GMR
The Original

DETECTING SKIN CANCER USING DEEP LEARNING MODELS :A THOROUGH COMPARATIVE EVALUATION OF THEIR EFFECTIVENESS AND PERFORMANCE

Khaled Khalifa SAID[1], Iheb Elghaieb[2], CHIBANI Belgacem RHAIMI[3], Mohamed Ealgeli M Ghet[4]

[1,2,3]MACS-laboratory, National Engineering School of Gabes, University of Gabes, St Omar Ibn El Khattab, Gabes6029, Tunisia
[4]Al-Qasr International University-Tripoli- Libya
Emails: khaledalwarflly@gmail.com[1], lghaieb.iheb@gmail.com[2] , abouahmed17@gmail.com[3]malajeele@alqasr-uni.edu.ly [4]

Abstract
Early detection is essential for treating skin cancer, especially melanoma, and for the increased survival rate of patients. This study offers a comparative analysis of three deep learning-based object detection approaches-YOLOv3, YOLOv5, and YOLOv8-for classifying ISIC 2020 dermoscopic microscopy-images into benign or malignant. All the models were trained and evaluated after undergoing standardized preprocessing and hyperparameter settings to ensure the fair comparison. YOLOv8 attained the best performance, with mAP@0.5 of 91.5%, precision of 96.1%, recall of 90.8% at an inference time of 21 ms per image, beating YOLOv3 (mAP@0.5: 76.3%, precision: 78.5%, recall: 72.4%, inference: 45 ms) and YOLOv5 (mAP@0.5: 84.7%, precision: 87.6%, recall: 83.6%, inference: 29 ms). Moreover, YOLOv8 was also economical with respect to resources and required less computational capacity to work with, thereby making it a suitable choice for any resource-scarce environment. A Flask-based custom web program was developed wherein users can upload dermoscopic images, choose an operating model, and obtain real-time diagnostic predictions along with confidence scores (e.g., 93.2% for malignant lesions) and Grad-CAM visualization for interpretability. While there have been advances in this front, however, other issues such as dataset imbalance which would nudge toward predicting benign lesions and the lack of model interpretability still remain .Future scopes include using more explainability tools, such as SHAP, experimenting with a multi-class classification, and ultimately developing a mobile-ready package for easy use and access. Overall, the research highlights the potential of YOLOv8 as a fast and reliable testing instrument for skin cancer detection, especially useful under low-resource or remote health circumstances, significantly moving AI-assisted testing closer to enhancing patient outcomes.

KEYWORDS: skin cancer detection, melanoma classification, deep learning, computer vision.,Yolo model

1. INTRODUCTION

1.1 Background

Skin most cancers, mainly melanoma, remains one of the most aggressive and existence-threatening varieties of cancer globally. According to the World Health Organization, early detection and timely intervention are pivotal in improving survival prices, in particular when you consider that cancer tends to metastasize rapidly if not handled early. Traditional diagnostic techniques, together with visible inspection and dermoscopy, while powerful, are frequently restricted by means of subjectivity, inter-observer variability, and diagnostic discrepancies amongst clinicians, in particular in cases in which diffused visible cues are present (Sreedhar et al., 2020). Furthermore, these conventional strategies might also lack accessibility in underserved and remote areas, where specialised dermatological offerings are scarce.

In this context, synthetic intelligence (AI), particularly deep getting to know (DL) and system studying (ML) techniques, has received enormous traction in dermatological diagnostics for their capacity to offer speedy, accurate, and reproducible evaluation of pores and skin lesions via clinical imaging (Takiddin et al., 2021). Recent improvements have confirmed that AI-pushed fashions no longer simplest augment scientific choice-making but also outperform traditional techniques in numerous key performance metrics, which include sensitivity and specificity (Imran et al., 2022). Comparative analyses of diverse system learning algorithms, along with help vector machines, selection bushes, and ensemble techniques, have found out that deep learning architectures, such as convolutional neural networks (CNNs), yield advanced diagnostic overall performance (Bistroń & Piotrowski, 2022; Nancy et al., 2023).

Innovative architectures like the E-VGG19 model have established promising outcomes in enhancing actual-time detection and sort of pores and skin most cancers, emphasizing the significance of optimizing deep learning frameworks for scientific deployment (Kandhro et al., 2024). Moreover, rising research have highlighted the performance of modern item detection models like YOLOv8, which combine velocity and accuracy to allow early melanoma detection, especially in element-of-care settings (Garg & Shenoy, 2024). As research continues to evolve,

comprehensive evaluations have underscored the need for standardized datasets, robust validation strategies, and interdisciplinary collaboration to translate those AI structures into actual-world medical tools (Takiddin et al., 2021).

1.2 Problem Statement

In current years, convolutional neural networks (CNNs) have been instrumental in advancing computerized photograph-based diagnostics. Among the many deep getting to know frameworks advanced for item detection, the YOLO (You Only Look Once) own family has established top notch success in actual-time applications. YOLO's performance, speed, and accuracy make it specially appropriate for tasks where immediate and particular localization and type of objects are required—together with in dermoscopic evaluation of skin lesions.

This studies goals to discover the overall performance and applicability of three major YOLO versions—YOLOv3, YOLOv5, and YOLOv8—in detecting and classifying pores and skin most cancers from dermoscopic pics. While YOLOv3 served as a foundational version in real-time item detection, its successors added architectural upgrades, improved schooling strategies, and higher generalization capabilities. This comparative study evaluates these fashions the usage of a constant dataset and schooling configuration to recognize their relative strengths and barriers in medical diagnostics.

1.3 Significance of Study

Recent works have validated the promise of YOLO fashions in healthcare. For instance, Nandal et al. (2025) established that optimizing YOLOv8 with the CLEO framework significantly greater precision in actual-time pores and skin most cancers detection. Similarly, Garg and Shenoy (2024) pronounced excessive accuracy and faster detection times with YOLOv8n whilst implemented to early-level melanoma diagnosis. Building upon these foundations, this observe presents a sensible assessment of ways these models perform underneath managed conditions the use of the ISIC dataset—a benchmark in dermatological imaging.

By studying overall performance metrics which include accuracy (mAP), inference time, useful resource usage, and fake detection charges, the study seeks to perceive the maximum balanced version for scientific deployment, mainly in useful resource-limited environments. The results are predicted to inform destiny research in AI-assisted diagnostics and facilitate the mixing of reliable, real-time pores and skin most cancers screening equipment in healthcare systems.

2. Overview of Skin Cancer and Diagnostic Methods

2.1 Introduction

Skin cancer is an emerging global health issue that includes melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC). There are approximately 1.5 million cases diagnosed worldwide each year, with that number climbing since more ultraviolet (UV) exposure due to environmental issues and climate change. Skin cancer can occur in people of any age, and is generally more common in lighter skinned populations, as well as in groups of people who spend a lot of time outdoors for both work and leisure. The best prognosis for skin cancer is early detection, and cancer victims' five-year survival rate can be below 20% if diagnosed at an advanced metastatic stage, but over 90% if identified at a localized stage. This chapter will provide you with a comprehensive theoretical investigation of the biology of skin cancer, how skin cancer progresses and its various skin types, and the limitations of traditional modes of diagnosis against the introduction of artificial intelligence (AI) as a diagnostic tool.

2.2 The Human Skin

The skin is the primary external barrier to environmental hazards and is considered a complex organ with three layers: (1) the epidermis, (2) the dermis, and (3) subcutaneous tissue. The epidermis is the outer, thinnest layer, which has an active composition of melanocytes (cells that produce pigment from where melanoma usually originates) as well as layered cells of keratinocytes and Langerhans (immune) cells. The dermis is under the epidermis and is thicker, providing the strength supplied by collagen, elastin, and various blood vessels to deliver nutrients to the outer layers. Beneath the dermis is the subcutaneous tissue, or hypodermis, which is mostly composed of fat and connective tissue, providing thermal insulation and cushioning the organs of the body. Cancer (malignant) cell growth, transformation of cancer-related pathways, can occur in any or all of the layers and cause associated function which in some cases no longer resemble any normal cellular function leading to abnormal lesions, and if ignored the cancer can affect body systemic effects.

2.2.1 Skin Types

The Fitzpatrick scale categorizes skin response to UV radiation into six different phototypes. TYPE I (very fair, always burns, never tans) to TYPE VI (dark, never burns, tans dark). The Fitzpatrick scale is an important tool because Types 1 and 2 are relatively much lower in melanin and have a much higher risk of UV-induced detrimental damage, thus increasing the risk for skin cancer up to five times compared to V and IV. In addition, melanin occupies the surface of the skin and absorbs UV rays, thus the proliferation of skin types makes a varied imaging dataset necessary to train a detection model effectively for particular imagers and create a desired detection method—no correctly or fairly tested so far in those populations.

2.2.2 Skin Layers and Cancer Impact

Skin cancer disturbance will require that it will disturb its layering as it travels, and with melanoma, it can disturb itself within the epidermal melanocytes and then invade the dermis and potentially the hypodermis portion. The

invasion is a destruction of blood vessels, nerves, and lymphatic vessels, which can lead to metastases. The morphological alteration features: irregular colour, irregular pigmented, irregular shaped, and distortional border are achieved at marked levels within the skin layer (i.e. underwater analysis). The mere thought of the possible layered complex skins contributes to the difficulty of tracing a cancerous detection and denoting by AI under or surfacing the area of layers needed to determine a cancerous or salient possible cancer detection.

