

Deep Learning Framework For Skin Disease Detection Using Dynamic Filter, SAM Segmentation, YOLO Detection And Efficientnet-B5

Kavitha B¹, Dr. Kusuma Kumari B M²

¹Research Scholar, Department of Studies and Research in Computer Applications, Tumkur University, Tumakuru, Karnataka

²Assistant Professor, Department of Studies and Research in Computer Applications, Tumkur University, Tumakuru, Karnataka

¹Orcid ID: [0009-0006-1042-5514], ²Orcid ID: [0000-0003-2961-7752]

ABSTRACT

The reliable diagnosis and clinical decision making can highly depend on the early and accurate detection of the affected regions in medical images. The article displays a comprehensive Image Enhancement, Segmentation and Classification using publicly available datasets with the help of deep learning approaches. Our proposed framework takes large amounts of data i.e. around 10000 images. The scale of the dataset can increase this period. The first action in the workflow. Next, three different filters are applied and the best filter is selected using four quantitative image evaluation metrics and they are Peak Signal to noise Ratio (PSNR), Structural Similarity Index Measurement (SSIM), Mean Square Error (MSE), Edge Preserving Index (EPI). The improved visuals are then supplied to Segment Anything Model (SAM), which very precisely generates masks of affected region and retrieves the region in a very robust automatically. The masked results are further processed and compiled as YOLO (You Only Look Once) inputs to generated only affected region. For classification task, efficientnet-b5 deep learning architecture is used to deal with spatial complexities. The performance is evaluated based on accuracy (96.45%), precision (96.26), recall (96.17), and f1-score (96.19). Scripting language employed is python.

Keywords: Skin disease, HAM10000, dynamic filter, SAM, YOLO, classification.

1. INTRODUCTION

One of the largest sense organs in the human body is the skin. An area of around 1.5-2 square meters may be covered by it. The skin's main job is to protect the body from the elements. It has many important functions, including protection from germs, balancing body temperature, feeling, and immune system. The skin is the first thing to come into contact with sunlight, pollution, chemicals, and microorganisms, which makes the skin vulnerable to many types of diseases and disorders. Skin disorders are a very common global health issue and affect people of all ages. Skin disorders can lead to serious health issues and may even become life-threatening if not treated in time.

The World Health Organization (WHO) states that skin diseases are often a common global problem and skin cancer has become a threatening disease in many regions in the world. WHO stresses improvement in awareness, screening, and diagnosis to reduce the impact of skin disorders. The magnitude of the problem of malignant skin diseases like melanoma is that late diagnoses mean that the cancer has spread to other parts of the body. Therefore, the early identification and management of skin lesions improve the survival of patients and lessens the burden on the healthcare system.

There are many various kinds of skin disease that are studied in dermatology. The most fatal and aggressive of them is the melanoma (MEL) skin disease. Some of the others studied are basal cell carcinoma (BCC) which is the most typical form of skin disease, benign keratosis-like lesions (BKL), actinic keratosis and intraepithelial carcinoma (AKIEC) which are precancerous that can lead to cancers, dermatofibroma (DF), vascular

lesions (VASC), and the melanocytic nevi (NV) which are moles. Some of these lesions may look like others that are cancerous and can mimic other that is why diagnosing may take a lot of time and experience.

Advances in artificial intelligence and deep learning have had a significant effect on medical imaging's Computer Aided Diagnostics (CAD) systems. CAD systems have proven to aid clinicians in the detection of lesions and abnormalities in medical images, reliably and with great precision. A popular example of medical imaging datasets is the HAM10000 dataset which is made up of over 10,000 dermatological images taken from a multitude of cross-sectional, clinical-based populations. The dataset houses images of 7 categories of skin lesions including melanoma, basal cell carcinoma, benign keratosis, dermatofibroma, vascular, actinic keratosis, and melanocytic nevi. The diversity and volume of the HAM10000 dataset have privileged it as a benchmark dataset for training and testing the classification of skin diseases using machine learning.

Developing automated skin disease detection means improving both precision of diagnosis and decreasing the amount of work dermatologists have to do. Automated systems help to diagnose skin lesions as either cancerous or non-cancerous. This is done by employing sophisticated algorithms with deep learning models together with seamless image processing and lesion segmentation. This technique is also useful to underdeveloped countries with poor access to dermatological help. The objective of this research is to create a system that can intelligently read dermatoscopic images of skin lesions from the HAM10000 dataset. The proposed method aims to improve patient care by enhancing the identification and categorization of skin lesions, which in turn helps doctors make more informed decisions.

Contribution of paper

1. Compared three filters for the pre-processing step and designed a dynamic filter method where best filter for every image is detected. Best filter is opted by considering metrics such as PSNR, SSIM, EPI and MSE
2. Detection is improved by removing hair from the image. Binary masks and lesion region is extracted using Segment anything model (SAM)
3. YOLO is provided with binary masks and lesion segmented parts, it segments only affected area of skin

Organization of paper

This section provides bird eye view of the paper, section I describes about importance of skin, diseases related to skin and WHO guidelines. Section II provides details about research carried on skin disease detection by various research scholars. Section III briefs about methodology employed in detection of skin diseases. Section IV describes about result screenshot obtained during implementation. Lastly conclusion for current research is provided along with reference.

2. LITERATURE SURVEY

The last decade has seen rapid developments of methods for dermoscopic images to detect different skin diseases. The implementation of deep learning method has helped detect skin diseases more rapidly and more accurately. The combination of the HAM10000 and other large public datasets has helped researchers develop more intelligent diagnostic tools for classifying skin lesions.

Alam et al. (2022) focused the HAM10000 dataset's class imbalance by developing a deep learning-based classifier. According to his research, boosting the identification accuracy of underrepresented classes requires a mix of data augmentation and balancing strategies.[1]

Behara et al. (2024) applied a hybrid deep learning framework that combines the interpretability enhancing, explainable artificial intelligence approach, Grad-CAM, tools to improve the skin cancer detection systems. [2]

In a review of the various deep learning techniques for segmentation and classification of skin lesions, Debelee et al. (2023) reported that combining segmentation and classification tasks led to better results. [3]

The HAM10000 dataset spawned multiple studies on various deep learning models for multi-class skin lesion detection. For example, Gessert et al. (2020) proposed an ensemble approach for dermoscopic image classification that combined the models EfficientNet, SENet, and ResNet. Their ensemble model achieved a significant classification improvement over the individual models. [4]

The classification of limited medical datasets using transfer learning became a hot topic of study. By comparing several transfer learning methods, Jain et al. (2021) assessed the HAM10000 dataset's ability to classify skin lesion types: ResNet50, DenseNet, MobileNet, and VGGNet. Their work showed that transfer learning approaches are superior to traditional machine learning approaches in the classification of skin lesions. [5]

Kassem et al. (2021) researched the use of several ML and DL methods to the problem of skin lesion identification and classification. When it comes to analysing dermoscopic images for the purpose of melanoma and other skin lesion detection, the study found that deep convolutional neural networks performed the best. [6]

Kreouzi et al. (2024) illustrated that by means of automated imaging analysis, artificial intelligence can empower clinicians to diagnose melanoma at an earlier stage and thus have a more positive impact on the survival of patients. [7]

Later, the focus of the research in this area shifted to hybrid deep learning and explainable AI methods [8].

