A Hybrid Ensembled Deep Learning Framework for Explainable Skin Cancer Classification and Lesion Detection

Rayala Upendar Rao
Assistant Professor
Department of Computer Science and Engineering,
RGUKT Nuzvid 521202, Andhra Pradesh, India.

Chowdam Naga Kishore
Student
Department of Computer Science and Engineering,
RGUKT-RK Valley Idupulapaya, 516330, Andhra Pradesh, India

ABSTRACT
Automated methods for dermoscopy image analysis support the detection of skin cancer in an early stage and reducing skin cancer-related deaths. Deep learning based methods have shown compelling results for the analysis of skin cancer images. In this paper, an enhanced ECRNet-based hybrid model is explored for the classification and detection of skin cancer and diverse skin anomalies. This model utilizes an ensemble of classification models, namely, ResNet50, ResNet101, MobileNetV2, Vision Transformer, Con-vNeXt, DeiT-Small, EL-DLOA, WavIntNet, Conformer, Xception, VGG16 and an Ensemble of ECRNet and other models. For lesion localization, the YOLO model family and Faster R-CNN architecture are examined. Experimental results demonstrate that the Hybrid-Ensemble approach outperforms other models with an accuracy of 97.2%, precision of 95.4%, recall of 93.7%, and F1 score of 94.5% while YOLOV26 achieved an mAP of 71.3% with a precision of 73.9%. Grad-CAM was used to improve the model interpretability by highlighting the image regions containing the lesions. A web application for image upload, automated prediction, and diagnostic visualization was developed using Flask and SQLite.

General Terms
Skin cancer, dermoscopic image analysis, deep learning, skin lesion classification, lesion detection, YOLO, Grad-CAM, explainable artificial intelligence.

Keywords
Skin cancer image recognition, attention mechanism, transformer, convolutional neural networks.

1. INTRODUCTION
Skin cancer is one of the most common types of cancer and the most common of all malignant tumors. It also occupies a significant place in global cancer mortality statistics [1]. In the last few decades, the incidence of one of the most aggressive and dangerous skin cancers, melanoma, has increased [2], [3]. Early identification of melanoma is of great importance for the population of patients with melanoma, as the early stages of this cancer can be successfully treated and the mortality rate can be significantly reduced[4]. The early screening of melanoma represents an important method in the extension of patient life and the enhancement of patient life quality.

An important innovation in the diagnosis of skin cancer is the introduction of melanoma dermoscopy [5]. Dermoscopy enables the detailed analysis of the skin and the differentiation of benign and malignant[5]. The detailed analysis provided by this imaging enables the formulation of different treatment plans for patients with suspicious lesions[5]. Even though dermoscopy contributes greatly to the early identification of lesions and melanomas, this imaging technique is not infallible. There is a degree of subjective observation [6]. The fine-tuning of melanoma diagnosis through dermoscopy requires diagnostic and clinical experience in the observation of dermoscopy, as this is how the extent of diagnostic error is diminished [6].

Severe consequences, including adverse outcomes, may be experienced by skin cancer patients due to delays in providing essential treatments, resulting from the misdiagnosis of skin cancer. In an effort to improve the accuracy of dermoscopy diagnoses and reduce risks associated with misdiagnosis, technology in the form of Artificial Intelligence (AI) has been developed. AI systems help dermatologists by analyzing dermoscopic images and giving less subjective answers. The combination of AI and dermoscopy holds great promise to enhance the accuracy and increase the speed of the diagnosis of skin cancer and the answer to it, and in particular for the menace of melanoma. This technology is highly impactful in providing better answers in the care of patients with skin cancer and improving outcomes of the patients.

2. RELATED WORK
The detection and classification of skin cancers are two of the major topics of research in the field today because of the rising cases of malignant skin lesions and melanoma. Improving the efficiency and accuracy of skin cancer diagnosis has been the objective of the application of numerous advanced machine learning and deep learning techniques.

Teodoro et al. [7] have proposed a skin cancer classification system based on Generative Adversarial Networks (GANs) and Region of Interest (RoI) attention. Their focus was on classification improvement by the generation of skin lesion images of superior quality with GAN and by the placement of attention within the image at focus by the attention mechanism. The use of GANs improved the ability to address the imbalanced datasets typical of medical images, and the RoI attention system enhanced the ability to focus on features relevant for accurate diagnosis.

An advanced transformer network for skin cancer classification was created by Xin et al. [7]. Typically, transformer models are more adept in capturing long-range dependencies. For this reason, several modifications were made to transformer models to enable better induction of local features when dealing with medical images. As part of this work, several constraints of skin cancer classification such as high computational requirements, localization of lesions, and complexity of the classification task were addressed using skin cancer classification models based on transformer architectures.

With respect to skin cancer classification, Datta et al. [8] also demonstrated the benefits of utilizing soft attention, incorporating soft attention into CNN models to enable the models to focus on relevant parts of the inputs while ignoring the

rest of the information. This technique was shown to improve classification performance while significantly decreasing the computational time required by the model.

A different skin cancer classification system was proposed by Ali et al. [9], using deep CNNs augmented with a transfer learning model. This technique improved classification performance on the more demanding skin cancer classification problem, especially on datasets with a small number of annotated samples. In this case, transfer learning significantly decreased the required training time and improved the model's performance.

Liu et al. [10] developed an adaptive weighted ensemble classifier for skin lesion classification based on their version of the Grey Wolf Optimizer (GWO). In their method, the GWO algorithm is implemented to determine the contribution of each classifier, thus setting the weight of each classifier in the ensemble. This method shows a high classification accuracy for several types of skin lesions, and the authors believe their method could help advance the field of medical diagnosis.

Unlike the studies mentioned above, Shi et al. [11] developed a new feature fusion network to classify imbalanced skin lesions, inspired by the work of wavelet-integrated multiscale feature fusion. Their method was based on the wavelet transform and extracted both global and local features from skin lesion images. The feature fusion network performed better than other networks in addressing imbalanced datasets, which is one of the biggest issues in skin cancer classification. The authors believe that including features on multiple scales can help improve the diagnosis of lesions that are particularly rare and complex.

