medical imaging further supports these findings. Ma et al. [26] demonstrated the effectiveness of pre-trained convolutional neural networks for thyroid nodule diagnosis, illustrating the transferability of deep learning techniques across different medical imaging modalities. Li and Shen [27] conducted comprehensive skin lesion analysis towards melanoma detection using deep learning networks, establishing foundational methodologies that have influenced subsequent object detection approaches in dermatology. Furthermore, classical classification approaches have paved the way for the evolution of more advanced object detection frameworks that provide both classification and positioning capabilities.

In conclusion, compared to the 75.4% mAP score achieved by AlSadhan et al. [25] with YOLOv7 or the high hardware requirements of hybrid models, YOLOv11 demonstrates a more balanced performance profile in terms of detection performance, recall, and computational efficiency. With its 10-class broad coverage and recall rate reaching 92%, this study confirms that YOLOv11 is an optimized option suitable for dermatology practice, balancing speed and accuracy. The comparative analysis across multiple YOLO architectures and hybrid models underscores the continuous evolution of object detection methodologies in clinical dermatology, with YOLOv11 representing a significant step forward in achieving an optimal trade-off between detection performance and computational efficiency.

Declaration of Ethical Standards

This study does not involve any experiments with human participants or animals. The dataset used in this study is publicly available and does not contain any personal or identifiable patient information. Therefore, ethical approval and informed consent were not required.

Credit Authorship Contribution Statement

All authors contributed equally to this work. Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, and visualization were performed jointly by all authors. All authors have read and approved the final version of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding / Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The dataset used in this study is publicly available and can be accessed from the Roboflow platform at: https://universe.roboflow.com/capstone-7x5vm/skin-disease-7u8pj

References

[1] Leelavathy, S., Jaichandran, R., Shobana, R., Vasudevan, S. S. P., & Prasad, N. (2020). Skin disease detection using computer vision and machine learning technique. *European Journal of Molecular & Clinical Medicine*, *7*(4), 2999-3003.

[2] R. Talukdar, S. Dutta, and S. Das, "Enhancing skin disease diagnosis through convolutional neural networks and YOLO v8 object detection," in *Proc. 7th Int. Conf. Electron. Mater. Eng. Nano-Technol. (IEMENTech)*, Kolkata, India, Dec. 2023, pp. 1–6, doi: 10.1109/IEMENTech60402.2023.10423482.

[3] G. Poornima and D. S. Sakkari, "Diagnosis of skin diseases based on deep learning and machine learning approach: Technical review," *Int. J. Health Sci.*, pp. 4224–4240, Jul. 2022, doi: 10.53730/ijhs.v6nS6.10608.

[4] V. B. Kumar, S. S. Kumar, and V. Saboo, "Dermatological disease detection using image processing and machine learning," in *Proc. 3rd Int. Conf. Artif. Intell. Pattern Recognit. (AIPR)*, Lodz, Poland, Sep. 2016, pp. 1–6, doi: 10.1109/ICAIPR.2016.7585217.

[5] T. R. Anik, P. Talukder, I. Faruki, I. S. I. Rahman, and E. Hossain, "Analysis of automated skin disease classification exploiting different machine learning techniques," in *Proc. IEEE 13th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Las Vegas, NV, USA, Mar. 2023, pp. 1145–1151, doi: 10.1109/CCWC57344.2023.10099354.

[6] M. S. Junayed, A. N. M. Sakib, N. Anjum, M. B. Islam, and A. A. Jeny, "EczemaNet: A deep CNN-based eczema diseases classification," in *Proc. IEEE 4th Int. Conf. Image Process.*

Appl. Syst. (IPAS), Genova, Italy, Dec. 2020, pp. 174–179, doi: 10.1109/IPAS50080.2020.9334929.

[7] D. K. A. Al-Saedi and S. Savaş, "Classification of skin cancer with deep transfer learning method," *Comput. Sci.*, Sep. 2022, doi: 10.53070/bbd.1172782.

[8] A. Mehra, A. Bhati, A. Kumar, and R. Malhotra, "Skin cancer classification through transfer learning using ResNet-50," in *Emerging Technologies in Data Mining and Information Security*, A. E. Hassanien *et al.*, Eds., Singapore: Springer Nature, 2021, pp. 55–62, doi: 10.1007/978-981-33-4367-2_6.

[9] Y. Nie, M. Carratu, M. O'Nils, P. Sommella, A. U. Moise, and J. Lundgren, "Skin cancer classification based on cosine cyclical learning rate with deep learning," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Ottawa, ON, Canada, May 2022, pp. 1–6, doi: 10.1109/I2MTC48687.2022.9806568.

[10] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin cancer classification using deep learning and transfer learning," in *Proc. 9th Cairo Int. Biomed. Eng. Conf. (CIBEC)*, Cairo, Egypt, Dec. 2018, pp. 90–93, doi: 10.1109/CIBEC.2018.8641762.

[11] M. Upadhyay, J. Rawat, and S. Maji, "Skin cancer image classification using deep neural network models," in *Evolution in Computational Intelligence*, V. Bhateja *et al.*, Eds., Singapore: Springer Nature, 2022, pp. 451–460, doi: 10.1007/978-981-16-6616-2_44.

[12] A. Rezvantalab, H. Safigholi, and S. Karimijeshni, "Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms," 2018, *arXiv:1810.10348*.

[13] J. Amin, M. Azhar, H. Arshad, A. Zafar, and S.-H. Kim, "Skin-lesion segmentation using boundary-aware segmentation network and classification based on a mixture of convolutional and transformer neural networks," *Front. Med.*, vol. 12, p. 1524146, Mar. 2025, doi: 10.3389/fmed.2025.1524146.