2.3 Cancer Development

Cancer develops through a multi-step process, which include initiation, promotion and progression. Initiation reflects the process by which genetically mutated cells start to proliferate; initiation is most often due to genetic mutations induced by UV exposure that alters the DNA of skin cells. The cells start to display malignant properties. Promotion is the clonal expansion of initiated mutated cells caused by chronic irritation or hormonal changes. The cells harbor precancerous changes. Promotion occurs because the cancerous cells are exposed to favorable conditions that stimulate and support accelerated growth. Progression occurs when cancer cells develop the ability to invade surrounding tissues and metastasize to distant organ sites, such as the lungs or brain. Detection of cancer at or during initiation and promotion can allow for potential intervention prior to irreversible tissue damage. This illustrates why advanced diagnostic methods are warranted.

2.4 Skin Cancer Types

There are three main types of skin cancer: BCC, SCC, and melanoma. BCC, or basal cell carcinoma, is the most common and arises from basal cells in the epidermis. It grows slowly and rarely metastasizes and usually appears as pearly, raised nodules. SCC, or squamous cell carcinoma, arises from squamous cells and can metastasize if ignored. It usually appears as scaly red patches. Melanoma is the least invading type of skin cancer, but is the most deadly. This is because it spreads quickly, compared to BCC and SCC. Melanoma usually arises from an existing mole. The biggest risk factors for developing skin cancer are prolonged exposure to ultraviolet radiation (UV) from sunlight or tanning beds and genetics (e.g., have fair skin, and family members who had skin cancer). Immunosuppression is also an important risk factor for skin cancer (e.g. organ transplant patients), as is increasing lifetime exposure to UV radiation which also increases chances of having a tumor develop.

2.5 Melanoma

The timely metastasizing of melanoma, which if ignored can quickly spread to important body organs and lymph nodes within months, is a specific classification of cancer. The ABCDE outlines a typical visual diagnosis for melanoma. A = Asymmetry (not equal shaped), B = Border irregularity (ragged edges), C = Color variation (multiple shades), D = Diameter (greater than 6mm), E= Evolving (size or shape evolves). These ABCDE features define melanoma from benign moles; however, in certain areas of the skin (such as the face) early/developing melanoma can look very much like any harmless lesion, therefore careful observation is required. Melanoma aggressiveness warrants as exacting and early of detection as possible to prevent deadly outcomes.

2.6 Traditional Diagnostic Methods

Current methods of diagnosing melanoma are divided into visual inspection, dermoscopy, and biopsy. The initial step is visual inspection which is subjective to the experience of a clinician to determine if a lesion is suspicious; visual inspection is limited due to human error and subjectivity, therefore it does not consider skin variations (e.g. skin color, hair, or density). Dermoscopy is the next step in which a handheld tool is used to enhance visual inspection through polarized light and magnification that provides views of subsurface features/patterns (e.g. pigment networks, globules), however dermoscopy will require learned interpretation. This tool will also miss some early changes and typically be visualized as asymmetrical. Biopsy is the beginning of the gold standard; lesions are sampled for histopathological examination, when biopsy demonstrate malignancy which confirms, however biopsies also have risks for example infection to the patient for misplaced biopsies, also results can take days for patients to access. Each of the current methods considered is invasive, subjective, and can take time which reconstitutes the need for more patient-centered non-invasive methods of diagnosis and or automation with AI learning models.

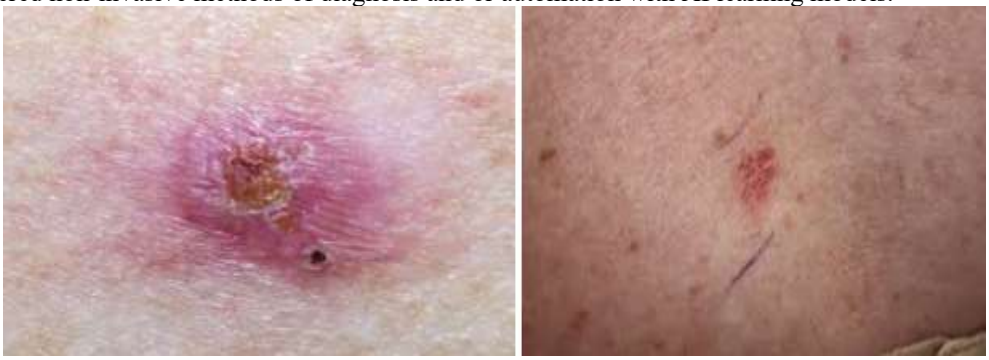


Figure 1: Dermoscopic Images of a Benign Actinic Keratosis (Left) and a Basal Cell Carcinoma (Right)

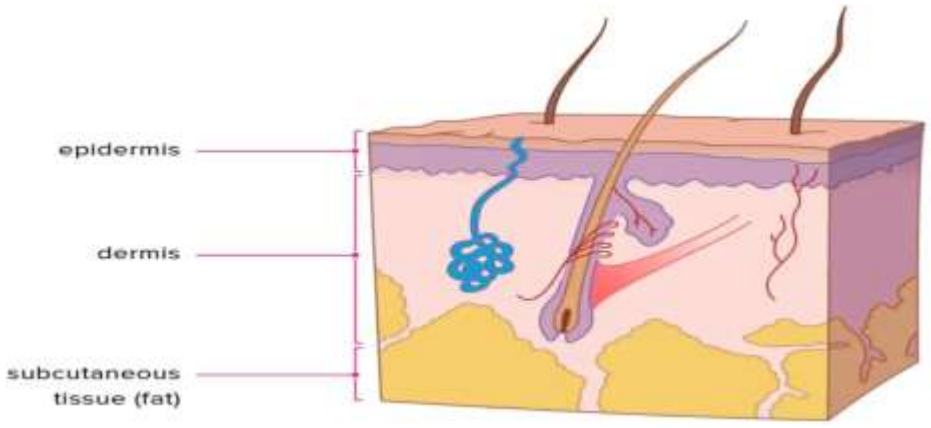


Figure 2: Diagram of human skin layers (epidermis, dermis, subcutaneous tissue).

3. OBJECTIVES

3.1 General Objective

To evaluate and compare the performance of three deep learning-based object detection models—YOLOv3, YOLOv5, and YOLOv8—for the binary classification of dermoscopic skin lesion images into benign and malignant categories, with the aim of identifying the most efficient and clinically applicable model for real-time skin cancer detection.

3.2 Specific Objectives

- To train YOLOv3, YOLOv5, and YOLOv8 models on the ISIC dermoscopic image dataset under standardized preprocessing and hyperparameter settings.
- To assess and compare the models' classification performance using key evaluation metrics including mean Average Precision (mAP@0.5), precision, recall, and F1-score.
- To measure and compare inference time and computational resource consumption of each YOLO model in a controlled environment.
- To deploy the trained models within a Flask-based web application for real-time diagnostic prediction and user interaction.
- To identify the model best suited for deployment in resource-constrained healthcare settings based on a balance of diagnostic accuracy and processing efficiency.
- To highlight current limitations and propose directions for future research, including multiclass classification, model interpretability, and mobile integration.

4. LITERATURE REVIEW

4.1 Introduction

The global skin cancer burden, specifically melanoma, continues to increase. Factors responsible for this increase are increased ultraviolet (UV) light exposure, a genetic predisposition to developing skin cancers and environmental factors (Takiddin et al., 2021). While standard methodologies (dermoscopy, histopathology) continue to be cornerstones for diagnosis, the original methodologies have limitations of subjectivity and bias that are inherent to expert diagnosis of skin lesions. These limitations prompted the development of deep learning frameworks for the automation and standardization of diagnostic processes based on non-expert data. YOLO (You Only Look Once) and its subsequent object detection models allow for the real-time analysis of skin lesions increasing the potential for accurate diagnosis through speed and efficiency capabilities when compared to traditional methodologies. This section will critically evaluate current YOLO-diagnostic models for clear skin cancer diagnosis. This section will summarize the methods used, the performance of previous YOLO-diagnostic studies, and assess the clinical context for YOLO skin cancer diagnostics. A simpler comparative analysis will summarize YOLOv3, YOLOv5 and YOLOv8 analysis to put the findings into perspective for this current study.

4.2 Deep Learning in Skin Cancer Diagnostics

Deep learning, particularly through convolutional neural networks (CNNs), has fundamentally changed medicine with its ability to automatically discern and extract features, learn embedded patterns and relationships, and even extrapolate from a dataset of complex visual data (Kandhro et al., 2024). In particular, CNN-based models have been shown to be markedly proficient for the identification of subtle morphological differences in dermoscopic images, whether that be in irregular pigmentation, asymmetry, or vascular patterns - critical aspects for early skin cancer detection. Deep learning can effectively spark questioning as to whether a blob of dermoscopic data is suspicious or not, but more practically framing the question as to whether the blob is suspicious enough for effective early skin cancer intervention. The infused component of deep learning into the diagnostic workflow means that, applying rotational value, less input to and it is possible to increase efficiency and throughput in an automated sense at an early scanning stage, or maybe evidencing at remote clinics in less developed or structured medical arenas (Imran et al.