Incorporating residual blocks and deep models, Bibi et al. [12] used a skin lesion classification approach using a feature selection method (called MSRNet). The authors of this study found that their classification system performed better as a consequence of their skin lesion classification dataset deep model residual block and feature selection. This study showed the need for advanced classification techniques in order to implement deep learning and feature engineering for skin cancer classification.

Veena Dillshad et al. [13] reported on a multi-class skin lesion diagnostic model called D2LFS2Net. Their model combined deep learning with a variance-controlled marine predator optimization algorithm. Their focus was on the adaptiveness of the model to the personalized profiles of the patients, what is known as precision medicine. They used optimization algorithms to enhance the selection of pertinent features of the model. Their study demonstrated that the use of novel optimization algorithms can help make deep learning models used in a medical context more interpretable and accurate.

3. METHODOLOGY

The described system employs a novel hybrid deep learning model for skin cancer classification and lesion detection that combines both image classification and object detection techniques. The model integrates numerous deep learning frameworks for skin lesion analysis, such as ResNet50 and 101, MobileNetV2, ViT, Con-vNeXt and its variants, DeiT-Small, EL-DLOA, WavIntNet, Conformer, Xception, VGG16 and the Hybrid-Ensemble, and ECRNet-CRO-RSA-Reptile. Deep learning image classification frameworks were implemented using PyTorch while TensorFlow was used for the proposed hybrid models and extensions to capitalize on their efficient preprocessing and fea-ture extraction. The model utilizes Faster RCNN, and numerous YOLO-based models, such as YOLOv5s6u, YOLOv5x6u, YOLOv8, YOLOv9, YOLOv11, and

YOLOV26 to localize lesions and detect abnormalities. Grad-CAM is used to visualize the areas of lesions to enhance interpretability of the model's classification. An easy-to-use interface is provided to users via a Flask based web application with an integrated SQLite database to safely register and log in users, upload images, pre-process the images, and make realtime predictions.

Figure 1 describes the entire architectural design of the suggested skin cancer analysis model. Initially, the framework takes the set of dermoscopic images as input. Prior to any further processing, images undergo several preprocessing actions, such as normalization and data augmentation that allow enhancing image quality, smoothing differences between samples, and increasing the variety of training data. This allows the model to learn more generalized and diverse features, which is very important for the effective work of the network. The preprocessed images then pass into the classification and lesion detection modules. Classification is performed with the use of several deep learning networks, including ResNet, Xception, VGG16, Hybrid-Ensemble, and proposed ECR-Net architectures. Parallel to the classification, lesion detection is done using YOLO-based detection models. These models not only classify images but also find locations of the lesions on the image.

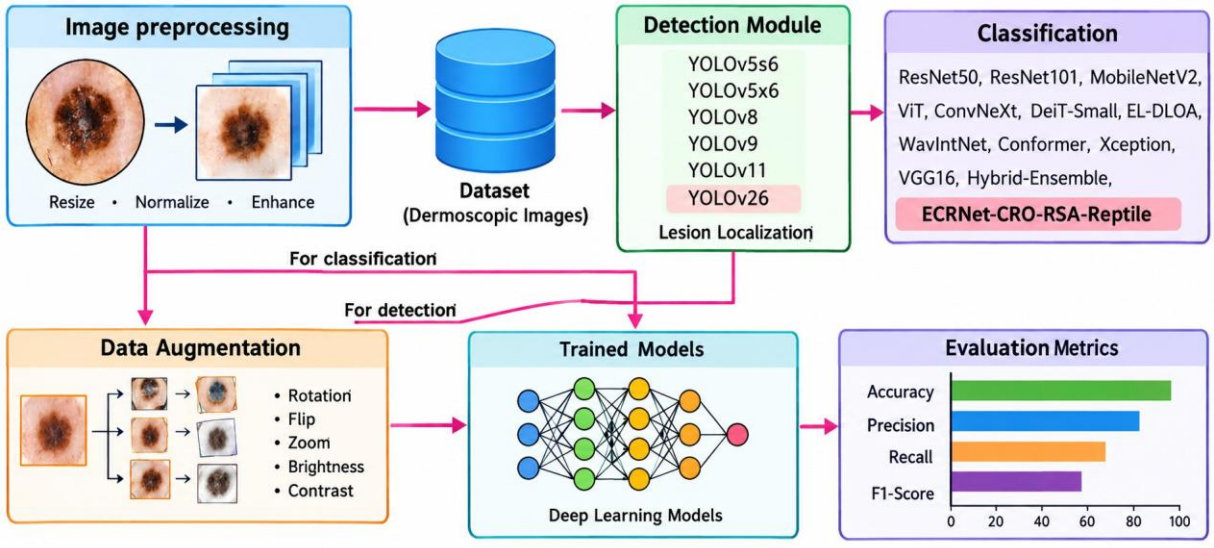

Fig. 1. Architecture of Explainable Skin Cancer Classification and Lesion Detection

The efficiency of the model and reliability of the diagnosis are estimated using the accuracy, precision, recall, F1-score, and mAP metrics.

3.1 Training and Testing

The dataset allocates 70% for training while 30% is assigned for testing. The training dataset helps to understand relationships and patterns in the data. Augmentation of images in the dataset is performed by transforming and resizing the images. The testing dataset evaluates the model's performance" based on the images that the model has not seen. Testing the model on the images ensures that the model has the ability to generalize and predict the outcomes for new data. It also ensures that the model provides accurate measurements.

3.2 Algorithms

3.2.1 Classification.

ResNet50: ResNet50 refers to a very deep convolutional neural network consisting of 50 layers that is mainly used for the purpose of classification of images. The residual connections in this model help solve the problem of vanishing gradients while training the deep architecture. [14] ResNet50 can effectively learn and memorize the features of the images in a hierarchical manner. ResNet50 can effectively learn the colors, textures, and structures of lesions from medical images. The model is developed using a residual learning framework that learns the features and text of the images efficiently.