[14] Y. Zhang, Y. Xie, H. Wang, J. C. Avery, M. L. Hull, and G. Carneiro, "A novel perspective for multi-modal multi-label skin lesion classification," 2024, *arXiv:2409.12390*.

[15] X. He *et al.*, "Computer-aided clinical skin disease diagnosis using CNN and object detection models," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Los Angeles, CA, USA, Dec. 2019, pp. 4839–4844, doi: 10.1109/BigData47090.2019.9006528.

[16] M. Yaseliani, A. I. Maghsoodi, E. Hassannayebi, and U. Aickelin, "Diagnostic clinical decision support based on deep learning and knowledge-based systems for psoriasis: From diagnosis to treatment options," *Comput. Ind. Eng.*, vol. 187, p. 109754, Jan. 2024, doi: 10.1016/j.cie.2023.109754.

[17] S. Gül, G. Cetinel, B. M. Aydin, D. Akgün, and R. O. Kara, "YOLOSAMIC: A hybrid approach to skin cancer segmentation with the segment anything model and YOLOv8," *Diagnostics*, vol. 15, no. 4, p. 479, Feb. 2025, doi: 10.3390/diagnostics15040479.

[18] R. A. Diptho and S. Basak, "Enhancing dermatological diagnosis through medical image analysis: How effective is YOLO11 compared to leading CNN models?" *NDT*, vol. 3, no. 2, p. 11, May 2025, doi: 10.3390/ndt3020011.

[19] N. Aishwarya, G. S. Y. Kannaa, and K. Seemakurthy, "YOLOSkin: A fusion framework for improved skin cancer diagnosis using YOLO detectors on Nvidia Jetson Nano," *Biomed. Signal Process. Control*, vol. 100, p. 107093, Feb. 2025, doi: 10.1016/j.bspc.2024.107093.

[20] S. G. Malik *et al.*, "High-precision skin disease diagnosis through deep learning on dermoscopic images," *Bioengineering*, vol. 11, no. 9, p. 867, Aug. 2024, doi: 10.3390/bioengineering11090867.

[21] Y.-P. Hsiao *et al.*, "Identification of skin lesions by using single-step multiframe detector," *J. Clin. Med.*, vol. 10, no. 1, p. 144, Jan. 2021, doi: 10.3390/jcm10010144.

[22] H. Ding *et al.*, "Automatic identification of benign pigmented skin lesions from clinical images using deep convolutional neural network," *BMC Biotechnol.*, vol. 22, no. 1, p. 28, Oct. 2022, doi: 10.1186/s12896-022-00755-5.

[23] M. Elshahawy *et al.*, "Early melanoma detection based on a hybrid YOLOv5 and ResNet technique," *Diagnostics*, vol. 13, no. 17, p. 2804, Aug. 2023, doi: 10.3390/diagnostics13172804.

[24] D. Datta, H. Prakash, and P. Singh, "Skin cancer detection with edge devices using YOLOv7 deep CNN," in *Proc. Data Anal. Manage.*, Singapore: Springer Nature, 2023, pp. 55–63, doi: 10.1007/978-981-99-6550-2_5.

[25] N. A. AlSadhan, S. A. Alamri, M. M. B. Ismail, and O. Bchir, "Skin cancer recognition using unified deep convolutional neural networks," *Cancers*, vol. 16, no. 7, p. 1246, Mar. 2024, doi: 10.3390/cancers16071246.

[26] J. Ma, F. Wu, J. Zhu, D. Xu, and D. Kong, "A pre-trained convolutional neural network based method for thyroid nodule diagnosis," *Ultrasonics*, vol. 73, pp. 221–230, Jan. 2017, doi: 10.1016/j.ultras.2016.09.011.

[27] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, p. 556, Feb. 2018, doi: 10.3390/s18020556.

[28] Roboflow, "Skin Disease Dataset," Roboflow Universe. [Online]. Available: https://universe.roboflow.com/capstone-7x5vm/skin-disease-7u8pj

[29] O. Kivrak and M. Z. Gürbüz, "YOLOv3, YOLOv4 ve YOLOv5 algoritmalarının performans karşılaştırması: Kümes hayvan tanıma için bir vaka çalışması," *Eur. J. Sci. Technol.*, Aug. 2022, doi: 10.31590/ejosat.1111288.

[30] A. A. Murat and M. S. Kiran, "A comprehensive review on YOLO versions for object detection," *Eng. Sci. Technol. Int. J.*, vol. 70, p. 102161, Oct. 2025, doi: 10.1016/j.jestch.2025.102161.

[31] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," 2016, *arXiv:1612.08242*.

[32] M. Hussain, "YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol. 11, no. 7, p. 677, Jun. 2023, doi: 10.3390/machines11070677.

[33] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," 2023, *arXiv:2304.00501*.

[34] Ultralytics, "YOLOv11 Model Documentation." [Online]. Available: https://docs.ultralytics.com/models/yolo11/

[35] R. Padilla, S. L. Netto, and E. A. Da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. 2020 Int. Conf. Syst. Signals Image Process. (IWSSIP)*, Niteroi, Brazil, 2020, pp. 237–242.

[36] İ. E. Parlak and E. Emel, "Deep learning-based detection of aluminum casting defects and their types," *Eng. Appl. Artif. Intell.*, vol. 118, p. 105636, Feb. 2023, doi: 10.1016/j.engappai.2022.105636.