Raghavendra et al. (2023) developed an improved model for skin lesion detection and classification using deep convolutional neural networks, outperforming more traditional approaches that rely on transfer learning. Their study proved that medical image categorization might benefit from specific CNN architectures. [9]

With the use of the HAM10000 dataset, to improve the accuracy of skin lesion categorization, Shetty et al. (2022) used specific image preprocessing and data augmentation methods. [10]

There was a significant advance in skin lesion studies with Tschandl et al. and the HAM10000 dataset. It offered over 10,000 dermatoscopic images of pigmented skin lesions from various clinics and spanning seven categories of skin diseases. It quickly became a major benchmark dataset for the training and evaluation of skin cancer detection models. [11]

Numerous studies appeared in 2022 with the aim of improving classification accuracy with the use of novel preprocessing techniques, data augmentation, and segmentation methods [12].

Alotaibi et al. proposed a computationally efficient and accurate convolutional neural network framework that detects melanoma based on the HAM10000 dataset. [13]

In the same vein, Akinrinade et al. (2025) addressed the significance of deep learning models in improving early skin cancer diagnosis, especially in areas with limited access to dermatologists [14]

The rapid improvement of deep learning techniques, especially Convolutional Neural Networks, has automated the skin lesion analysis. The ability of deep learning techniques to learn from medical images has shifted focus to automated analysis. Better results were achieved by employing the use of VGGNet, ResNet, and Inception.

Between 2024 and 2025, researchers primarily focused on using ensemble approaches, transformer-based frameworks, and multimodal deep learning architectures. In order to detect complex visual patterns in images of skin lesions, these advanced techniques combine attention processes with convolutional neural networks and transformers. Other recent studies have focused on multimodal and transformer-based methodologies. These studies use a combination of dermoscopic images and clinical metadata to improve accuracy in diagnosis. The research demonstrated that the multimodal models are more effective than conventional methods based on CNNs for classifying skin lesions.

3. METHODOLOGY

The goal of the proposed methodology is to create an automated system for classifying skin lesions based on dermoscopic images. This framework combines image preprocessing, lesion segmentation, and classification through deep learning, improving skin disease detection accuracy. Overall, the workflow includes acquisition and organization of the dataset, image preprocessing and filtering, artifact removal, lesion segmentation via advanced models, and classification using deep learning. Proposed steps of model are summarised in below Fig.1.

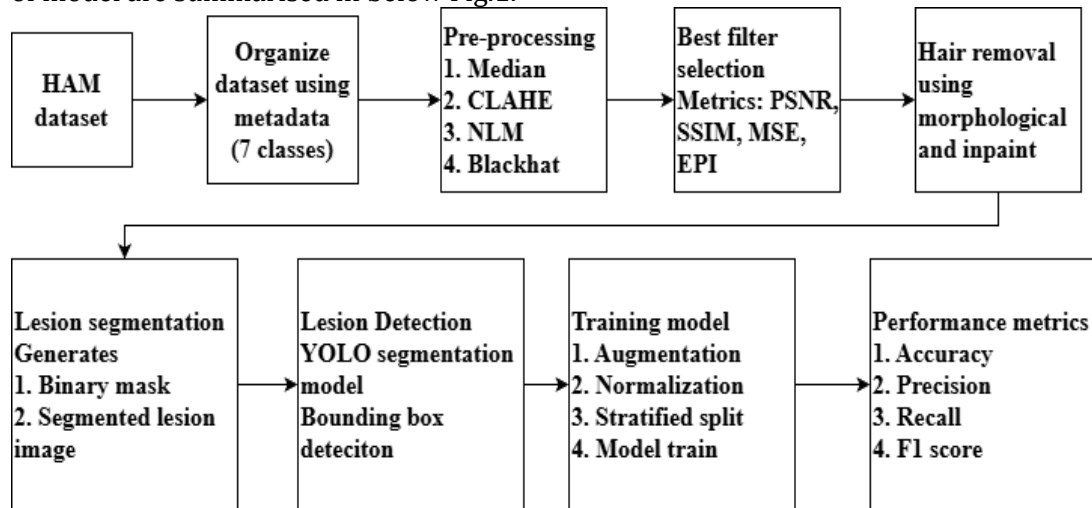


Fig. 1: System architecture of proposed model

A. Collection of Dataset

First step is to collect the data for any AI based research. In our research we are using the HMA10000 as the dataset.

Dermoscopic images used in the analysis have been obtained from the Internet Archive with permission from Kaggle. Dermoscopy images of pigmented skin lesions from more than 10,000 different clinical locations makes up the HAM10000 collection. There is information included with the dataset that describes the lesion kind and diagnosis for each image.

The dataset's images of skin lesions are classified into the following seven categories:

- Melanoma (mel) – an aggressive form of skin cancer.
- Melanocytic Nevi (nv) – benign lesions of the skin, commonly referred to as moles.
- Basal Cell Carcinoma (bcc) – skin cancer that is the most frequently found.
- Benign Keratosis-like Lesions (bkl)
- Actinic Keratosis / Intraepithelial Carcinoma (akiec)
- Dermatofibroma (df)
- Vascular Lesions (vasc)

The inclusion of images from both debilitating and non-debilitating dermatological conditions renders the dataset ideal for multi-class classification challenges in the analysis of dermatology.

B. Organization of dataset

When the dataset was received from Kaggle, the images were first organized into class-specific folders by diagnosis, which were received from the metadata file. Each image was classified into one of the seven classes based upon the metadata file. This was done to prepare the dataset for supervised learning algorithms.

The images were organized into seven folders, each corresponding to the seven categories of lesions. This made it possible to load images quickly when training the model, ensuring that each class was able to be processed separately during the preprocessing and segmentation of images.