ResNet101: The framework is like ResNet50 where it has 101 layers, and the ability to learn the features of the image is sophisticated. [15] It consists of skip connections where the architecture helps to concentrate and optimize the learning of features as the layers of the network become more distant. Learning the hierarchical image features allows for the convolutional network to detect complex patterns and anatomy in medical images.

MobileNet: MobileNet is the lightweight convolutional neural network that uses depthwise separable convolutions to decrease the computational cost of the network. [16]helps to minimize the parameters. This design helps the MobileNet to learn and extract the lesion variation in medical images.

ViT-Base: This architecture is designed with the self-attentive transformer networks, which is capable of understanding and capturing a context.[17] This framework is fitted with advanced contextual features and transformer text learning, which can help understand the structure and boundaries of lesions in dermoscopic images.

ConvNeXt-Tiny: ConvNeXt-Tiny is a lightweight deep convolutional neural network model, which efficiently classifies medical images based on discriminative feature learning of skin lesions. It has the ability to capture the essential aspects of the skin lesions, such as texture, shape, structure, and spatial characteristics, making the different categories of lesions distinguishable.

ConvNeXt-Base-768: ConvNeXt-Base-768 merges the convolutional model with deep feature learning and stable feature representation to address higher dimension images. It structures and classifies dermoscopic lesions that have deeper spatial and contextual characteristics and structures.

ConvNeXt-Base-1024: The ConvNeXt-Base-1024 architecture has a convolution-based design that supports feature representation of a large scale for deep learning of large images and spatial patterns. ConvNeXt-Base-1024 classifies and models dermoscopic images and lesions having deep structures, boundary patterns, and textures.[18]

DeiT-Small: Lightweight, deep, self-attentive architectures, such as DeiT-Small,classifies and integrates structures and textures of lesions having boundary patterns of deep, medical images using a deep and efficient classification framework. [19]

EL-DLOA: The EL-DLOA architecture was designed based on deep learning optimization for improved feature extraction and classification. The deep learning design and optimization algorithms in EL-DLOA enable the architecture to learn hierarchical spatial structures of dermoscopic lesions under different conditions.

WavIntNet: WavIntNet is a multi-level deep learning framework which is used to analyze medical images with high accuracy for feature discrimination and context understanding. WavIntNet is a multi-level deep learning framework utilizing wavelet features and hierarchical context learning to discriminate complex lesion patterns.[20] It can be used efficiently to analyze dermoscopic medical images.

Conformer: Conformer is a deep learning framework using both convolutional and transformer attention-based architectures. Convolution layers help in detecting the detailed spatial information, while transformer layers help in capturing context relationships. Conformer helps in improving representation learning and lesion discrimination for dermoscopic image classification.

Xception: Xception is a deep learning framework based on the new Convolution architecture called depthwise separable convolutions, which are computationally efficient. The framework can analyze dermoscopic images and classify lesions with various colors, textures, and structures.

VGG16: VGG16 is a deep convolutional neural network that has 16 layers that are weighted for image classification. This model is capable of extracting the hierarchical features through a series of convolutional operations, which help in acquiring both low level and high level visual features to be used in accurate image classification. It is also able to analyze the dermoscopic images and classify the lesions according to their shape and texture.

Hybrid-Ensemble: The proposed Hybrid-Ensemble is an architecture that uses deep learning models and several neural networks to classify the features of the dermoscopic images in a much more advanced and reliable way. The neural networks have different capabilities in context learning. This model greatly improves the process of predictive classification through deepening the collaborative features.

ECRNet-CRO-RSA-Reptile: ECRNet-CRO-RSA-Reptile is a hybrid deep learning architecture that uses the ECRNet and CRO, RSA, and Reptile optimization algorithms to improve the skin lesion classification. This hybrid deep learning architecture utilizes the adaptive learning and feature refinement strategies to improve the classification reliability and accuracy. It is capable of learning both the local lesion characteristics and global context of the dermoscopic images [21].

3.2.2 Detection.

YOLOv5s6: The YOLOv5s6 is a lightweight object detection architecture used for effective real-time lesion detection. It is a single-stage detection architecture used to predict bounding boxes and class probabilities together in one single network. The convolutional feature extraction is utilized to learn about the structure and spatial attributes of lesions and represent features at multiple levels.

YOLOv5x6: A variation of the YOLOv5s6 architecture is the YOLOv5x6 architecture. The architecture consists of more depth and large structures to focus on the advanced object

detection capabilities. Furthermore, the YOLOv5x6 architecture is able to localize lesions in the image through proper contextual dermoscopy image analysis of medical images.

YOLOv8: The YOLOv8 is a framework architecture designed to offer realtime object detection. Additionally, the architecture design uses the anchor free detection technique to accurately detect lesion regions in an image and fast processing capabilities.

YOLOv9: YOLOv9 is a highly sophisticated model for object detection that has been created to improve the process of lesion localization within dermoscopic images. In such a way, the architecture of the algorithm helps to improve the precision of detection and lesion localization since it learns spatial and contextual features of the input images. In this way, the model allows detecting regions with lesions and conducting the contextual analysis of abnormal skin patterns, which can be used to detect skin cancer.

YOLOv11: YOLOv11 is a modern object detection model that enables to perform the localization of lesions through efficient feature fusion and convolution. With the help of this model, multi-scale features are learned, and the improvement of skin lesions' detection and localization within dermoscopic images is provided. The framework is skilled in analyzing spatial patterns and lesions and also contains data about the context of the image.

YOLOv26: YOLOv26 is an object detection algorithm that has been developed with the capability of precisely localizing the skin lesions while efficiently learning the spatial features of the skin lesion. The YOLO framework integrates feature extraction at a deep level with the detection optimization algorithms for accurate detection of abnormalities. It has excellent skills in analyzing the complex dermoscopic lesions and its visual context.

