Conditional Adversarial Segmentation and Deep Learning Approach for Skin Lesions Subtyping from Dermoscopic Images

Mirunalini P[1†], Karthik Desingu[1†], Aswatha S[1], Deepika R[1], Deepika V[1], Jaisakthi S M[2*†]

[1]Department of Computer Science Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, 600130, Tamil Nadu, India.

[2*]School of Computer Science & Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, 600127, Tamil Nadu, India.

*Corresponding author(s). E-mail(s): jaisakti.murugaiyan@vit.ac.in;
Contributing authors: miruna@ssn.edu.in; karthik19047@cse.ssn.edu.in; aswatha18026@cse.ssn.edu.in; deepika18041@cse.ssn.edu.in; deepika18042@cse.ssn.edu.in;
†These authors contributed equally to this work.

Abstract

Automatic skin lesion subtyping is a crucial step for diagnosing and treating skin cancer and acts as a first level diagnostic aid for medical experts. Although, in general, deep learning is very effective in image processing tasks, there are notable areas of the processing pipeline in the dermoscopic image regime that can benefit from refinement. Our work identifies two such areas for improvement; First, most benchmark dermoscopic datasets for skin cancers and lesions are highly imbalanced due to the relative rarity and commonality in the occurrence of specific lesion types. Deep learning methods tend to exhibit biased performance in favor of the majority classes with such datasets, leading to poor generalization. Second, dermoscopic images can be associated with irrelevant information in the form of skin color, hair, veins, etc. And limiting the information available to a neural network by retaining only relevant portions of an input image has been successful in prompting the network towards learning task-relevant features by improving its performance. Hence, this research work augments the skin lesion characterization pipeline in the following ways. First, it balances the dataset to overcome sample size biases. Two balancing methods, Synthetic Minority Over-sampling Technique (SMOTE) and Reweighting are applied, compared, and analyzed. Second, a lesion segmentation stage is introduced before classification to retain only the region of interest (ROI). A baseline segmentation approach based on Bi-Directional ConvLSTM U-Net (BCDU-Net) is improved using conditional adversarial training for enhanced segmentation performance. Finally, the classification stage is implemented using EfficientNets, where the B2 variant is used to benchmark and choose between the balancing and segmentation techniques, and the architecture is then scaled through B7 to analyze the performance boost in lesion classification. From these experiments, we find that the pipeline that balances using SMOTE and uses the adversarially trained segmentation network achieves the best baseline performance of 91% classification accuracy with EfficientNet B2 baseline. Based on the scaling experiments, we find that optimal performance is reached with the B6 architecture, that classifies with a 97% accuracy, which results in 6.6% increase with respect to the B2 baseline model. Furthermore, the proposed pipelined model for the characterization of lesion outperforms the state of the art when compared with the same data set.

Keywords: Skin Lesion, EfficientNet, SMOTE, Re-weighting, BCDU-Net, CGAN, Deep learning, Adversarial Segmentation

1 Introduction

The number of people diagnosed with skin cancer is increasing annually. Skin lesions are caused by various dermatological disorders and can be benign and malignant types. Therefore, it is important to discriminate benign and malignant lesions for proper treatment since early detection of malignant lesions can significantly reduce the risk of fatality. However, manual inspection of skin lesions is prone to failure, time consuming, and subjective, since most skin lesions classes have high similarities. Routine clinical settings achieved less than 80% accuracy in detecting melanomas from dermoscopic images [49]. Therefore, an automatic method is essential to classify the types of skin lesion. This can be used as a preliminary diagnosing tool by the medical expert for treating the patients.

The occurrence of different types of skin lesions is highly disproportionate, some being extremely rare and others common. As a result, the dermoscopic image data sets available for skin lesions also exhibit an imbalance. Diagnostic systems, especially those based on machine learning techniques, are prone to bias when trained using unbalanced data sets [10, 11, 37]. Hence, it is imperative to ensure that such bias is eliminated when developing such diagnostic algorithms.