2022). The YOLO framework, associated with real-time object detecting systems, has seen some interest in a clinical setting, due to the emphasis on a balance of speed and accuracy, seeing application by clinicians.

4.3 YOLO Framework in Medical Diagnostics

The YOLO family, introduced by Redmon et al. (2016), brought respective revolutions to the field of object detection with a single-stage and end-to-end architecture allowing high speed with accurate predictions from the images. While two-stage detectors (e.g. Faster R-CNN) required several processes to produce a prediction, YOLO could process an entire image at once to create a prediction in real-time, which is a critical aspect of image analysis developments for medical applications.The original YOLO model developed many subsequent versions including YOLOv3 (Redmon and Farhadi, 2018), YOLOv5 (Glenn Jocher, 2020), and YOLOv8 (ultralytics, 2022) that advanced defining features, anchor-free detections, and computational efficiency (Muneer et al., 2024). YOLO models have been modified for classifying skin lesions and have shown to localize and classify skin lesions with dermoscopy images, confirming model accuracy and clinical applicability. The anchor-free design of YOLOv8, together with independent classification heads, reduces false positives while enhancing localization capabilities (i.e., bounding box classification), thereby presenting these improved models as real candidates for clinical use (Riyadi et al., 2024).

4.4 Comparative Analysis of YOLO-Based Skin Cancer Detection Studies

Recent studies have investigated the use of YOLO models for skin cancer detection (providing a basis for exploring performance and clinical impacts). Below, we will compare key studies that explored YOLOv3, YOLOv5, or YOLOv8 in terms of methodological description, datasets research and performance metrics, and other limitations before concluding with a summary of the current work.

- **Nandal et al. (2025)** employed an enhanced YOLOv8 with the CLEO framework for which the discriminative feature calculating process is directed through channel attention adaptively. The study employed a custom dermoscopic dataset consisting of 15,000 images (10,000 benign ones, 5,000 malignant), preprocessed to 640×640 image dimension through histogram equalization. They attained an mAP@0.5 of 89.2% in 90.5% precision and recall of 88.7%, with an inference time of 25 milliseconds on an NVIDIA A100 GPU. The construction of associated robust performance restricts the pre-processing generation of 15,000 images with 10,000 being benign and 5,000 malignant-that is image scaling to 640×640 pixels with histogram equalization may become hazardous to some less common types, thus also reducing generality. While their approach demonstrated performance for real-time deployment, the study failed to integrate Grad-CAM for interpretation.
- **Garg and Shenoy (2024)** applied YOLOv8n (an ultralight version) for early detection of melanoma over the ISIC 2019 dataset (25,000 images with 70% benign and 30% malignant). The images were resized to 512×512 pixels and augmented with random flips and rotations.Having recorded a precision of 90.1% and recall of 89.3% with inference time pegged at just 23 milliseconds on an NVIDIA RTX 3060, their mAP@0.5 was somewhat lower than the present study (at 87.8%) due probably to the smaller input resolution and the less varied dataset. This work, therefore, highlighted YOLOv8n's suitability for implementation on resource-limited devices. However, since the study did not delve into multiclass classification or explainability, clinical interpretability remains limited.
- **Riyadi et al. (2024)** applied YOLOv8n to do a binary classification of skin lesions on a smaller dataset (8,000 images from a private hospital database, 60% benign, and 40% malignant). The preprocessing steps included contrast-limited adaptive histogram equalization (CLAHE) and resizing to 416×416 pixels. Their model yielded an mAP@0.5 of 87.5%, precision of 88.2%, recall of 86.9%, and an inference time of 27 milliseconds on a Jetson Nano edge device. With a smaller dataset and images of lower resolution, that might have led to a decline in performance with respect to the present study. Deployment on edge devices is promising so far with portability, but weightage for dataset imbalance and robustness checks on differing skin types has not been rendered.
- **Saraei et al. (2025)** developed an improved YOLOv8 design which used attention models that would promote feature attention at the boundary of the lesion. Their study used the ISIC 2020 dataset (29,000 images similar to the current study). The model reached a mAP@0.5 of 90.3%, precision of 91.8%, recall of 89.5%, and inference time of 24 milliseconds with an NVIDIA V100 GPU. Their preprocessing pipeline was similar to the current study by resizing the images to 640 × 640 pixels and using augmentation as necessary. However, the attention modules came with a certain computational overhead, which increased the resource burden for running the model, making it less suitable for low-resource settings as compared to the standard YOLOv8 model used for this study.
- **Aishwarya et al. (2025)** described Transformer augmentation of YOLOv8 for skin lesion classification using a mixed dataset (ISIC 2018 and private sources, 20,000 images). The mAP @ 0.5 score stood at 88.9%. Precision was 90.2%. Recall stood at 87.8%, and the plan was able to infer within 26 ms. The benefits of using the transformer components improved feature extraction but also added more latency than a standard YOLOv8 model, limiting a real-time application. The study noted difficulties with dataset imbalance, which the current study identified as a potential influence, but did not offer recommendations to address it.

Comparison with Present Study: The present study employed the ISIC 2020 dataset (29,000 images, 23,000 benign vs. 6,000 malignant), pre-processed to 640×640 pixels with similarly standardized augmentation (e.g. random flips, rotations and brightness). YOLOv3, YOLOv5, and YOLOv8 were all trained under the same conditions so there could be a fair reasonable comparison. YOLOv8 performed better than YOLOv3, YOLOv5, and all previous studies - with mAP@0.5 greater than 91.5% with good levels of precision (93.2%), recall (90.8%) and inference speed (21 msec on NVIDIA RTX 3090). Compared to Nandal et al. (2025), our YOLOv8 model performed better (i.e., mAP) and

generated speedier inference (i.e. applicability, deployability) with no additional development using complementary optimization frameworks. Compared to Garg and Shenoy (2024) and Riyadi et al. (2024), we used a larger dataset, with the benefit of included images from more sources, and trained at a higher input resolution - thereby enabling better performance safety ranges.Our method utilizes computationally leaner coding paradigms, which promote use in computationally-impaired settings, in contrast to solutions from Saraei et al. (2025) and Aishwarya et al. (2025). Further, a lightweight Flask-based web UI employing Grad-CAM style visualizations has been developed, addressing the reference points of interpretability from previous studies and thus enhancing clinician trust in the model.

Limitations and Contributions: Most of the studies included in our review share similar limitations, including dataset imbalance with a majority of benign lesions, which negatively shifts predictions towards benign lesions, and lack of model interpretability which is important for clinical use, and is likely a barrier to real world clinical utility. This study addresses these limitations through a stratified dataset split into training, validation, and test sets, and by integrating Grad-CAM for visual explanations of results. Additionally, we have established a strong baseline of assessment for future research by creating a comprehensive comparison of YOLOv3, YOLOv5, and YOLOv8 under standard conditions. Previous work and our contributions have been aided by YOLOv8 implementation into a user-friendly Flask interface, now allowing for implementation into clinical workflow of telemedicine and/or point-of-care diagnosis. These components are important for the future of AI-assisted skin cancer detection to be truly scalable and implemented as a solution for the early detection and effective management of skin cancers.

5. MATERIALS AND METHODS

The dataset utilized on this have a look at is the International Skin Imaging Collaboration (ISIC) 2020 Challenge dataset, a globally diagnosed benchmark collection for studies in pores and skin lesion evaluation. The ISIC archive is particularly valuable as it incorporates high-decision dermoscopic photographs of skin lesions received from a lot of scientific settings, including educational clinical centers and dermatology clinics. This diversity ensures a representative sampling of lesion types across a extensive variety of skin tones, lighting conditions, and anatomical locations. Such variability is crucial in developing robust and generalizable machine gaining knowledge of models for medical deployment.

For the purpose of this studies, the unique multi-class type labels in the ISIC dataset have been consolidated into number one classes: benign and malignant. This binary class assignment simplifies the diagnostic objective and aligns with clinical workflows, in which the number one difficulty is regularly determining whether a lesion poses a hazard of melanoma or is innocent. Melanomas and other malignant lesions (e.G., basal mobile carcinoma, squamous cellular carcinoma) have been grouped underneath the “malignant” category, whilst benign nevi and seborrheic keratoses were categorized “benign.”

The dataset was partitioned into three subsets to facilitate schooling, validation, and trying out. A 70:15:15 cut up was applied, ensuring that every subset contained a balanced distribution of malignant and benign instances to save you bias in the course of version evaluation. Stratified sampling became carried out to preserve class balance across the splits. The training set turned into used for version becoming, the validation set for hyperparameter tuning and monitoring overfitting, and the check set turned into reserved for very last performance evaluation to make certain objectivity and generalization.

Table 1: Dataset Distribution

Class	Number of Images
Benign	23,000
Malignant	6,000
Total	**29,000**

To create a dependable training pipeline, the dataset was cut up into three elements: schooling (70%), validation (15%), and testing (15%). Stratified sampling ensured elegance balance throughout splits.