Procedure for placing images into their respective folders is explained below,

- The first step taken after getting access to the HAM10000 dataset uploaded on Kaggle was to sort the images based on the details provided in the metadata file (HAM10000_metadata.csv). The metadata file details important attributes for each dermoscopic image which include image_id, lesion_id, diagnosis (dx), age, sex, and localization. Of the attributes listed above, the diagnosis label (dx) was useful in sorting the dataset for classification.
- The first step in the above process was to sort the metadata file using a data processing library such as Pandas. From the sorted file, the image_id and dx columns were extracted. The dx column indicates the disease type for each image. With this information, seven different folders were made for each of the seven classes of skin lesions; mel (melanoma), nv (melanocytic nevi), bcc (basal cell carcinoma), bkl (benign keratosis-like lesions), akiec (actinic keratosis), df (dermatofibroma) and vasc (vascular lesions).
- Finally, each image in the metadata file was compared with the images present in the dataset folders. The program retrieved the image using the image_id and sorted it into one of the class folders as dictated by the dx label. This helped to ensure that all images categorized as the same disease were sequestered into one folder.
- The dataset is facilitated for loading and helps in the supervised training of models for skin lesion classification through the structuring of images into seven class-wise folders using the metadata labels.

C. Pre-processing

Dermoscopic Imaging involves many distractions, varying lighting, and additional obstacles that can diminish how accurate an image classification can be. To remedy these problems, multiple filters were implemented to increase image quality prior to classifying and segmenting the images. There were four filters that were examined:

- Median Filters – used to remove noise and still keep the edges of images intact.
- Contrast Limited Adaptive Histogram Equalization (CLAHE) – increases the local contrast, and allows for better viewing of the lesion structure.
- Non-Local Means Denoising (NLMD) – decreases noise and keep the dermoscopic images texture vital
- Morphological Black-Hat Filters – improves the look of lesion boundaries and highlights dark features (hairs).

The filters were used on the images, and their success was measured using multiple quality metrics such as MSE, PSNR, SSIM, EPI. Equations for metrics is provided below,

$$\text{PSNR} = 10 \log_{10} \frac{\text{Highest pixel intensity}^2}{\text{MSE between original and filtered image}} \quad [15]$$

$$\begin{aligned} \text{MSE} &= \frac{1}{\text{image dimension (m \times n)}} \sum_{i=1}^m \sum_{j=1}^n (\text{Original image}(i, j) - \text{Filtered image}(i, j))^2 \quad [15] \\ \text{SSIM (i, j)} &= \frac{(2\mu_i\mu_j + C_1)(2\sigma_{xy} + C_2)}{(\mu_i^2 + \mu_j^2 + C_1)(\sigma_i^2 + \sigma_j^2 + C_2)} \quad [15] \\ \text{EPI} &= \frac{\sum_{i=1}^N \nabla I_o^{(i)} \cdot \nabla I_p^{(i)}}{\sqrt{\sum_{i=1}^N (\nabla I_o^{(i)})^2} \cdot \sqrt{\sum_{i=1}^N (\nabla I_p^{(i)})^2}} \quad [15] \end{aligned} \quad (1)$$

There was a total value calculated, to identify which quality filter was the highest ranked. The filter that produced the most quality was chosen to be the quality filter for the rest of the images. Combined score calculation is provided below, based on the highest score image from that filter is saved for next process.

$$\text{Combined score for every filter} = (0.3 * \text{PSNR}) + (0.3 * \text{SSIM}) + (0.3 * \text{EPI}) - (0.1 * \text{MSE}) \quad (2)$$

Above formula ensures that reduction in noise, preservation of structure and edge. If there is excessive distortion then it is subtracted (-0.1 * MSE). For example, real time best filter selection calculation is as shown in below Fig.2. for respective image median filter is providing good results compared to other filters. Hence Median filter is chosen as best filter for it. In the same manner for rest of images different filters are selected based on score. Instead of saving four images from four filter only one best filtered image is chosen which reduces storage burden. As choosing filter is dynamic in nature for every image, hence we call it as dynamic filtering process.

class	image	filter	PSNR	MSE	SSIM	EPI	score
mel	ISIC_0028	median	34.5811	22.6449	0.87951	0.592693	8.5515
mel	ISIC_0028	clahe	24.95038	207.9897	0.797045	0.917646	-12.7995
mel	ISIC_0028	nlm	34.45201	23.32811	0.805884	0.671648	8.44605
mel	ISIC_0028	blackhat	4.852118	21274.95	0.004753	-0.99008	-2126.34
mel	ISIC_0028	BEST					median

Fig. 2 Selection of best filter based on score

- Justification for choosing weight to PSNR, SSIM, EPI and MSE

The suggested approach for the analysis of different filtering techniques has divided image quality figures of merits. As each figure of merit expresses a different quality aspect of the image, a weighted scoring approach has been adopted in order to quantify their relative importance. The weights of 0.3, 0.3, 0.3, and 0.1 have been used to reflect the most critical features necessary to assess dermoscopic images.

First, PSNR (Peak Signal to Noise Ratio) predisposes the level of noise that has been eliminated while retaining the image quality. In medical image processing, retaining original information is extremely important. Pattern analysis is extremely critical, and excessive filtering is often avoided; thus, a weight of 0.3 is assigned here.

Second, SSIM (Structural Similarity Index) assesses the degree of likeness in structure of the original image to that of the filtered version. Segmentation and classification are primarily focused on retaining structural parts, so the EPI score is assigned a weight of 0.3. Dermoscopic images contain important structural features, such as lesions, that assist in the identification of diseases and facilitate accurate disease diagnosis.

Edge Preservation Index (EPI) quantifies how well the filtering process works in preserving edge details within the image. Edges are very important for skin lesion analysis, as the borders of a lesion in comparison to the normal skin encapsulate important detail in the severity and type of the disease. Segmentation models are

able to accurately locate lesion borders when edge preservation is accurate. For these reasons, EPI is given a weight of 0.3, making it important for the edges of the lesions to be preserved after filtering.

Mean Squared Error (MSE) quantifies the differences, at the pixel level, of the original and filtered images. While MSE provides some important information regarding the level of distortion in the image, it is not a good indicator of how good or bad the perceptual quality of the image is. For instance, perceptually an image may be better and on an MSE score it may achieve a higher score compared to a different image with lower quality structural features, which results in a lower MSE. MSE is therefore assigned a very small weight of 0.1 and is used as a penalty term in the scoring function.

The scoring function made to describe a filter's performance optimal to noise reduction, structure, and lesion boundary preservation, which are most important in skin lesion segmentation and classification, by giving more weight to PSNR, SSIM, and Edge preservation, and less weight to MSE. This balanced weighting strategy helps to choose a filter that optimally enhances image quality without obscuring clinically significant lesion characteristics.

D. Hair removal

Dermatoscopic images sometimes include hair structures that impact segmentation and classification of lesions. To address this issue, a hair removal algorithm using morphological operations was implemented.