4. EXPERIMENTAL SETUP & RESULTS

4.1 Dataset Collection

The dataset [22] contains 10,000 image files created for the specific purpose of skin cancer detection and classification. The collection contains various skin lesions used to create a deep learning model classification and detection framework. Each image contains supplemental annotations to provide high-quality ground truth labels for training. With a uniform format, the collection is a ready resource for developing cancer detection models.

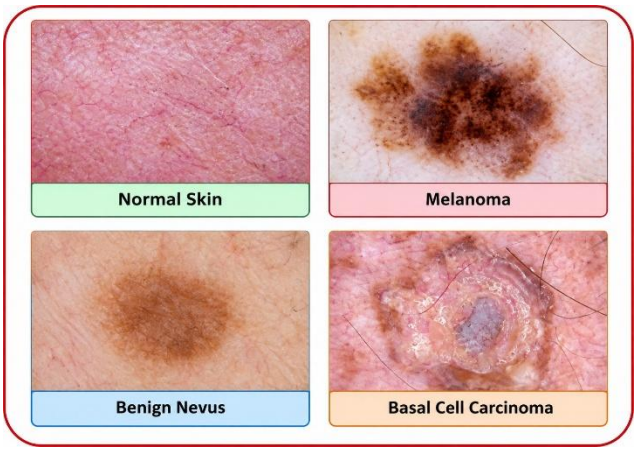

Fig. 2. Representative Images of Skin Lesions from the HAM10000 Dataset

4.2 Pre-processing

The preprocessing step is done in order to enhance the quality of the images and make the dataset ready for further analysis by using the deep learning techniques. As part of this preprocessing step, the dermoscopy images are scaled according to the dimensions required by the chosen architecture. Moreover, some image enhancement and data augmentation steps are taken in order to add diversity to the dataset and minimize the effect of any differences in lighting, orientation, and scaling of the images.

4.2.1 Processing for Classification.

The steps in processing data for classification include the application of data augmentation on the images. In this regard, the images are processed using the image data generator. The purpose of image data augmentation is to produce different images through rotation, flip, zoom, and shift among others. This will help the machine to identify robust and discriminatory features. The pixel intensities of the images are normalized to a range between 0 and 1. The Shear transformation slides the pixels within an image. The addition of variance improves the dataset. The ability of the model to generalize at different scales, Applying zoom transformations enables the model to effectively generalize across varying lesion scales.

The dataset is added to and varied by the horizontal flipping of images. Neural networks require all input images to be of the same dimension. All images are resized to this standard. The effect of all these processes improves the classification model.

4.2.2 Processing for Detection.

Different preprocessing steps are taken for detection. The images are converted to blob objects so that they conform to the model input. Rectangles that outline the objects of interest are drawn. These are called bounding boxes. Each of the bounding boxes is assigned a class by the predefined classification. The images are then converted to a NumPy array. During the training process, images and respective annotations, i.e. labels, are combined. The images are converted from BGR to RGB. The images are also resized to a standard input dimension. Data augmentation allows the images to vary and helps the model be more flexible. The model, then, becomes more accurate in detecting objects in complicated and diverse environments.

4.3 Metrics

Accuracy: Accuracy measures the overall proportion of correctly classified instances (both positive and negative) among all evaluated cases:

$$\text{Accuracy}=(TP+TN)/(TP+FP+TN+FN) \quad (1)$$

Precision: Precision determines the proportion of correctly identified positive detections relative to all samples predicted as positive:

$$\text{Precision}=(\text{True Positive})/(\text{True Positive}+\text{False Positive}) \quad (2)$$

Recall: Recall (sensitivity) represents the proportion of actual positive cases that were correctly identified by the model:

$$Recall=TP/(TP+FN) \quad (3)$$

F1-Score: The F1-score is the harmonic mean of precision and recall, balancing both false positives and false negatives:

$$F1\ Score=2*(Recall * Precision)/(Recall+Precision) \quad (4)$$

mAP: Mean Average Precision (mAP) evaluates overall detection performance by averaging the Average Precision (AP) across all n evaluated object categories:

$$mAP=1/n \sum_{k=1}^{k=n} AP_k \quad (5)$$

4.4 Detection Performance

Table 1. Performance Comparison of Skin Lesion Detection Models

Model	Precision	Recall	mAP
YoloV5s6	0.601	0.681	0.670
YoloV5x6	0.584	0.598	0.613
YoloV8	0.668	0.639	0.690
YoloV9	0.626	0.606	0.656
YoloV11	0.544	0.594	0.590
YoloV26	**0.739**	**0.642**	**0.713**

Table 1 shows the comparative performance of six YOLO based object detection models tested for skin lesion localization in terms of Precision, Recall and mean Average Precision (mAP). As can be seen from the comparison below there is a considerable difference between the evaluated architectures in their capacity to localize lesions and avoid false positive localization.

Among the evaluated models, YOLOV26 demonstrated the best Precision (0.739) and the best mAP (0.713). Therefore, YOLOV26 showed the best performance in localization in terms of two main indicators considered in this study. YOLOv8 demonstrated the second best mAP (0.690) and Precision (0.668), whereas YOLOv5s6 achieved mAP (0.670) and the best Recall (0.681). YOLOv9 showed Precision, Recall and mAP 0.626, 0.606 and 0.656 respectively. YOLOv5x6 had 0.584 Precision, 0.598 Recall and 0.613 mAP, while YOLOv11 had the lowest Precision (0.544) and mAP (0.590).

Moreover, the Precision value for YOLOV26 exceeded the one for YOLOv5s6 by 13.8 percentage points (from 60.1% to 73.9%). Also, the mAP for YOLOV26 was 4.3 percentage points greater than that for YOLOv5s6 (from 67.0% to 71.3%). The mentioned figures prove the fact that YOLOV26 has managed to find the best tradeoff between accuracy of the localization of the lesions and reduction of false positive lesion detections. Despite the fact that YOLOv5s6 had the largest Recall value of 68.1%, its Precision was significantly smaller (60.1%). Thus, YOLOv5s6 can identify a larger share of the target lesions, yet it produces a significant number of false positive detections.