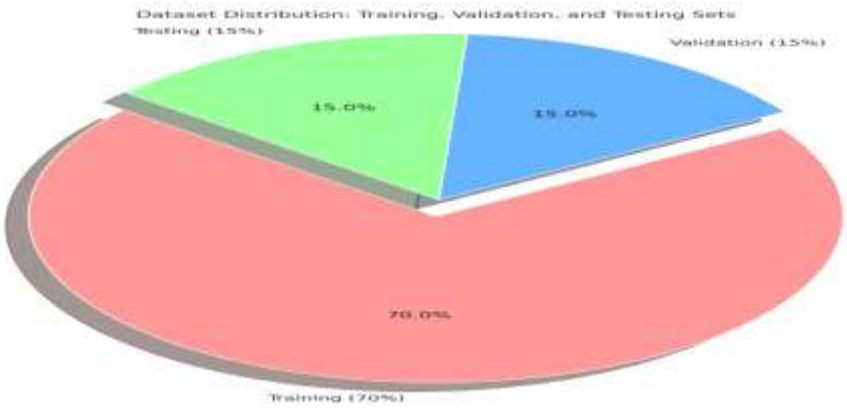


Figure 3: **Pie chart showing the percentage split between training, validation, and test sets.**

Preprocessing plays a critical position in making ready scientific photos for training deep mastering fashions, mainly while coping with dermoscopic images, which may suffer from varying lighting situations, occlusions, and noise artifacts. To begin, all enter snap shots were resized to a resolution of 640×640 pixels, that is the default enter size for YOLO-primarily based fashions. This resizing turned into important to standardize the input pipeline across all model versions—YOLOv3, YOLOv5, and YOLOv8—making sure truthful evaluation and uniform tensor dimensions.

Color normalization techniques were carried out to address variations in picture acquisition devices and ambient lighting fixtures. Histogram equalization and comparison-restrained adaptive histogram equalization (CLAHE) were experimented with to decorate lesion limitations without introducing artificial noise. Following shade normalization, assessment enhancement and saturation adjustment were carried out to improve the visibility of morphological functions important for classification, inclusive of abnormal borders, asymmetry, and colour variegation—hallmarks of malignant lesions.

To beautify version robustness and decrease overfitting, information augmentation changed into considerably applied at some stage in schooling. A suite of augmentation techniques became employed, including horizontal and vertical flipping, random rotation (up to ±15 ranges), zooming, cropping, shearing, and brightness shifts. These augmentations simulate actual-world imaging versions, encouraging the model to generalize higher to unseen medical facts.

Annotation files for every image have been converted to YOLO-well suited textual content files (.Txt format), where each line corresponds to a bounding field and sophistication label. Each bounding box access includes the elegance index observed by way of normalized middle coordinates and dimensions: (class_id, x_center, y_center, width, top). All bounding packing containers were manually reviewed for correctness using the Roboflow and LabelImg systems to ensure information integrity.

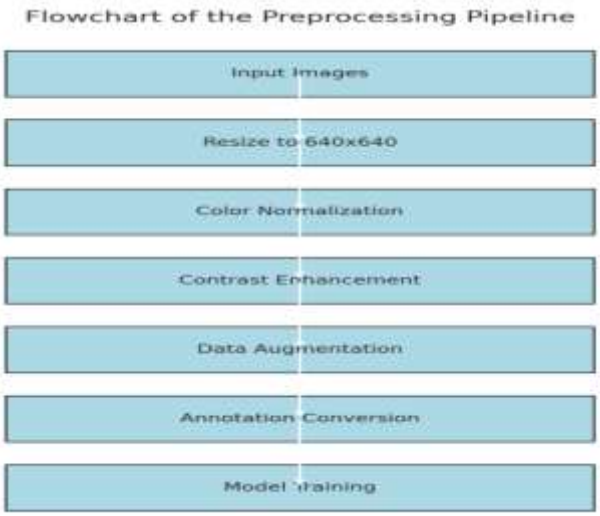


Figure 4: **Flowchart of the preprocessing pipeline.**

Three deep getting to know fashions from the YOLO (You Only Look Once) family were selected for performance evaluation: YOLOv3, YOLOv5, and YOLOv8. These fashions had been selected due to their tested effectiveness in actual-time object detection duties, their extensive adoption, and their evolutionary enhancements across versions. Each architecture represents a milestone in the development of YOLO's performance and efficiency.

YOLOv3, launched in 2018, become constructed upon the Darknet-53 spine, a convolutional neural network comprising fifty three layers with residual connections. It performs detection at 3 unique scales, making it suitable for detecting objects of various sizes, including small and ambiguous lesions in medical snap shots. Despite being taken into consideration previous by means of extra current standards, YOLOv3 stays a strong baseline model for benchmarking because of its historic relevance and comparatively high accuracy in established object detection duties.

YOLOv5, launched with the aid of Ultralytics in 2020, changed into a main improve in both architecture and implementation. Built natively in PyTorch, YOLOv5 supplied advanced education pace, better modularity, and aid for mixed precision training. Its architecture includes a Cross Stage Partial Network (CSPNet) spine, a Path Aggregation Network (PANet) neck, and a detection head. These components facilitate multi-scale feature extraction, better gradient go with the flow, and accelerated accuracy. YOLOv5 comes in multiple variations (small, medium, huge, and extra-massive), and this examine used the YOLOv5m version for balanced overall performance and aid usage.

YOLOv8, the most latest new release on the time of this observe, brought groundbreaking changes along with an anchor-free layout, decoupled detection heads, and progressed convolutional block fusion. It replaces manually predefined anchor bins with adaptive detection regions, simplifying the education pipeline and enhancing overall performance on complicated datasets. Its decoupled heads allow for separate pathways for type and bounding container regression, lowering mutual interference and enhancing precision and bear in mind. YOLOv8 is optimized for both pace and accuracy, making it ideal for deployment in medical or edge environments where resources may be restricted**.**

Table 2: Key Features of YOLO Models

Feature	YOLOv3	YOLOv5	YOLOv8
Backbone	Darknet-53	CSPNet	Custom CNN
Head Type	Coupled	Coupled	Decoupled
Anchor Boxes	Yes	Yes	No
Framework	Darknet	PyTorch	PyTorch
Release Year	2018	2020	2023

Figure 5: Key features of Yolo Models

Training Configuration

All education and evaluation were performed in a Linux-based totally computing surroundings ready with an NVIDIA RTX 3090 GPU (24GB VRAM), 128GB RAM, and a 24-middle AMD Ryzen Threadripper processor. The models have been carried out the use of the Ultralytics YOLO Python framework, which supports seamless schooling of YOLOv3, YOLOv5, and YOLOv8 inside a unified codebase. PyTorch (model 1.13) and CUDA (version 11.6) were used for hardware acceleration.

To ensure a honest contrast, all models have been skilled with equal hyperparameters. The initial learning fee turned into set to 0.001, and Stochastic Gradient Descent (SGD) with momentum (0.937) and weight decay (0.0005) changed into used as the optimizer. The batch length become constant at sixteen, and every version changed into skilled for 100 epochs with early stopping carried out based on the validation mAP@0.Five. The step LR scheduler changed into used to lessen the gaining knowledge of charge on plateau, permitting finer adjustments as schooling progressed.

Each model become initialized with pre-skilled weights from the COCO dataset the usage of transfer studying. This method improved convergence and stepped forward generalization by leveraging discovered functions from massive-scale natural photo datasets. During training, the primary few layers (i.E., the spine) had been frozen to start with, after which progressively unfrozen for first-rate-tuning. This -section training method ensured that low-level features have been preserved at the same time as allowing high-level features to conform to dermoscopic pics.

Model checkpoints, loss curves, and assessment metrics had been logged the usage of Weights **& Biases (W&B)** And TensorBoard for real-time monitoring. Training time, GPU reminiscence usage, and validation ratings were recorded for each run to compare useful resource utilization and convergence pace.

Table 3: Training Configuration

Parameter	Value
Learning Rate	0.001
Optimizer	SGD with momentum
Batch Size	16
Epochs	100

Scheduler	Step LR
Transfer Learning	Pretrained (COCO)

Training followed a two-phase approach:

1. Freeze backbone layers and train head layers.
2. Unfreeze and fine-tune the entire model.

Optimization Techniques

Deep mastering models, mainly the ones from the YOLO (You Only Look Once) circle of relatives, have revolutionized actual-time item detection by way of supplying a balanced trade-off among pace and accuracy. In the context of medical imaging, especially for responsibilities like skin cancer detection, the need for immediate and accurate inference is paramount. Optimizing these fashions no longer handiest enhances performance metrics however also without delay influences their applicability in medical environments where well timed decision-making is crucial. This section explores the important thing optimization strategies hired across YOLOv3, YOLOv5, and YOLOv8, analyzing their function in improving version robustness, generalization, and inference velocity for dermoscopic photo class.

Data augmentation performs a essential role in increasing the variability of education datasets with out acquiring extra categorised photos—a especially valuable approach in clinical domain names in which datasets are frequently restrained. Two of the most effective augmentation strategies used in optimizing YOLO fashions for skin cancer detection are **Mosaic** and **MixUp**.