The algorithm consisted of hair pixel value detection via black-hat morphological filtering and subsequent inpainting to restore the modified image areas. The approach effectively removed hair artifacts while keeping important lesion features, leading to segmentation and analysis images with more clarity.

Equations for black hat and closing morphological operations,

$$Black_{hat}^{Input\ image} = Input\ images * Structuring\ element - Input\ image \quad (3)$$

$$I.S = (Input\ image \oplus Structuring\ element) \ominus Structuring\ element \quad (4)$$

Where $\oplus$ represents dilation operation and $\ominus$ represents erosion operation

Closing operation on image: it performs dilation and erosion morphological operation on image. Equation for it is as shown, these operations is going to highlight dark structure in input image (usually hair).

After the Blackhat morphological operation we are going to generate mask which detects hair pixels in image. To perform this, we need a threshold. Had chosen threshold as 10 because if zero is chosen it detects white pixels, if higher value of 40 removes important features from image. Less than 10 may detect small contrast pixels. With trial and error method have chosen the optimum threshold value as 10. Detected hair region pixel is marked as one. Equation for hair mask generation is given below,

$$M(x,y) = \begin{cases} 1 & \text{if } Blackhat(x,y) > Threshold \\ 0 & \text{for other values} \end{cases} \quad (5)$$

Inpainting method is employed to remove hairs by using procedure stated here; after generating the hair mask, the hair pixels are erased by an image inpainting algorithm that uses the surrounding pixels of the detected regions to fill in the gaps using interpolation. Below equation states in simpler way

$$I_{clean}(x,y) = Inpaint\ to\ (I(x,y) \times M(x,y)) \quad (6)$$

Where, I_{clean} it is the image where no or less traces of hairs are present, I is the input image and M is the masked hair region.

E. Lesion segmentation using Segment any Model (SAM)

Post preprocessing and depilation, lesion segmentation was conducted to separate the area of interest from the surrounding dermal tissue. Good segmentation assists the classifier to concentrate on the lesion area and minimizes the effect of background noise. For this study, segmentation was done using the Segment Anything Model (SAM). SAM is a very robust foundation model for producing segmentation masks for a wide range of objects in images. A centre prompt was provided to direct the model to the lesion area for each dermoscopic image.

This process of segmentation generated two outputs and these outputs were fed into the analysis and classification purpose.

1. Binary lesion masks – indicating the precise position of the lesion.
2. Segmented color lesion images – consisting of just the lesion area and the background skin was attenuated.

F. Using YOLO lesion detection

To improve the identification and localisation of lesions, we also considered an object detection method using YOLO. It is an algorithm that performs object detection and provides object location in real time with a high degree of accuracy.

The YOLO model was provided with training data that included annotated and segmented images of lesions. The model was able to learn the dermoscopic images and to draw bounding boxes around areas that warranted further investigation. This process assisted in the refinement of localisation of lesions and provided the classification model with concentrated areas of input. Steps for implementing yolo is provided below,

- Dataset preparation as per YOLO acceptance format
- Validation and training splitting of images
- Preparation of YAML format for routing dataset
- Model training

Dataset preparation: YOLO needs image and annotation files telling where objects are for each individual image. The lesion regions were identified after preprocessing and segmentation and were converted into bounding boxes for YOLO training. For every dermoscopic image, there was a text file with bounding box coordinates for the lesion. These annotation files were saved in .txt format and complied with the standard YOLO annotation format.

The YOLO annotation format is given in Tab.1.

Table 1 Annotation format as per YOLO

ID which specifies class It is the numerical label which gives corresponding lesion class	Centre at x axis Centre at x and y axis is centre of bounding box.	Centre at y axis	Width & Height It is going to provide normalized height and width of boundary box of respective image
--	---	------------------	--

All values for coordinates are normalised between 0 and 1. This helps the model to generalise across images that are of different sizes. Bounding boxes are created automatically from the lesion segmentation masks created from the segmentation stage. These are calculated to be the minimum rectangle that contains the segmented lesion region and are converted to YOLO bounding box coordinates.

When training YOLO, the training files need to be organised in a certain way. The files for this dataset were split into two separate folders. The folders contain the images, and the other for the labels that annotate the images. The following shows how the dataset will be organised for YOLO training. For each image file, there is one annotation file that has the same name, but ends with .txt as shown in Fig.3.

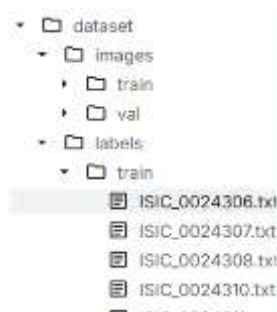


Fig. 3 Preparation structure for YOLO dataset

Validation and training split: In order to analyse how the YOLO model works, the data set has to be split into two sets; one for training, and one for validation.

Classifying photos is done via stratified split. To enable the YOLO model to learn, there is a training set, and a validation set to check the model's performance and prevent it from overtraining.

YOLO YAML file generation: In order to use YOLO, you need to create a YAML configuration file that details where the dataset is located, where the training and validation directories, and what the class labels present. This YAML configuration file is necessary for informing the YOLO model training information. Sample YAML configurations is shown below Fig.4.:

```
path: /kaggle/input/datasets/amoghabandrikalli/skin-yono-ready/dataset
train: images/train
val: images/val

names:
  0: MEL
  1: NV
  2: BCC
  3: AKIEC
  4: BKL
  5: DF
  6: VASC
```

Fig. 4 YAML file preparation

Training the model: Once the dataset and the YAML configuration file were finished, the YOLO model was trained on the dataset that was created. The YOLO architecture analyses and processes the entire image in one go. The model predicts bounding boxes and provides probabilities of classes for the identified objects. For better efficiency, the YOLO model was trained using pre-trained weights. Thanks to transfer learning, the model is also able to use learned visual features, which improves the convergence speed and the detection accuracy. The most important training parameters were set as follows:

Table 2 YOLO training parameters

Image size; 224*224 pixels
Batch size; 16
Number of epochs; 50
Optimiser; Stochastic Gradient Descent (SGD) or Adam optimiser

While training, the model improves on three main loss components:

- Localization Loss; determines how accurate the predicted bounding box coordinates are.
- Objectness Loss; determines whether or not an object exists in a certain area.
- Classification Loss; determines the accuracy of predicted class labels.

Through the use of back propagation, the YOLO model will continuously change its parameters until the loss function converges. Once the model is trained, trained model is employed to segment all the images present in dataset.

Classification

For this study, skin lesion classifications were carried out with the EfficientNet-B5 Convolutional Neural Network (CNN) architecture, implemented with the PyTorch framework. EfficientNet is a cutting-edge architecture of deep CNNs that achieves a high level of accuracy with a reduced number of parameters because of its unique compound scaling approach.