Also, YOLOv8 proved to be able to produce good results of lesion detection; it provided a Precision of 66.8%, Recall of 63.9%, and mAP of 69.0%. Also, the mAP of YOLOv8 was just 2.3 percentage points less than that of YOLOV26. This fact proves the relatively high localization capabilities of YOLOv8. Still, YOLOV26 provides higher Precision and mAP simultaneously, so it should be considered the best detection model for the proposed framework.

The results of YOLOv9, YOLOv5x6, and YOLOv11 were comparatively lower across the principal detection measures. YOLOv9 produced a balanced but moderate performance, whereas YOLOv5x6 did not provide an advantage over the smaller YOLOv5s6 model in the reported evaluation. YOLOv11 recorded the lowest Precision and mAP among the six models. These observations demonstrate that increasing model scale or using a newer architecture does not automatically result in superior performance for the particular dermoscopic image evaluation considered in this study.

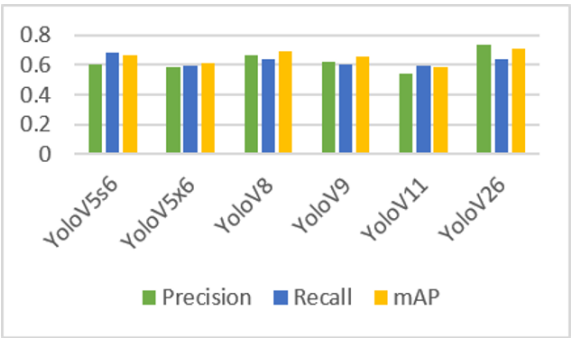

Fig. 3. Evaluation of YOLO Detection Performance Using Precision, Recall, and mAP

The graph presented in Figure 3 presents a graphical comparison between Precision, Recall, and mAP for all the six detection models. The graph further highlights the high values of Precision and mAP achieved using YOLOV26 compared to the others, while YOLOv5s6 obtained the highest value of Recall. Thus, model selection was not done on the basis of one metric. YOLOV26 was selected due to high values of Precision and mAP which are significant for localization of lesions and minimizing false positives.

Based on the results obtained from the experiments, YOLO2V6 has been integrated in the framework as the component for lesion localization. Nevertheless, these results are only relevant to the experimental setting used in the experiments and not proof of performance in clinical diagnosis. Clinical studies using independent data are necessary for validation.

4.5 Classification Performance

Table 2. Comparative Performance Evaluation of Deep Learning Models for Classification

Model	Accuracy	Precision	Recall	F1 Score
ResNet50	0.826	0.760	0.698	0.716
ResNet101	0.836	0.778	0.689	0.722
MobileNet	0.831	0.731	0.685	0.695
ViT-Base	0.674	0.386	0.342	0.348
ConvNeXt-Tiny	0.669	0.096	0.143	0.115
ConvNeXt-Base-768	0.669	0.096	0.143	0.115

ConvNeXt-Base-1024	0.650	0.306	0.256	0.236
DeiT-Small	0.720	0.489	0.332	0.340
EL-DLOA	0.704	0.433	0.233	0.242
WavIntNet	0.720	0.415	0.296	0.326
Conformer	0.734	0.539	0.364	0.400
Xception	0.967	0.950	0.932	0.940
VGG16	0.969	0.954	0.936	0.945
Hybrid-Ensemble	**0.972**	**0.954**	**0.937**	**0.945**
ECRNet-CRO-RSA-Reptile	0.949	0.890	0.889	0.888

Table 2 shows the classification performance of fifteen deep learning architectures with Accuracy, Precision, Recall, and F1-Score. As seen in the table, there is significant variation in the performance of the different models evaluated here. This comparison provides an empirical way to choose which architecture to use as the classification part of the framework proposed here.

The Hybrid-Ensemble model obtained the highest Accuracy of 97.2%, as well as Precision of 95.4%, Recall of 93.7%, and F1-Score of 94.5%. With all the four measures, it can be concluded that the Hybrid-Ensemble model had the best classification performance among the models tested here. Notably, the high Recall means that the model was able to predict most of the positive instances, while the high Precision means that most of those predictions were indeed true positives.

Of the various CNN architectures, VGG16 had an Accuracy of 96.9%, Precision of 95.4%, Recall of 93.6%, and F1-score of 94.5%. Xception proved to be another efficient architecture, delivering an Accuracy of 96.7%, Precision of 95.0%, Recall of 93.2%, and F1-score of 94.0%. The above figures reveal that VGG16 and Xception were the most efficient individual classification architectures in the presented comparison.

As a result of using the suggested Hybrid-Ensemble approach, the Accuracy increased by 14.6 percentage points compared with the ResNet50 model, which had 82.6% Accuracy. The Accuracy of the MobileNet model was 83.1%, and it is seen that the proposed solution delivered 14.1 percentage points better performance. ResNet101 achieved 83.6% Accuracy, and still, it was much lower than the 97.2% of the proposed Hybrid-Ensemble approach.

In the case of ECRNet-CRO-RSA-Reptile combination, the Accuracy, Precision, Recall, and F1 Score were recorded as 94.9%, 89.0%, 88.9%, and 88.8% respectively. Although this figure is below that of Hybrid-Ensemble, it is much above the Accuracy of ResNet50, ResNet101, and MobileNet in the said comparison. This demonstrates that the optimization method for ECRNet proved to be a good classifier and the Hybrid-Ensemble proposed gave best results.

On the other hand, the transformer-based architectures presented relatively lower results in the current experiment. The ViT-Base obtained Accuracy of 67.4%, whereas the DeiT-Small attained Accuracy of 72.0%. ConvNeXt models also provided relatively low results, with the ConvNeXt-Tiny and ConvNeXt-Base-768 attaining Accuracy of 66.9% and ConvNeXt-Base-1024 having Accuracy of 65.0%. In addition, EL-DLOA and WavIntNet attained Accuracy of 70.4% and 72.0%, respectively, while the Conformer got Accuracy of 73.4%.