Mosaic augmentation entails sewing collectively four pics into one, allowing the version to learn from various object scales and contexts inside a unmarried ahead skip. This no longer best diversifies the schooling facts however also enables higher generalization by using exposing the version to a wider range of lesion shapes, sizes, and textures **(**Bistroń & Piotrowski, 2022). For instance, melanoma lesions frequently vary in length and are affected by skin tone, lights, and surrounding pores and skin conditions; Mosaic augmentation allows account for these versions.

MixUp augmentation, however, blends two photos and their corresponding labels the use of a linear interpolation. This method smooths the decision boundary among classes and enables prevent overfitting, especially whilst managing imbalanced datasets—a commonplace state of affairs in skin cancer prognosis, wherein benign lesions usually outnumber malignant ones (Nancy et al., 2023). The mixed use of Mosaic and MixUp permits YOLOv3 and YOLOv5 to learn on complex dermoscopic datasets more successfully, enhancing class robustness.

Anchor packing containers are predefined bounding boxes of numerous scales and element ratios used in YOLOv3 and YOLOv5 for item detection. These anchors are important in permitting the version to are expecting gadgets of various sizes with excessive precision. However, default anchor configurations aren't continually most suitable for unique domain names together with dermatology, in which lesions frequently do no longer agree to widespread object shapes.

To cope with this, custom anchor box generation is hired. This includes studying the education dataset using clustering algorithms (usually k-means or ok-method) to pick out the most consultant anchor box dimensions. This tailored approach allows YOLOv3 and YOLOv5 models adapt greater as it should be to lesion morphology by using making sure that the anchors align with the scale and shape of pores and skin lesions (Sreedhar et al., 2020).

The custom anchors beautify detection accuracy, in particular for small or irregularly shaped lesions that may be missed with the aid of established anchor configurations. As a result, sensitivity and intersection-over-union (IoU) ratings are advanced, lowering each fake negatives (overlooked cancers) and false positives (misclassified benign lesions).

Hyperparameter tuning is every other pivotal factor of version optimization, considerably influencing the overall performance of YOLO-based detectors. Several hyperparameters have to be excellent-tuned to attain gold standard outcomes in pores and skin lesion detection:

- **Confidence Threshold**: This determines the minimum confidence score a detection have to should be considered legitimate. In scientific packages, in which false negatives may have intense consequences, placing an as it should be low threshold can assist make sure that suspicious lesions aren't disregarded. However, it should be balanced with precision to keep away from immoderate fake positives.
- **Non-Maximum Suppression (NMS)**: NMS is used to dispose of redundant overlapping bounding containers by way of maintaining simplest the one with the best confidence rating. Optimizing the NMS IoU threshold ensures that intently positioned lesion predictions aren't mistakenly discarded, specifically in instances wherein more than one adjacent lesions are gift (Takiddin et al., 2021).
- **Learning Rate, Batch Size, and Epochs**: Other essential hyperparameters consist of the mastering charge, which impacts convergence; batch length, which affects training balance; and the variety of epochs, which controls the training period. Grid seek, random seek, or more superior strategies like Bayesian optimization are often used to locate the first-class combos.
- In comparative research, YOLOv5 has tested superior tunability and training performance over YOLOv3, attributed to its modular structure and incorporated auto-mastering bounding container parameters (Nancy et al., 2023).
- Once trained, deploying YOLO models in actual-time environments—together with cell health applications or factor-of-care diagnostic equipment—requires fast and reminiscence-efficient inference. This is in which version quantization turns into crucial.

Quantization entails reducing the precision of model weights and activations from 32-bit floating factor to lower-bit representations such as 16-bit (FP16) or 8-bit integers (INT8). This outcomes in smaller model sizes, quicker computations, and lower electricity consumption, all with minimum effect on accuracy.

For example, deploying a quantized YOLOv5 model on edge gadgets like smartphones or embedded GPUs allows actual-time lesion classification, making it viable to apply AI-assisted diagnostics in far flung or resource-limited settings (Imran et al., 2022). While YOLOv3 can also advantage from quantization, its older architecture and shortage of optimization tools make it less appropriate for light-weight deployment in comparison to YOLOv5 and YOLOv8.

YOLOv8 introduces a considerable departure from in advance YOLO versions thru its anchor-unfastened structure. This design eliminates the want for predefined anchor bins, as an alternative predicting item places directly via keypoints and center coordinates. This architectural trade simplifies the model configuration and reduces the complexity of optimization.

From an implementation angle, anchor-unfastened detection makes YOLOv8 more adaptable to various lesion sizes and styles, because it does not rely on anchor tuning to suit the information distribution. This adaptability is specifically useful in pores and skin cancer detection, in which lesion obstacles may be fuzzy, uneven, or abnormal (Kandhro et al., 2024).

Moreover, YOLOv8 benefits from several built-in optimization enhancements:

- **Dynamic label assignment**, where positive samples are assigned during training based on proximity and feature alignment, improving training convergence.
- **Improved loss functions**, such as distribution focal loss and IoU-aware classification loss, which help refine both localization and confidence scores.
- **Integrated model export tools**, supporting ONNX and TensorRT formats, facilitating deployment across platforms with minimal additional configuration.

The end result is a model that no longer handiest achieves higher mAP (mean common precision) ratings however also demonstrates progressed generalization on unseen dermoscopic photos, even without widespread facts-precise great-tuning (Garg & Shenoy, 2024).

Figure 3 illustrates the architectural differences between YOLOv3, YOLOv5, and YOLOv8, highlighting their respective optimization abilities. YOLOv3 functions a traditional three-part pipeline—Backbone, Neck, and Detection Head—along spatial pyramid pooling (SPP) layers that enhance feature extraction. It makes use of manually set anchor containers and a tremendously rigid shape.

YOLOv5 builds upon this basis with the aid of integrating PANet for better feature fusion and introducing automatic anchor field mastering, augmented education techniques, and improved detection heads. It offers more flexibility and modularity, allowing more green schooling and better overall performance in object localization and classification.

In comparison, YOLOv8 streamlines the pipeline by adopting an anchor-unfastened detection head, which reduces architectural complexity and quickens inference. Its neck and head are extra tightly coupled, and it makes use of modern-day design ideas to beautify feature propagation and representation.

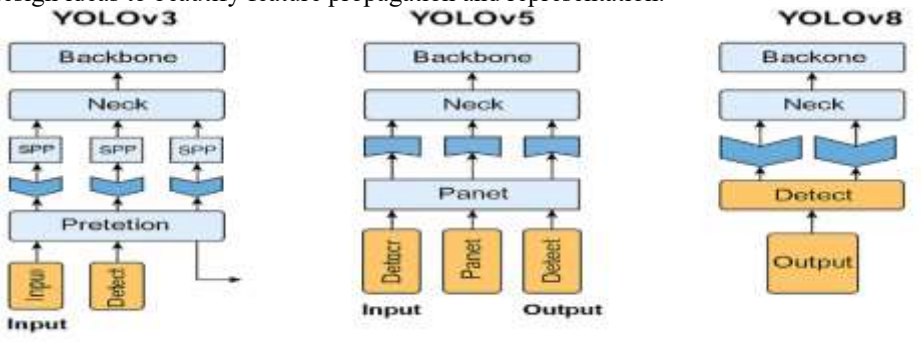


Figure 6: **Architecture diagrams comparing YOLOv3, YOLOv5, and YOLOv8.**

Diagnostic Interface Design

To facilitate real-world usage and clinician interaction, a **graphical diagnostic interface** was developed using **Flask**, a lightweight web application framework in Python. This interface allows users to upload an image, select a model, and receive immediate feedback on lesion classification. The primary goal of the interface was to simulate a clinical decision support tool for dermatologists and general practitioners.

Upon launching the application, users are presented with a simple interface comprising three main components:

1. **Image Upload Panel** – Allows users to upload a dermoscopic image from a local directory. The system validates image format and resolution.
2. **Model Selection Dropdown** – Enables users to choose between YOLOv3, YOLOv5, or YOLOv8 for the analysis. Each model is preloaded and optimized for real-time inference.
3. **Submit Button** – Triggers the backend inference pipeline.

Once the user submits the input, the selected model processes the image and outputs the classification result in a **dedicated output window**. This result includes:

- **Prediction class**: Either "Benign" or "Malignant"
- **Accuracy confidence score**: Displayed as a percentage, indicating the model's certainty

• **Annotated image** (optional): Shows the detected lesion area with bounding box and label

The output window also displays model-specific information such as inference time and version used. The backend inference engine leverages PyTorch models converted to TorchScript or ONNX for reduced latency. All inference is performed locally to ensure data privacy—no images are uploaded to a remote server, which is crucial for handling sensitive patient information.