Several layers of convolutional stages make up EfficientNet-b5 and are built to gradually extract features hierarchically from images. Typically, architectures consist of the following components.

Input layer: The first layer receives dermoscopic images resized to 224 x 224 x 3. Images are normalised to increase stability during training.

Convolutional layer: The first layer in the network performs a regular convolution operation to obtain low-level features of the image such as edges and textures. From here, the initial feature map is sent to the next deeper layers.

Layers = RELU activation function (Batch normalization (Conv(Input image)))

(7)

At the centre of EfficientNet are the Mobile Inverted Bottleneck Convolutions (MBConv) which come from the MobileNet architecture. Each of these blocks has

- An Expansion Layer: Features are expanded here by using pointwise convolutional. The equation for the same is $\text{Function} = \text{convolutional } 1 \times 1 (\text{Input}_{\text{image}})$
- A Depth wise Convolution: In depth wise convolution, spatial filtering is performed separately for each channel. $\text{Function} = \text{Convolutional}_{\text{depthwise}} (\text{F}_{\text{Expansion layer}})$. This equation reduces dimensions of input image which helps in increasing chances of efficient features
- A Squeeze-and-Excitation (SE) Module: The Squeeze-and-Excitation module enhances feature representation through channels by understanding the importance weights of channels. Afterwards, the original feature maps are adjusted by these weights.
- A Projection Layer: To round off, the expanded features are brought back to lower dimensions using a pointwise convolution.

After these block an average pooling layer is placed, where input features are batched with window size. Perform average of intensity pixels of window sized elements. This averaged intensity value is placed at central region of window size.

Last layer is SoftMax layer, it gives the probability of seven classes (N=7). Highest probability classes are the predication of respective input image. It is provided by below equation,

$$\text{Probability}_{(\text{class } i)} = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}}$$

(8)

Once all models are designed, model is set to train for the input images with the parameters provided in below Tab.3.

Table 3 Model training parameter

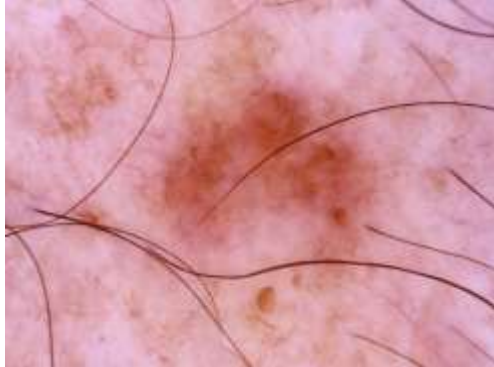
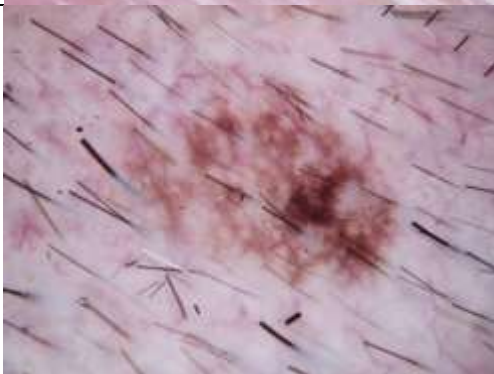
Layer name	Value of it
Input image	224 x 244
Batch size	16
Epochs	25
Learning rate	0.00003

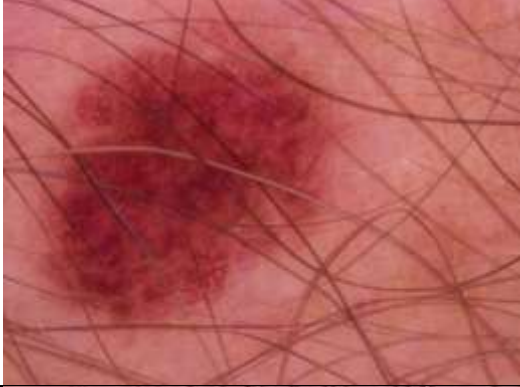
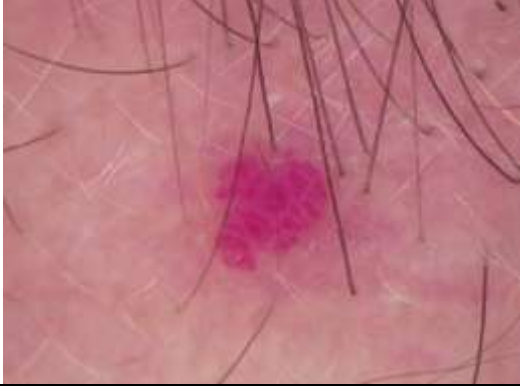
Optimizer	ADAM
Loss function	Focal loss
Output classes	Seven

4. RESULT AND DISCUSSION

The proposed skin lesion analysis framework was tested on the HAM10000 dataset containing dermoscopic images that are classified into seven different skin disease types such as basal cell carcinoma, dermatofibroma, vascular lesions, actinic keratosis, benign keratosis-like lesions, melanoma and melanocytic nevi. The dataset images were structured in accordance with the provided metadata labels and were utilised in both training and evaluation of the proposed system. Below Tab.4. shows the instance of employed dataset,

Table 4 Examples of dermoscopic images showing different types of skin lesions from the HAM10000 dataset.

Name of class	Image of respective class
AKIEC	
BCC	
BKI	

DF	
MEL	
NV	
VASC	

(Image Quality) Median filtering, CLAHE, Non-Local Means Denoising and black-hat morphological filtering were all utilised, and in order to select the best performing filter for each of the remaining processing steps, these were taken into consideration of the following image quality metrics: PSNR, MSE, SSIM, and Edge Preservation Index. Below

Tab.5. shows the selection of best filter based on score and also shows results of filtered images. Below figure shows before filter application and after processing the image.