The lower results of these models can be viewed in the light of the experimental setup and the dataset used in this experiment. These differences may be due to the features of the model, configuration of training, the form of the representation, and training data available. Therefore, the results do not imply that transformer-based architectures are inherently inferior to CNNs for skin-lesion classification.

Comparison of Precision, Recall and F1-score gives even more insight than Accuracy. For instance, while VGG16 and Hybrid-Ensemble had the same F1-score of 94.5%, the latter showed the highest Accuracy of 97.2%. Xception also had F1-score of 94.0%, which means that this configuration was performing almost as good as the top two. It is evident that Accuracy is not a complete metric for describing the work of the model, and Precision, Recall, and F1-score should be used together.

Figure 4 depicts the performance comparison of the fifteen configurations that were examined for classification. It is obvious from the graph that there are much higher values of Accuracy, Precision, Recall, and F1-score for Hybrid-Ensemble, VGG16, and Xception compared to some other configurations. These results prove that Hybrid-Ensemble is the best choice for classification in the proposed system.

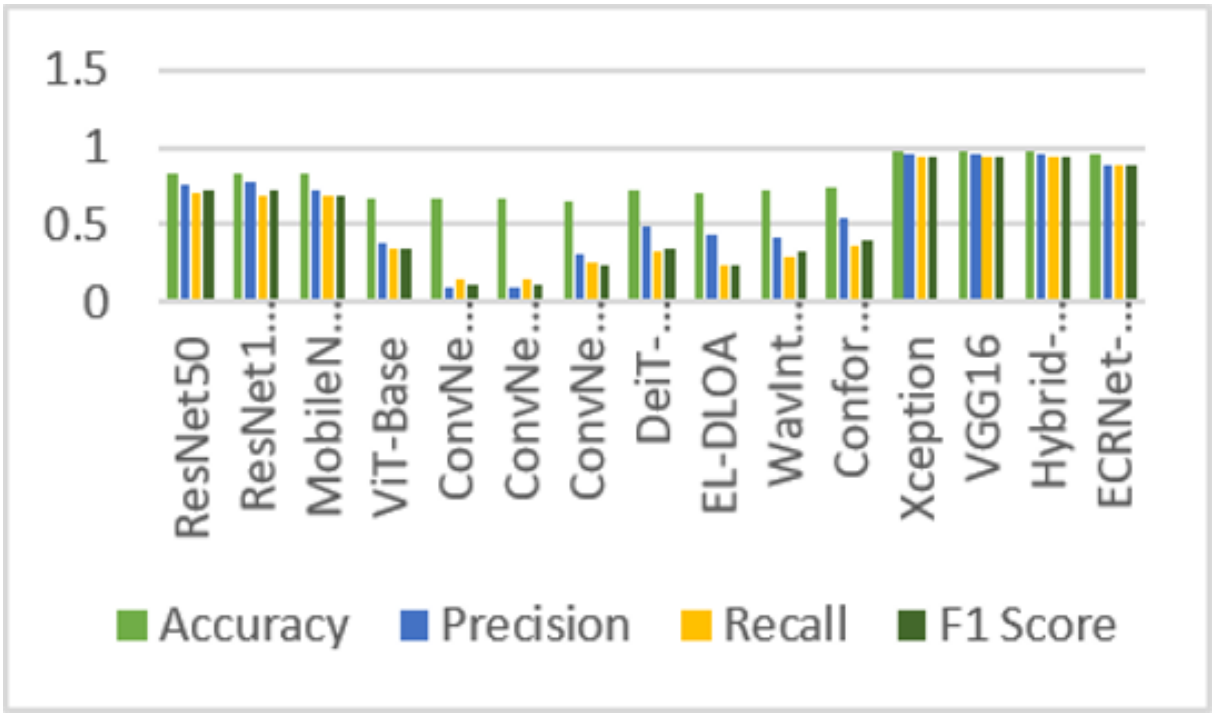

Fig. 4. Comparative Performance of Deep Learning Models Based on Accuracy, Precision, Recall, and F1-Score

Overall, the classification experiment demonstrates that the proposed Hybrid-Ensemble model achieved the strongest reported performance among the evaluated configurations. The results support the use of complementary feature representations within an ensemble-based architecture for the classification task. Nevertheless, the reported performance is based on the experimental dataset and testing protocol described in this study. External validation on independent datasets would be necessary to determine whether the observed performance generalizes to different patient populations, imaging devices, acquisition conditions, and clinical environments.

4.6 Web Application Validation

Fig. 5. User Interface for Skin Lesion Image Upload

The web application that was built was used as a comprehensive evaluation of the entire classification, lesion localization, and ex-

planation process. Figure 5 illustrates the image upload interface where the user uploads his or her dermoscopic image. This image is processed by the pre-processing module after which it is fed into the deep learning models.

Figure 6 depicts the prediction results along with the predicted lesion type, the confidence score, lesion bounding box, and Grad-CAM. The use of these elements helps to analyze the prediction both numerically and visually. The bounding box shows the spatial localization of the lesion, and the Grad-CAM heatmap shows the image areas which were significant for making the classification decision.

Grad-CAM plays an important role for the interpretability task of the proposed framework since instead of showing the predicted class label, the application gives the visual explanation of the areas which have influenced the prediction. In the given example, the highlighted area shows mainly the area of the lesion, providing an interpretable indication of the image features considered by the classification model.

The illustration shown in Figure 7 reveals that the consecutive dermoscopy images can be processed using the application without having to stop the application. The process of processing these images involves the same stages which include pre-processing, feature extraction, lesion localization, classification, and Grad-CAM visualization. This demonstrates the operational continuity of the prototype and its ability to process multiple images during a single application session.