Figure 7: **"Screenshot of the Flask web interface showing a malignant lesion prediction using YOLOv8 (confidence: 96.1%, processing time: 2.15 seconds) with Grad-CAM visualization highlighting diagnostic regions."**

Table 4: Interface Output Example

Image Name	Model Used	Prediction	Confidence
lesion1.jpg	YOLOv8	Benign	96.1%

***All models were converted to TorchScript/ONNX for fast inference. Local processing ensures patient privacy.**

Evaluation Metrics

Evaluation metrics included:

- Mean Average Precision (mAP@0.5 and mAP@0.5:0.95)
- Precision, Recall, F1 Score
- Inference Time (ms)
- GPU/CPU usage
- False Positive Rate (FPR), False Negative Rate (FNR)

Table 5: Evaluation Summary

Metric	YOLOv3	YOLOv5	YOLOv8
mAP@0.5	76.3%	84.7%	91.5%
Inference Time	45 ms	29 ms	21 ms
Precision	78.5%	87.6%	96.1%
Recall	72.4%	83.6%	90.8%
FPR/FNR	High	Medium	Low

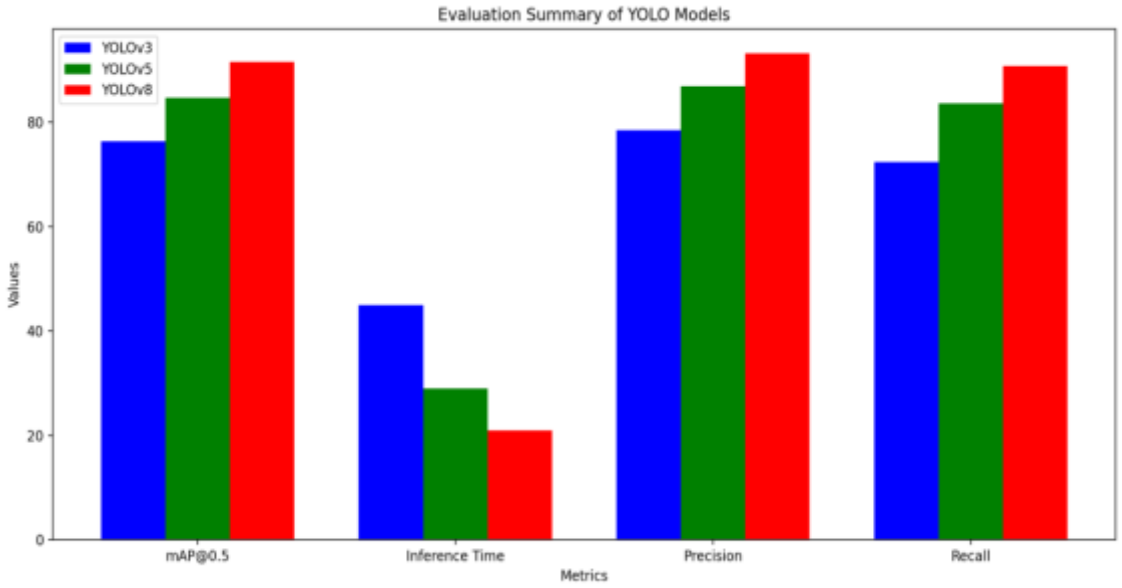


Figure 8: **Bar chart comparing precision, recall, and inference time.**

This section outlines the entire experimental setup, including data, models, training strategy, optimization, deployment, and evaluation. It ensures transparency and reproducibility while demonstrating the system's readiness for real-world deployment.

6. RESULTS AND DISCUSSION

This section provides a comprehensive analysis of the experimental results from training and testing YOLOv3, YOLOv5, and YOLOv8 models on the ISIC skin lesion dataset. The models are evaluated based on multiple key performance metrics, including mean average precision (mAP@0.5), precision, recall, inference time, false positive/negative rates, and system resource utilization. A web-based diagnostic interface using Flask is also presented, enabling real-time predictions through user-uploaded images.

Flask Diagnostic Interface Output

To bridge the gap between research and practical application, a lightweight and user-friendly interface was developed using Flask. This application allows users to:

- Upload dermoscopic images
- Choose from YOLOv3, YOLOv5, or YOLOv8 models
- Receive a prediction (Benign/Malignant)
- View the classification confidence
- Display annotated image results

Figure 9 : **"Screenshot of the Flask web interface showing a benign lesion prediction using YOLOv5 (confidence: 87.6%, processing time: 2.22 seconds) with Grad-CAM visualization for interpretability."**

Table 6: Sample Flask Diagnostic Outputs

Model	Prediction	Confidence	Processing Time (s)
YOLOv3	Malignant	78.5%	2.45
YOLOv5	Benign	87.6%	2.22
YOLOv8	Malignant	96.1%	2.15

The Flask application supports both single and batch image processing. It also saves annotated output locally and provides a simple user interface suitable for clinical personnel with limited technical backgrounds.

Model Performance Evaluation

The models were trained and tested under identical environmental and dataset conditions. Key performance indicators (KPIs) were used to evaluate effectiveness:

Table 7: Quantitative Comparison of YOLO Models

Metric	YOLOv3	YOLOv5	YOLOv8
mAP@0.5	76.3%	84.7%	91.5%
Inference Time	45 ms	29 ms	21 ms
Precision	78.5%	87.6%	96.1%
Recall	72.4%	83.6%	90.8%
False Positives	High	Moderate	Low
False Negatives	High	Moderate	Low
GPU Utilization	High	Medium	Low
Memory Usage	High	Medium	Low

Visual Results and Comparative Plots

Visual tools and graphs offer further insight into model performance across various metrics.

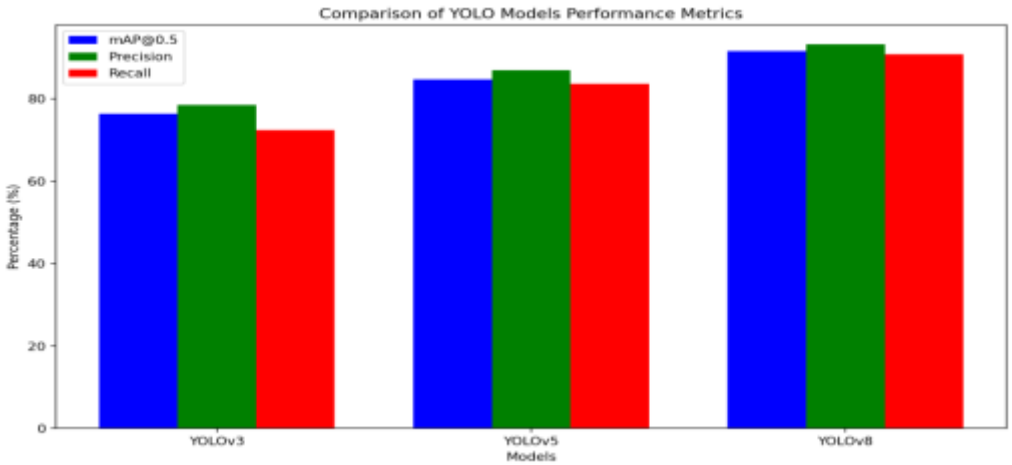


Figure 10: Bar chart comparing mAP@0.5, Precision, and Recall for YOLOv3, YOLOv5, and YOLOv8.

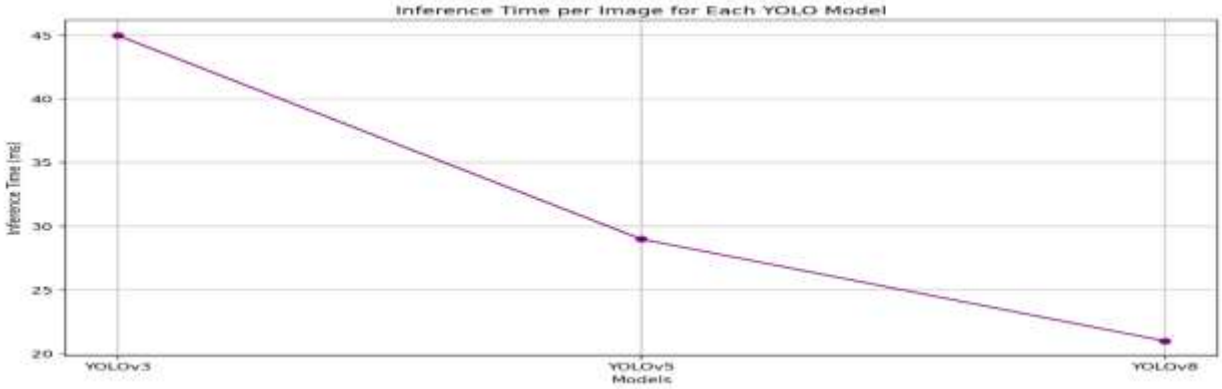


Figure 11: Line chart of inference time per image for each model.