Table 5 Filtering results and selecting optimum filter for input image

Input image		Processed image					
							
Score for selecting best filtered image is given here							
class	image	filter	PSNR	MSE	SSIM	EPI	score
mel	ISIC_0028	median	34.5811	22.6449	0.87951	0.592693	8.5515
mel	ISIC_0028	clahe	24.95038	207.9897	0.797045	0.917646	-12.7995
mel	ISIC_0028	nlm	34.45201	23.32811	0.805884	0.671648	8.44605
mel	ISIC_0028	blackhat	4.852118	21274.95	0.004753	-0.99008	-2126.34
mel	ISIC_0028	BEST					median

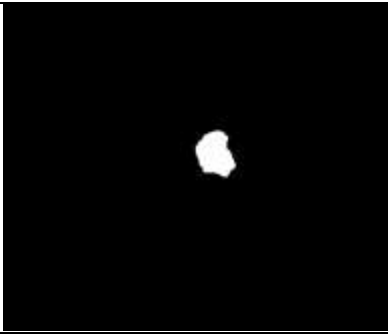
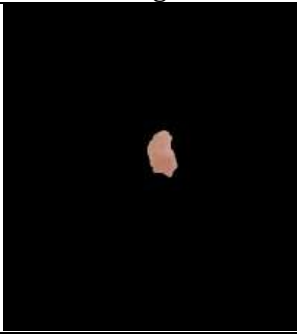
The visibility of skin lesions, and the associated detracting noise that may arbitrarily affect the accuracy of the segmentation, was enhanced by the removal of hair artefacts from dermoscopic images using a combination of morphological operations and inpainting. Removal of hair is clearly showed before and after the process is shown below Tab.6.,

Table 6 Processed hair removed image

Filtered image	After hair removed images
	

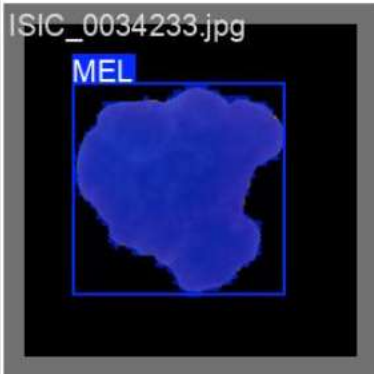
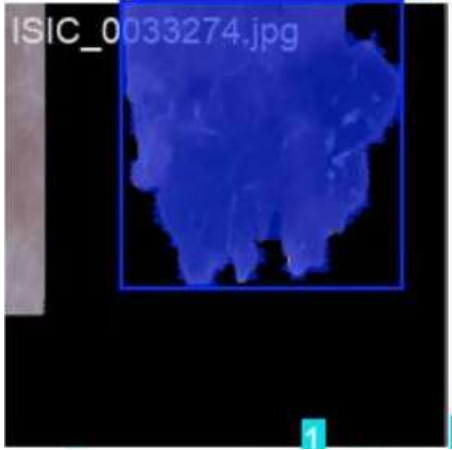
The Segment Anything Model was used to create precise segmentation for the lesions and the segmented lesions are further processed by the YOLO object detection framework for bounding box prediction. The SAM segmented images are showed in below Tab, which includes original image, binary mask of lesion and segmented region of it is as shown in Tab.7.

Table 7 Process of masks generation using SAM

Hair removed image	Binary masked image	Lesion segmented image
		

YOLO segmentation process, its score and confidence is provided in below Tab.8. It takes SAM boundary boxes, hair removed images as input.

Table 8 YOLO segmented process

Hair removed image	Description
	 Boundary boxed images with class name by YOLO
	 Boundary boxes at the affected region from YOLO

The EfficientNet-B5 Deep Learning Model was used to classify skin lesion images. The classifier was optimised through training via transfer learning in conjunction with data augmentation for improved generalisation. As for the experiment, the initial proposed

pipeline improved detection and classification of lesions to high accuracy over multiple categories of skin diseases. The integrated pipeline of image improvement, segmentation, detection, and deep learning classification provides enhanced consistency for automated skin lesion diagnosis. Below figure shows accuracy achieved per epochs, training accuracy, confusion matrix for trained model. Had employed GPU P100 in Kaggle to execute the proposed work and took one hour forty two minutes fifty one seconds to complete training process. Novelty concerns in the implementation introduced,

1. Augmentation – introduced horizontal flip, rotation with 20 degree angle, normalise image, color jitter.
2. Instead of normal split of images between classes had employed stratified split. To address NV imbalance in dataset to manage imbalance to optimum level stratified split is employed.

Performance metrics for evaluating the trained model for the testing data considered are, accuracy, f1-score, recall and precision. In accordance with the metrics confusion matrix gives clear image regarding where model performance is good and where its performance is bad. In addition to this accuracy per class is provided in this sub-section let's see those one by one.

The model's demonstrated overall classification accuracy stands at 0.96 (96.45%) and means that almost all dermoscopic images were correctly classified into their respective categories. The definition of accuracy is that it is the number of correctly predicted samples when compared to the total samples in the dataset. Because this model returned such a high accuracy value, it is reasonable to conclude that the model is able to differentiate between a multitude of skin lesion types with great clarity.

When a model predicts a class of skin disease, the model's reliability for that prediction is measured by the model's precision value. The precision value of 0.9623 (96.26%) means that of the predicted positively classed samples, the majority are indeed correct. Medical diagnosis relies heavily on a good precision score to ensure that a lesion that is healthy is not diagnosed as a diseased lesion; otherwise, it will contribute to a high number of false positive predictions.

The model's recall value of 0.9626 (96.17%) pertains to the number of positive samples in the class of skin lesion that the model is able to correctly identify. Recall is defined as the model's ability to identify true cases at the positive end of the sample. Since the recall value is high, it indicates that the model possesses the ability to identify a large number of genuine skin lesion cases, and a small number of false cases will also be identified.

The F1 score, which measures the harmonic mean of recall and accuracy, was 0.9619, or 96.19%. The F1 score considers both the good and negative outcomes of a test. If a model's F1 score is high, it means that its accuracy and recall are well-balanced, which leads to excellent performance. The tabular data presents an overview of the accuracy, recall, and F1 score for each class as tabulated in Tab.9.

Table 9 Comparison of precision, recall, F1score for each class

Class ID	Precision	Recall	F1 score
Akiec	100	100	100
Bcc	98	98	98
Bkl	91	98	94
Df	100	100	100
Mel	88	91	89
Nv	97	93	95
vasc	100	100	100