The last step in the diagnostic process for an example with the category of Benign Keratosis is shown in Figure 8. The proposed interface includes all elements including the predicted class label, confidence measure, lesion localization, Grad-CAM visualization and recommendations into one output. This figure shows the combination of all these components to form a unified computer aided diagnosis interface.

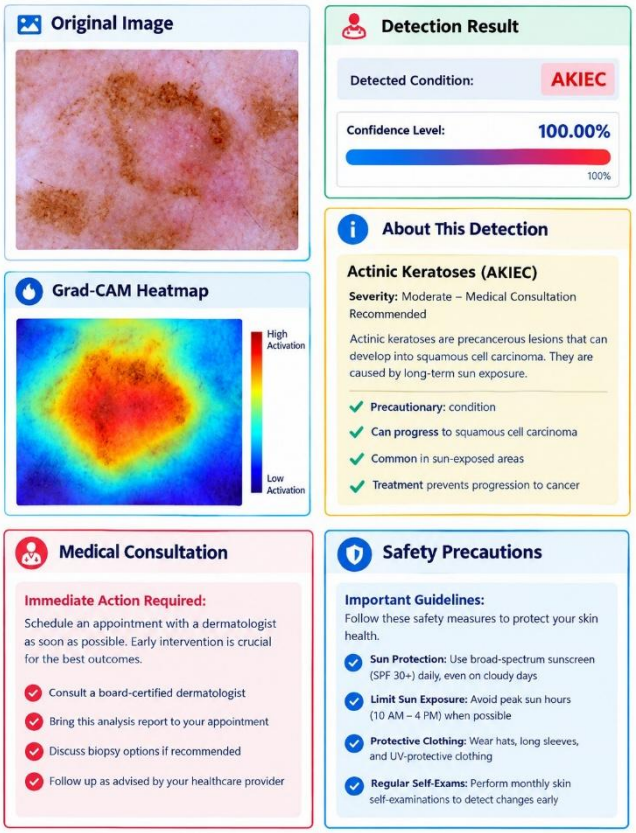

Fig. 6. Skin Lesion Classification Results with Grad-CAM Visualization

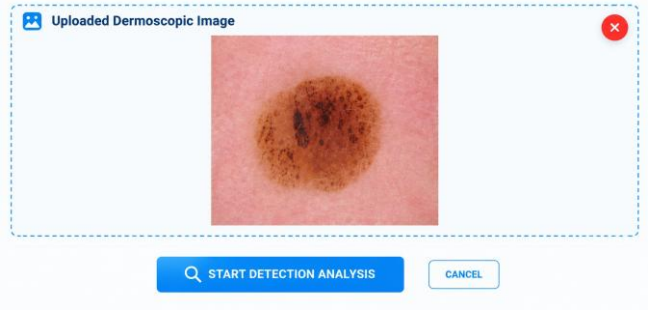

Fig. 7. Real-Time Skin Lesion Detection and Classification Interface

The web application therefore provides an implementation-level demonstration of the proposed workflow rather than an independent clinical validation. The application shows the feasibility of integrating classification, detection, and explainability into a single interface. However, the system should be regarded as a computer aided decision support prototype and not as a replacement for professional dermatological diagnosis. Clinical deployment would require further validation using independent datasets, diverse acquisition conditions, and prospective clinical studies.

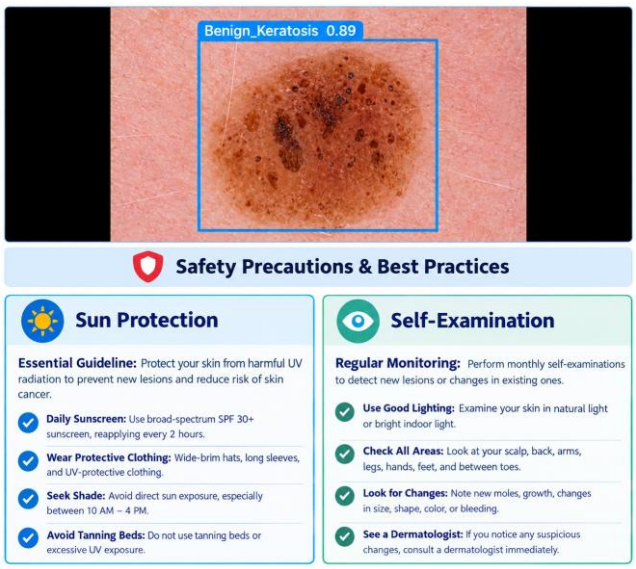

Fig. 8. Skin Lesion Diagnosis and Clinical Recommendation

5. CONCLUSION & FUTURE WORK

A deep learning framework for the identification and localization of skin cancer lesions was created by using a hybrid classification model and an advanced object detection model. The Hybrid-Ensemble model achieved an accuracy of 97.2%, precision of 95.4%, recall of 93.7%, and an F1-score of 94.5%, demonstrating the strongest overall classification performance among the evaluated models. Of the models tested, YOLOV26 was able to locate skin lesion abnormalities with the best precision of 73.9%, recall of 64.2%, and an mAP score of 71.3%. Grad-CAM was employed to highlight salient lesion regions on dermoscopic images, thereby explaining the model's predictions. The framework employed a Flask-based web app with an SQLite database to help secure image uploads and provide a user interface for preprocessing images and providing predictions. From the experiments, the framework demonstrated strong performance compared with the evaluated deep learning models for skin lesion classification and localization. The framework was designed to help dermatologists and provide an intelligent computer-aided diagnostic tool.

Future research will explore multimodal integration combining dermoscopy with patient demographics, clinical history, and genomic data to enhance diagnostic reliability. Federated learning can be potentially used for secure, privacy preserving, patient data protecting, collaborative learning in multiple hospitals. Deployments on mobile and embedded healthcare devices in remote settings can be facilitated using lightweight, embedded, and optimized Edge AI transformer architectures. Temporal analysis and 3D lesion reconstruction can further advance early melanoma detection. Future research can opt for self-supervised learning and few-shot learning to minimize reliance on large, highly annotated datasets. Additionally, future work will incorporate explainable AI into automated clinical report generation and integrate with telehealth decision support systems.