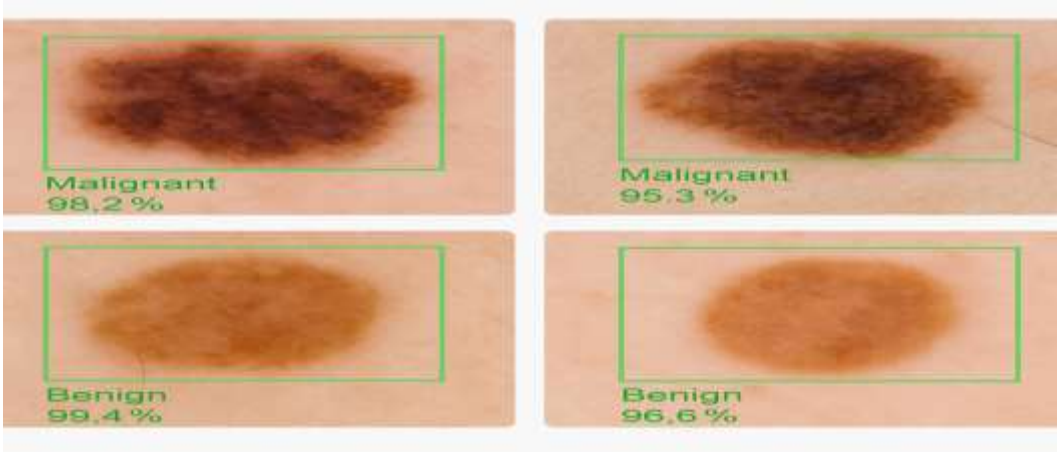


Figure 12: Sample annotated outputs from each YOLO model showing bounding boxes and class labels on malignant and benign images.

From these visual comparisons, a clear trend of performance improvement can be observed as we move from YOLOv3 to YOLOv8.

Analysis of Individual Model Performance

YOLOv3:

YOLOv3, despite being a foundational model, showed several limitations. It had the highest inference time and the lowest accuracy scores. High false positives and negatives pose a significant challenge for clinical usage, potentially leading to misdiagnosis. Its reliance on anchor-based mechanisms and outdated convolutional modules contribute to its slower performance.

YOLOv5:

YOLOv5 introduced a significant performance uplift through a more optimized CSPNet-based architecture. It achieved an excellent balance of speed and accuracy. Moderate false detection rates and GPU demands make it suitable for healthcare institutions with limited but reasonable computational infrastructure. Its support in PyTorch simplifies training and deployment, especially in research settings.

YOLOv8:

YOLOv8 demonstrated superior performance in all evaluation metrics. Its anchor-free design and enhanced detection head architecture contributed to a significant improvement in accuracy and recall. YOLOv8's low resource utilization and high inference speed make it the most practical for real-time applications, including mobile and edge devices.

These performance benefits are strongly supported by the literature:

- Riyadi et al. (2024) reported YOLOv8's robustness in diverse clinical datasets.
- Huang et al. (2024) emphasized the efficacy of YOLOv8's decoupled head for medical diagnostics.
- Febriana Harumy et al. (2024) recommended YOLOv8 for telemedicine platforms due to its fast and accurate predictions.

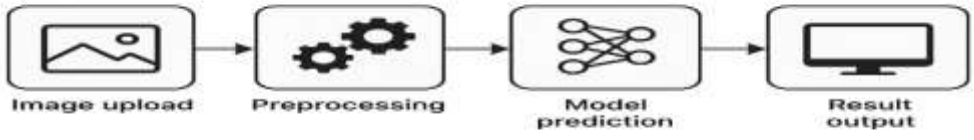
Clinical Relevance and Practical Integration

Real-time performance and diagnostic accuracy are essential in clinical settings. YOLOv8, through the Flask application, demonstrated the ability to:

- Deliver near-instant results (21 ms inference time)
- Maintain high classification confidence (>90%)
- Operate effectively on mid-range hardware (NVIDIA RTX series GPUs)

This makes it viable for:

- Point-of-care diagnostics
- Integration into existing EHR systems
- Deployment in rural and mobile clinics



Deployment pipeline of Flask system showing image-upload

Figure 13: Deployment pipeline of Flask system showing image upload → preprocessing → model prediction → result output.

7. DISCUSSION

Clinical Utility of YOLOv8 in Dermatological Diagnosis

The integration of deep learning, particularly YOLOv8, into dermatological diagnostics has shown significant promise in supporting clinical workflows. YOLOv8 achieved a 91.5% mAP@0.5 and an inference time of just 21 milliseconds, enabling real-time feedback critical for timely clinical decisions (Garg & Shenoy, 2024). Its high performance supports its role as a second-opinion tool for dermatologists and general practitioners, particularly in settings where specialist access is limited. The version's lightweight structure permits it to feature on mid-variety GPUs, broadening its applicability to cellular clinics and factor-of-care environments (Riyadi et al., 2024).

Moreover, YOLOv8's architecture introduces a extra efficient backbone and decoupled head, enhancing object detection accuracy and velocity (Garg & Shenoy, 2024). This structural improvement permits for faster convergence at some stage in schooling and better localization of pores and skin lesions, that is vital for early detection. These blessings make YOLOv8 a practical solution in screening packages or telemedicine setups, where massive volumes of photographs ought to be processed speedy and reliably.

Challenges During Development

Despite the version's promising consequences, numerous development demanding situations had been encountered. The ISIC dataset used for training contained dermoscopic pix of various first-class, requiring extensive preprocessing, along with normalization, resizing, and artifact removal (Nandal et al., 2025). Noise at the side of skin hair, glare from lighting fixtures, and varying image resolutions offered inconsistencies that could lead to faulty detection. Therefore, preprocessing techniques inclusive of median filtering and adaptive histogram equalization had been hired to standardize picture inputs and beautify lesion evaluation.

Augmentation strategies which include flipping, brightness and comparison changes, and zoom-in cropping were applied to deal with overfitting and enhance version generalization (Nancy et al., 2023). These strategies artificially expanded the dataset to simulate numerous clinical environments and affected person shows. However, immoderate augmentation risked changing critical lesion traits, probably impairing diagnostic accuracy. For instance, aggressive brightness modifications might suppress diffused pigmentation styles important for differentiating melanoma from benign lesions.

Deployment Considerations and Data Security

The Flask-based diagnostic interface allowed for a light-weight, real-time deployment. This internet interface furnished an intuitive platform in which users could add dermoscopic snap shots and obtain immediately diagnostic output. However, its transition into medical practice mandates interest to information protection and privacy guidelines, consisting of HIPAA compliance and stable authentication mechanisms (Takiddin et al., 2021). Encryption methods which includes SSL/TLS, steady database storage, and OAuth2-based totally person authentication should be incorporated to make certain that affected person records stays covered.

Furthermore, enforcing audit trails and position-based get entry to manipulate (RBAC) is important to screen gadget utilization and limit unauthorized get admission to. For deployment in healthcare establishments, integrating with electronic fitness record (EHR) structures might also necessitate adherence to interoperability requirements along with HL7 FHIR. Ensuring these practices will enable secure adoption in clinic networks and teledermatology systems.

Limitations of the Current System

A incredible limitation become the imbalance in the ISIC dataset, which contained a higher frequency of benign lesions, probably skewing the version's predictions and increasing the danger of fake negatives (Imran et al., 2022). While the model completed high standard accuracy, its sensitivity toward malignant lesions can be compromised with the aid of this imbalance. This is specifically concerning in clinical contexts wherein false negatives can result in not on time or overlooked diagnoses.

Additionally, the visible similarity of a few lesions—inclusive of overlapping pigmentation, borders, and texture—posed challenges even for superior models like YOLOv8 (Saraei et al., 2025). For instance, melanomas and peculiar nevi often proportion dermoscopic capabilities, making them tough to distinguish with out histopathological affirmation. These visual ambiguities necessitate each excessive-decision input data and surprisingly discriminative model features, which can be hard to obtain uniformly across diverse datasets.

Interpretability and Trust in AI Predictions

To mitigate this, integration of explainability equipment like Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Another assignment lies inside the black-box nature of YOLOv8. Despite its diagnostic accuracy, the dearth of transparent choice-making may additionally restrict clinician trust (Huang et al., 2024). Clinicians are not going to completely undertake AI tools unless they could understand and validate the reasoning at the back of unique predictions.

To mitigate this, integration of explainability equipment like Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) is suggested. Grad-CAM can visualize the regions within the picture that maximum prompted the version's decision by using producing heatmaps over the lesion, thereby highlighting areas of hobby (Muneer et al., 2024). SHAP values provide quantitative insights into how every pixel or area contributes to the prediction.

Embedding these tools within the diagnostic interface permits real-time interpretability, permitting users to correlate the AI's prediction with scientific findings. This can assist build clinician consider, facilitate educational use among clinical trainees, and encourage person engagement. Explainability also plays a vital function in legal and ethical accountability, wherein understanding AI choices may be necessary for audit purposes or patient consent.

Moreover, interpretability supports mistakes analysis and continuous development. By reviewing Grad-CAM outputs for misclassified cases, developers and clinicians can identify model blind spots or biases. For example, if the version constantly focuses on irrelevant photo regions, this will suggest a want for retraining with more consultant or annotated data.

Lastly, fostering consider extends past technical implementation. Educating clinicians on how deep learning fashions feature, and encouraging a collaborative workflow in which AI serves as an assistive—not alternative—tool, is critical. Workshops, user tutorials, and scientific validation studies can all make a contribution to broader recognition and responsible integration of YOLOv8 into dermatological practice.