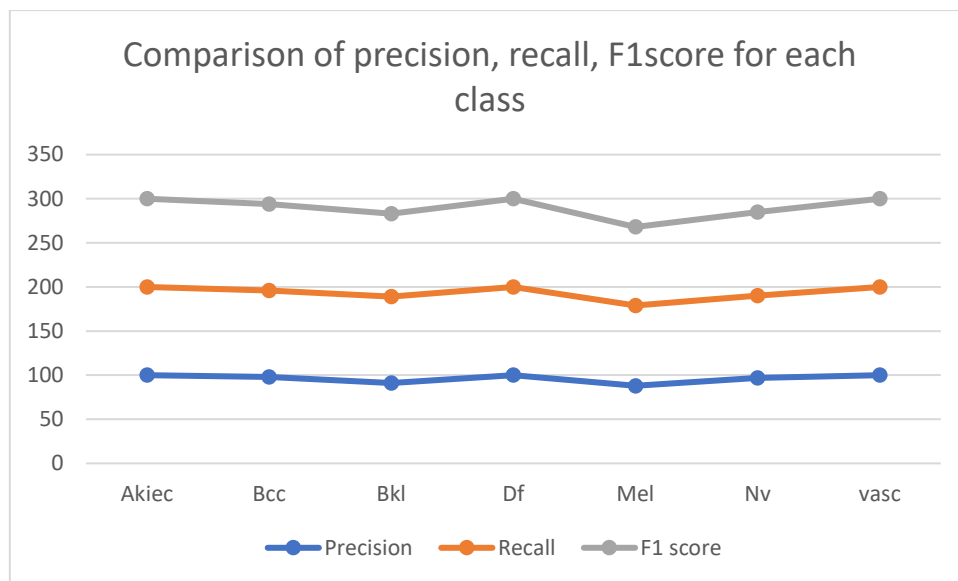


Fig. 5 Comparison of precision, recall, F1 score for every class

Confusion matrix: Figure 5 demonstrates the confusion matrix, which was used to assess the model's prediction capacity. Each category of skin lesions, including akiec, bcc, bkl, df, mel, nv, and vasc, was classified using the EfficientNet-B5 model. The rows of the matrix indicate the actual class label, while the columns indicate the class label as predicted by the model. Correctly predicted samples are denoted by the diagonal elements, whereas the off-diagonal elements indicate incorrect predictions. The confusion matrix shows that most of the predictions are accurate because the model demonstrates a significant amount of classification accuracy. For the actinic keratosis (akiec) category, the model correctly classified all 2616 samples. For the dermatofibroma (df) category, the model had a perfect score, which was 2105 predictions, signifying dependable model reliability. Likewise, the model is capable of accurately differentiating vascular lesions from other lesions by correctly classifying 2272 samples in the vascular lesion (vasc) category.

Basal cell carcinoma (bcc) class: Correctly classified: 2025, Incorrectly classified: 3 samples misclassified as something else (melanocytic nevi (nv)). Benign keratosis-like lesion (bkl) class: 2162 correct predictions, also misclassified as akiec, bcc, melanoma, and nv (small number of images, likely due to visual similarities between lesions) Melanoma (mel) class: 2164 correct, and misclassified to bcc, bkl, nv (mel) and benign lesions due to the same texture and colour patterns. Melanocytic nevi (nv) class contains the largest number of samples classified correctly (6592)

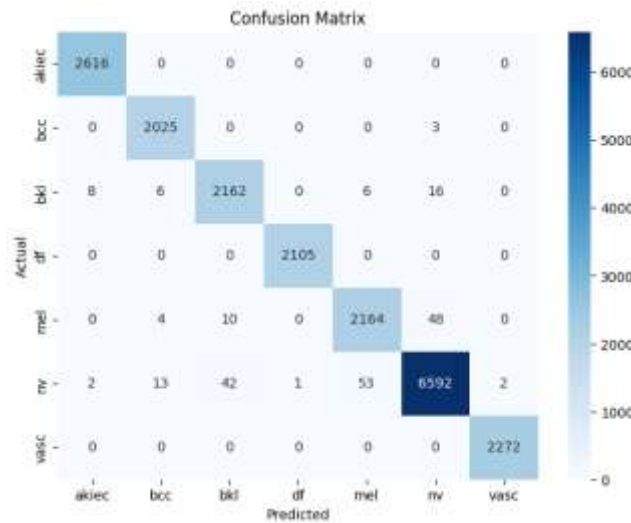


Fig. 6 Confusion matrix

To support robustness of model, we had generated accuracy per class is as shown in below Tab.10.

Table 10 Accuracy performance per class

Class name	Accuracy (in percentage)
Akiec	100
Bcc	99.85
Bkl	98.36
Df	100
Mel	97.21
Nv	98.31
Vasc	100

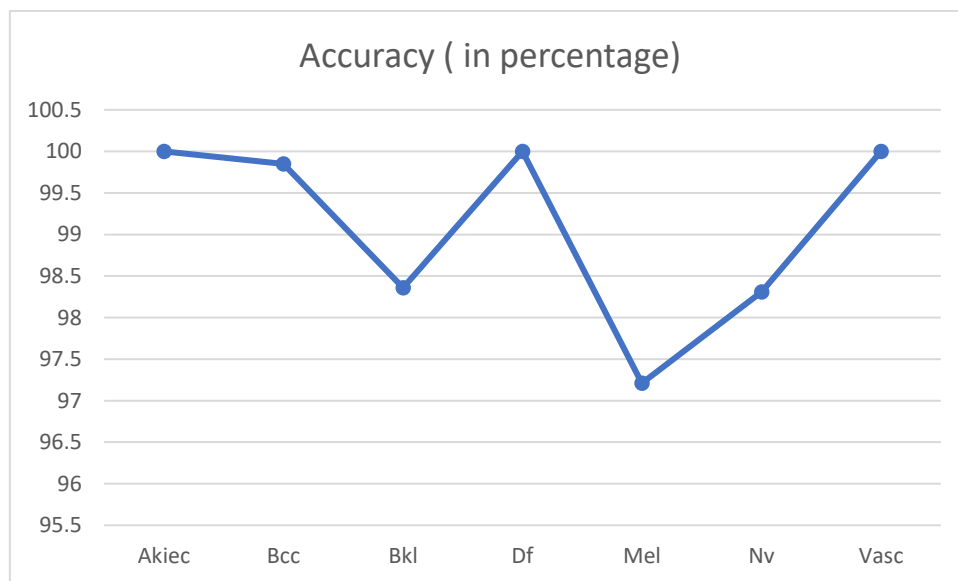


Fig. 7 Accuracy per epoch for every class

Comparison of different methods in detecting skin disease detection is provided in below Tab.11.