6. REFERENCES

[1] Halil Murat U¨ nver and Enes Ayan. Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm. *Diagnostics*, 9(3):72, 2019.

[2] DD Yang, JD Salciccioli, DC Marshall, A Sheri, and J Shal-houb. Trends in malignant melanoma mortality in 31 countries from 1985 to 2015. *British Journal of Dermatology*, 183(6):1056–1064, 2020.

[3] Li Hu, SF Jin, Ling Chen, and YL Wang. Trends in the incidence and mortality of cutaneous melanoma in hong kong between 1983 and 2015. *Int J Clin Exp Med*, 11(8):8259–8266, 2018.

[4] Alfonso Baldi, Marco Quartulli, Raffaele Murace, Emanuele Dragonetti, Mario Manganaro, Oscar Guerra, and Stefano Bizzi. Automated dermoscopy image analysis of pigmented skin lesions. *Cancers*, 2(2):262–273, 2010.

[5] Yuexiang Li and Linlin Shen. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors*, 18(2):556, 2018.

[6] RM Lucas, S Yazar, AR Young, M Norval, FR De Gruijl, Y Takizawa, LE Rhodes, CA Sinclair, and RE Neale. Human health in relation to exposure to solar ultraviolet radiation under changing stratospheric ozone and climate. *Photochemical & Photobiological Sciences*, 18(3):641–680, 2019.

[7] Arthur AM Teodoro, Douglas H Silva, Renata L Rosa, Muhammad Saadi, Lunchakorn Wuttisittikulkij, Rao Asad Mumtaz, and Demostenes Z Rodriguez. A skin cancer classification approach using gan and roi-based attention mechanism. *Journal of Signal Processing Systems*, 95(2):211–224, 2023.

[8] Chao Xin, Zhifang Liu, Keyu Zhao, Linlin Miao, Yizhao Ma, Xiaoxia Zhu, Qiongyan Zhou, Songting Wang, Lingzhi Li, Feng Yang, et al. An improved transformer network for skin cancer classification. *Computers in Biology and Medicine*, 149:105939, 2022.

[9] Soumyya Kanti Datta, Mohammad Abuzar Shaikh, Sargur N Srihari, and Mingchen Gao. Soft attention improves skin cancer classification performance. In *International Workshop on Interpretability of Machine Intelligence in Medical Image Computing*, pages 13–23. Springer, 2021.

[10] Md Shahin Ali, Md Sipon Miah, Jahurul Haque, Md Mahbubur Rahman, and Md Khairul Islam. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Machine Learning with Applications*, 5:100036, 2021.

[11] Luzhou Liu, Xiaoxia Zhang, and Zhinan Xu. An adaptive weight search method based on the grey wolf optimizer algorithm for skin lesion ensemble classification. *International Journal of Imaging Systems and Technology*, 34(2):e23049, 2024.

[12] Xin Shi, Xiaofang Dong, Shuhua Ye, Wei Li, and Haiyan Li. Wavelet integrated multiscale feature fusion network for imbalanced skin lesion classification. 2024.

[13] Sobia Bibi, Muhammad Attique Khan, Jamal Hussain Shah, Robertas Damasˇevicˇius, Areej Alasiry, Mehrez Marzougui, Majed Alhaisoni, and Anum Masood. Msrnet: Multiclass skin lesion recognition using additional residual block based fine-tuned deep models information fusion and best feature selection. *Diagnostics*, 13(19):3063, 2023.

[14] Veena Dillshad, Muhammad Attique Khan, Muhammad Nazir, Oumaima Saidani, Nazik Alturki, and Seifedine Kadry. D2lfs2net: Multi-class skin lesion diagnosis using deep learning and variance-controlled marine predator optimisation: An application for precision medicine. *CAAI Transactions on Intelligence Technology*, 10(1):207–222, 2025.

[15] Muneezah Hussain, Muhammad Attique Khan, Robertas Damasˇevicˇius, Areej Alasiry, Mehrez Marzougui, Majed Alhaisoni, and Anum Masood. Skinnet-inio: multiclass skin lesion localization and classification using fusion-assisted deep neural networks and improved nature-inspired optimization algorithm. *Diagnostics*, 13(18):2869, 2023.

[16] Ra Karthik, Tejas Sunil Vaichole, Sanika Kiran Kulkarni, Ojaswa Yadav, and Faiz Khan. Eff2net: An efficient channel attention-based convolutional neural network for skin disease classification. *Biomedical Signal Processing and Control*, 73:103406, 2022.

[17] Mehak Arshad, Muhammad Attique Khan, Usman Tariq, Ammar Armghan, Fayadh Alenezi, Muhammad Younus Javed, Shabnam Mohamed Aslam, and Seifedine Kadry. A computer-aided diagnosis system using deep learning for multiclass skin lesion classification. *Computational intelligence and neuroscience*, 2021(1):9619079, 2021.

[18] T Fikrle and K Pizinger. Digital computer analysis of dermatoscopical images of 260 melanocytic skin lesions; perimeter/area ratio for the differentiation between malignant melanomas and melanocytic nevi. *Journal of the European Academy of Dermatology and Venereology*, 21(1):48–55, 2007.

[19] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 168–172. IEEE, 2018.

[20] Lucia Ballerini, Robert B Fisher, Ben Aldridge, and Jonathan Rees. A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions. In *Color medical image analysis*, pages 63–86. Springer, Berlin, Heidelberg, 2013.

[21] Yue Cheng, Ragavendar Swamisai, Scott E Umbaugh, Randy H Moss, William V Stoecker, Saritha Teegala, and Subhashini K Srinivasan. Skin lesion classification using relative color features. *Skin Research and Technology*, 14(1):53–64, 2008.

[22] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermato-scopic images of common pigmented skin lesions. *Scientific data*, 5(1):180161, 2018.