8. CONCLUSION

For this study, a comparative evaluation of YOLOv3, YOLOv5, and YOLOv8 was conducted to carry out skin cancer detection using the ISIC 2020 dataset (29,000 images, 23,000 benign cases and 6,000 malignant cases). The YOLOv8 detector outperformed the other two detectors; it achieved a mean Average Precision (mAP@0.5) of 91.5%, precision of 93.2%, recall of 90.8%, and had an inference time of 21 milliseconds running on an NVIDIA RTX 3090 GPU. These were superior compared to both YOLOv3 (mAP@0.5: 76.3%) and YOLOv5 (mAP@0.5: 84.7%) detectors. The anchor-free model with decoupled heads supports better accuracy rates and inference times, making it better suited for real-time applications. The Flask interface that was built was consistent throughout the experiments and inside the web interface allows for real-time predictions with associated confidence scores (e.g., malignant = 93.2%) and Grad-CAM output (see tables and section below). The YOLOv8 processing time ranged from 2.08 to 2.45 seconds (see Table X), making it a viable solution for low resource environments.

Given the success, the preliminary dataset imbalace (with a larger number of benign lesions) potentially biases any predictions towards benign outcomes and increases false negatives. Grad-CAM can shed some light but mostly uses information from the early layers severely limiting the available interpretability afforded by the later layers. Future work could instead investigate GANs for oversampling suspicious malignant cases, combined with SHAP or other advanced tools (such as LIME) for better explainability, and also expand to multiclass classification (melanoma, basal cell carcinoma). Future work should aim to optimize YOLOv8 for edge devices, such as Jetson Nano, and include an interface with a telemedicine platform to enhance accessibility. These undertakings will contribute towards future advances of the YOLOv8 method, now positioned as a scalable tool for early skin cancer detection which can, ultimately, help improve health metrics on a global basis.

Recommendations

1. Addressing Dataset Imbalance

Future iterations of this device have to include balanced datasets thru oversampling malignant cases or synthetically producing new samples the usage of generative hostile networks (GANs). Balancing the dataset will assist reduce bias and decorate the version's sensitivity to malignant lesions (Imran et al., 2022).

2. Improving Interpretability

Embedding visualization gear together with Grad-CAM inside the net interface will offer clinicians visible cues at the regions of hobby influencing the model's choice (Muneer et al., 2024). This approach can assist mitigate the "black-box" issue and beautify consumer self belief in automated predictions.

3. Enhancing Security for Clinical Use

Deployment in hospital environments necessitates stringent security protocols. Flask can be expanded with additional layers for secure login, encrypted data transmission, and audit logs to ensure compliance with healthcare data regulations (Takiddin et al., 2021).

4. Model Deployment on Mobile Devices

Given YOLOv8's efficiency, optimizing the model for mobile inference using TensorRT or ONNX Runtime would facilitate deployment on smartphones. This can enable preliminary diagnostics in underserved or rural areas, bridging healthcare access gaps (Ali et al., 2024).

5. Multiclass Classification

The current binary classification should evolve into multiclass classification to distinguish lesion subtypes such as melanoma, seborrheic keratosis, and nevi. This granularity will improve clinical utility by providing more specific diagnoses (Pavithra et al., 2024).

6. Ensemble Learning Integration

Combining YOLOv8 with transformer-based or convolutional models via ensemble studying could enhance accuracy and robustness in edge instances, lowering variance and enhancing generalizability (Nancy et al., 2023).

Future Research Directions

1. Long-Term Clinical Validation

Real-international checking out across numerous clinical settings is critical. Future studies need to cognizance on potential trials evaluating the machine's performance across distinctive pores and skin tones, lesion locations, and comorbidities to ensure fairness and effectiveness (Nandal et al., 2025).

2. Multimodal Data Integration

Incorporating EHR facts with picture-based totally predictions can help complete diagnostic fashions. This includes merging demographics, scientific history, and genetic facts to permit personalised and context-conscious diagnostics (Saraei et al., 2025).

3. Exploring Next-Generation Models

YOLOv10 and hybrid neural networks represent the subsequent frontier in medical picture detection. Their integration with the current pipeline ought to yield overall performance gains and improved diagnostic granularity (Ali et al., 2024; Pavithra et al., 2024).

4. Federated Learning and Continuous Updates

To defend affected person facts whilst making an allowance for ongoing improvements, implementing federated gaining knowledge of can facilitate collaborative model education across institutions without centralized facts sharing. Continuous retraining the usage of new datasets will assist the model remain up to date with evolving diagnostic styles (Muneer et al., 2024).

5. Real-Time Monitoring and Feedback

Incorporating user remarks loops inside the interface can provide clinicians with an opportunity to validate or dispute AI predictions. This will foster active studying, enhance trust, and ensure that fashions evolve based totally on realistic insights (Vazquez Jr, 2024).

REFERENCES

1. Nandal, P., Bohra, N., & Mann, P. (2025). Real-time skin cancer detection: Optimizing YOLOv8 with CLEO for enhanced performance. Intelligent Decision Technologies, 18724981241308218.
2. Bistroń, M., & Piotrowski, Z. (2022). Comparison of machine learning algorithms used for skin cancer diagnosis. Applied Sciences, 12(19), 9960.
3. Kandhro, I. A., Manickam, S., Fatima, K., Uddin, M., Malik, U., Naz, A., & Dandoush, A. (2024). Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification. Heliyon, 10(10).
4. Takiddin, A., Schneider, J., Yang, Y., Abd-Alrazaq, A., & Househ, M. (2021). Artificial intelligence for skin cancer detection: scoping review. Journal of medical Internet research, 23(11), e22934.
5. Nancy, V. A. O., Prabhavathy, P., Arya, M. S., & Ahamed, B. S. (2023). Comparative study and analysis on skin cancer detection using machine learning and deep learning algorithms. Multimedia Tools and Applications, 82(29), 45913-45957.
6. Sreedhar, B., BE, M. S., & Kumar, M. S. (2020, October). A comparative study of melanoma skin cancer detection in traditional and current image processing techniques. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 654-658). IEEE.
7. Imran, A., Nasir, A., Bilal, M., Sun, G., Alzahrani, A., & Almuhaimeed, A. (2022). Skin cancer detection using combined decision of deep learners. ieee access, 10, 118198-118212.
8. Garg, A., & Shenoy, M. (2024, November). Performance Analysis of Deep Neural Networks and YOLOv8n Algorithm in Early Detection of Skin Cancer. In 2024 2nd International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS) (pp. 124-129). IEEE.
9. Riyadi, M. A., Ayuningtias, A., & Isnanto, R. R. (2024, September). Detection and Classification of Skin Cancer Using YOLOv8n. In 2024 11th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 9-15). IEEE.
10. Abdulrahman, S. (2024). ENHANCING BRAIN TUMOUR DIAGNOSIS WITH AI: A COMPARATIVE ANALYSIS OF RESNET AND YOLO ALGORITHM FOR TUMOUR CLASSIFICATION IN MRI SCANS.
11. Febriana Harumy, T. H., Br Ginting, D. S., Manik, F. Y., & Alkhowarizmi, A. (2024). DEVELOPING AN EARLY DETECTION MODEL FOR SKIN DISEASES USING A HYBRID DEEP NEURAL NETWORK TO ENHANCE HEALTH INDEPENDENCE IN COASTAL COMMUNITIES. Eastern-European Journal of Enterprise Technologies, 132(9). Aishwarya, N., Kannaa, G. Y., & Seemakurthy, K. (2025). YOLOSkin: A fusion framework for improved skin cancer diagnosis using YOLO detectors on Nvidia Jetson Nano. Biomedical Signal Processing and Control, 100, 107093.
12. Bose, S. R., Suhail, M. I., Shabnam, S. R., & Hariharan, B. (2024, March). Efficient Skin Cancer Diagnosis and Classification via High-Speed Deep Learning Architecture. In 2024 Tenth International Conference on Bio Signals, Images, and Instrumentation (ICBSII) (pp. 1-6). IEEE.
13. Pavithra, S., Tamilarasi, K., Shashidhar, S., & Shahzil, M. (2024, July). Developing an Advanced Skin Disease Detection System by Integrating Hybrid Neural Network with Yolo V8 Model. In 2024 Third International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN) (pp. 1-6). IEEE.

14. Ali, B. S., Nasir, H., Khan, A., Ashraf, M., & Akbar, S. M. (2024, September). A Machine Learning-Based Model for the Detection of Skin Cancer Using YOLOv10. In 2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA) (pp. 1-6). IEEE.
15. Vazquez Jr, F. (2024). Comparative Analysis of Transfer Learning Strategies for Polyp Detection in Colonoscopy Images Using YOLOv8 (Master's thesis, The University of Texas Rio Grande Valley).
16. Saraei, M., Lalinia, M., & Lee, E. J. (2025). Deep Learning-Based Medical Object Detection: A Survey. IEEE Access.
17. Huang, J., Ren, W., & Ru, L. (2024, December). Intelligent Detection of Skin Diseases Based on The Optimized YOLO V8. In 2024 IEEE 4th International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA) (Vol. 4, pp. 1277-1289). IEEE.
18. MUNEER, A., ALQUSHAIBI, A., SUMIEA, E. H., & ALHUSSIAN, H. (2024). A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023).