Table 11 Comparison of performance metrics with proposed method and existing state of art methods

Employed method	Accuracy	Precision	Recall	F1 score
SVM[16]	71	33.71	24.57	24.57
KNN [16]	55	23	21.7	22.71
DecisionTree [16]	57	22	24.29	23.86
CNN [17]	91.1	83.7	97	89.9
MobileNetV3-Small [18]	77	--	--	94
MobileNet V3 Large [18]	83	--	--	95
EfficientNetB4 [18]	81	--	--	95
ResNet50 [18]	77	--	--	96
DenseNet201 [18]	79	--	--	95
CNN [19]	87	--	--	--
Proposed Filtering, SAM, YOLO,	96.45	96.26	96.17	96.19

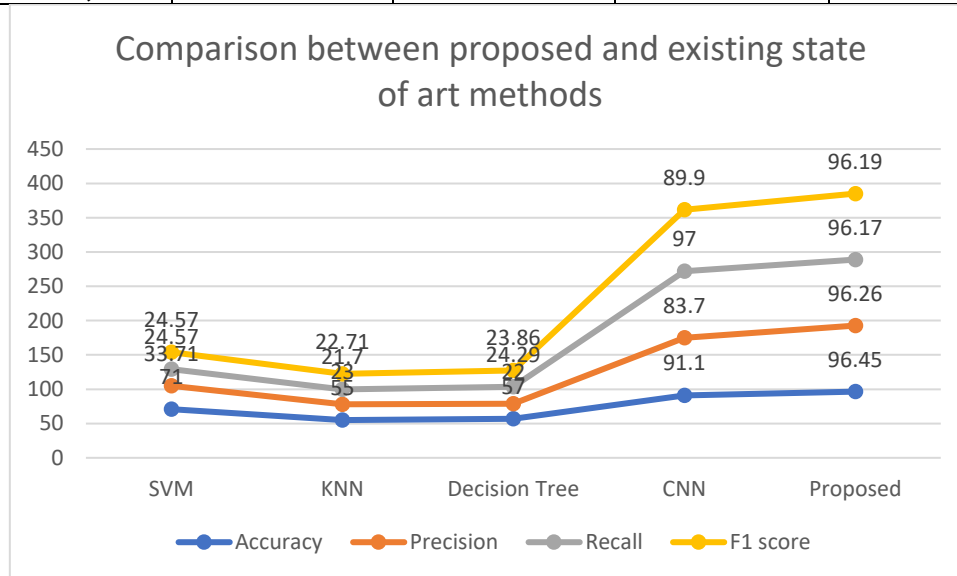


Fig. 8 Comparison of performance metrics between existing and existing state of art methods

5. CONCLUSION

In this research, a complete deep learning skin lesion detection and skin lesion classification system has been developed for automated analysis of dermoscopic images. The HAM10000 dermoscopic image dataset for skin lesions was used. This dataset comprises dermoscopic images belonging to seven different skin lesion categories. First,

Median filtering, CLAHE, Non-Local Means Denoising, and black hat morphological filtering were employed to enhance images and improve lesion visibility. The most effective dermoscopic image enhancement method was determined by selecting the filter with the most favourable Weighted Scoring mechanism, in terms of PSNR, SSIM, edge preservation index, and MSE. Lastly, morphological operations and inpainting techniques were used for hair artifact removal to enhance the lesion.

After preprocessing the data, we used the Segment Anything Model for lesion segmentation, which produced a binary mask and images showing the lesion while removing the lesion area from the skin image. The segmented images were processed with the YOLO object detection model, which generates bounding boxes around the lesion area and therefore accurately localises the area of interest. The isolated lesion images were classified using the EfficientNet-B5 model, applying transfer learning and data augmentation techniques. The pipeline demonstrated its value when it classified the lesion images and segmented, detected, and classified images, achieving a remarkable accuracy of 96.5%. This multi-stage model enhances dermatoscopic lesion localisation and diagnostics, supporting dermatologists in detecting skin diseases early and accurately.

Reference

- [1] Alam, T. M., et al. "An Efficient Deep Learning-Based Skin Cancer Classifier Using the HAM10000 Dataset." *Diagnostics*, 2022.
- [2] Behara, Kavita, Ernest Bhero, and John Terhile Agee. "An improved skin lesion classification using a hybrid approach with active contour snake model and lightweight attention-guided capsule networks." *Diagnostics* 14.6 (2024): 636.
- [3] Debelee, Taye Girma. "Skin lesion classification and detection using machine learning techniques: a systematic review." *Diagnostics* 13.19 (2023): 3147.
- [4] Gessert, N., et al. "Skin Lesion Classification Using Ensembles of Multi-Resolution EfficientNets." *Computers in Biology and Medicine*, 2020.
- [5] Jain, Satin, et al. "Deep learning-based transfer learning for classification of skin cancer." *Sensors* 21.23 (2021): 8142.
- [6] Kassem, M. A., et al. "Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review. *Diagnostics*. 2021; 11: 1390." 2021,
- [7] Kreouzi, Magdalini, et al. "Deep learning for melanoma detection: a deep learning approach to differentiating malignant melanoma from benign melanocytic nevi." *Cancers* 17.1 (2024): 28.
- [8] Naqvi, Maryam, et al. "Skin cancer detection using deep learning—a review." *Diagnostics* 13.11 (2023): 1911.
- [9] Raghavendra, Paravatham VSP, et al. "Deep learning-based skin lesion multi-class classification with global average pooling improvement." *Journal of Digital Imaging* 36.5 (2023): 2227-2248.
- [10] Shetty, Bhuvaneshwari, et al. "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network." *Scientific Reports* 12.1 (2022): 18134.
- [11] Tschandl, Philipp, Cliff Rosendahl, and Harald Kittler. "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions." *Scientific data* 5.1 (2018): 180161.
- [12] Wu, Yinhao et al. "Skin Cancer Classification With Deep Learning: A Systematic Review." *Frontiers in oncology* vol. 12 893972. 13 Jul. 2022, doi:10.3389/fonc.2022.893972

- [13] Alotaibi, Areej, and Duaa AlSaeed. "Skin Cancer Detection Using Transfer Learning and Deep Attention Mechanisms." *Diagnostics* (Basel, Switzerland) vol. 15,1 99. 3 Jan. 2025, doi:10.3390/diagnostics15010099
- [14] Akinrinade, Olusoji, and Chunglin Du. "Skin cancer detection using deep machine learning techniques." *Intelligence-Based Medicine* 11 (2025): 100191.
- [15] Sara, Umme, Morium Akter, and Mohammad Shorif Uddin. "Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study." *Journal of Computer and Communications* 7.3 (2019): 8-18.
- [16] Mostafiz Ahammed, Md. Al Mamun, Mohammad Shorif Uddin, A machine learning approach for skin disease detection and classification using image segmentation, *Healthcare Analytics*, Volume 2, 2022, 100122, ISSN 2772-4425, <https://doi.org/10.1016/j.health.2022.100122>.
- [17] Ajani, Atinuke Arinola, Ishaya Gambo, and Oluwasegun Faniyi. "Artificial Intelligence–Assisted Diagnosis of Atopic Eczema in Darker Skin Types: A CNN-Based Study." *Dermatology Practical & Conceptual* 16.1 (2026): 5897.
- [18] Tanantong, Tanatorn, et al. "Mobile-Based Deep Learning Framework for Classifying Common Skin Diseases in Thailand." *ICIC Express Letters Part B: Applications ICIC International* 15.5 (2024): 495-503.
- [19] Inthiyaz, Syed, et al. "Skin disease detection using deep learning." *Advances in Engineering Software* 175 (2023): 103361.