Dermoscopic images are frequently characterized by artifacts such as hair, vein, marker annotations, etc. that are irrelevant to the skin lesion. The presence of artifacts has a negative impact on the ability of machine learning-based classifiers to learn representative features that can distinguish types of lesion. Therefore, segmenting to retain only the relevant portions of the lesion from the dermoscopic images before feeding them to the classifier will improve its performance. However, since lesions typically occur in irregular shapes, the segmentation process is complex and challenging. In such scenarios, deep-learning-based segmentation methods are known to outperform traditional heuristic-based and algorithmic image processing methods.

To address these issues, we require more complex features than traditional hand-crafted features used in image processing methods. Hence, we propose a deep-learning-driven diagnostic process to leverage their proven ability to learn discriminative features from their input incrementally through their hidden-layer architecture. The proposed process is an automatic two-level diagnostic system that first segments skin lesions regions from dermoscopic images and then classifies the lesion type from these images.

The rest of this article is organized as follows: Section 2 presents a brief review of the related literature. The research gaps have been summarized in Section 3. The proposed methodology towards the research gap is outlined in Section 4. Section 5 presents experimental results and section 6 discusses the results obtained. Conclusion and comparison of the results with existing work are discussed in Section 7. The implementation of this research work is made available at https://github.com/karthik-d/lesion-characterization-using-cgan

2 Related Work

The accurate classification of skin lesions in their early stages enables dermatologists to treat patients well in time. The time-consuming and interevaluator bias in the manual diagnosis of skin cancer can be overcome with a computerized analysis system.

Before classification, skin lesion segmentation is an important step. The presence of artifacts such as hair and air bubbles, intrinsic factors such as size, texture, contrast, shape and color and variations in image acquisition conditions make lesion segmentation and classification a challenging task [26, 34]. Deep learning have been widely used for medical image analysis but there are some challenges in the accurate segmentation and classification of the skin lesions by deep learning models due to the following reasons such as availability quality data, unavailability of large balanced dataset, high inter-class and intra-class variations [44], the noise and the unwanted artifacts in the dermoscopic images impedes the feature extraction process by the network [4, 22].

A crucial step in the workflow for automated skin lesion analysis is segmentation. There are many ways to achieve the segmentation one most familiar method is ABCD algorithm (Asymmetry, Border, Color, Diameter of lesions) which is rule-based skin lesion segmentation system [36] and its derivatives ABCDE[1] (Asymmetry, Border, Color, Diameter and Evolving) and ABCDEF [25] (Asymmetry, Border, Color, Diameter, Evolving and Firm). Before deep learning models the segmentation and analysis of skin lesion was based on the image processing and machine learning techniques such as thresholding [7], active contours and region growing [6, 12] and by using clustering techniques [9]. Jaisakthi et al. [23], proposed an automatic skin lesion segmentation model using Grabcut and the K-means algorithm. The k-means algorithm helps improving the boundaries of the segments by utilizing the color features learnt from the training images.

All these approaches rely on hand-crafted feature extraction, which is very hard and limits discriminatory capability. These traditional methods may not yield good results for large and complex datasets. Since deep learning techniques smoothly blends feature extraction with learning it always demands for large datasets. Although deep learning approaches has demonstrated remarkable performance in medical image analysis its applicability and generalizability to unseen, out-of-distribution data are limited by its dependence on high-quality annotations [45].

One of the important steps in computerized analysis of skin lesions is to extract the region of interest, that is, the exact location of the pigmented lesion. Using a segmentation model, the skin lesion region can be separated from normal skin. Different studies proposed different models for the purpose of segmenting the ROI. Manu et al. [16], designed end-to-end ensemble segmentation methods to combine Mask R-CNN and DeeplabV3+ with pre-processing and post-processing methods to produce accurate lesion segmentation for the ISIC 2017 dataset. This approach accurately generates the bounding boxes for objects, its performance on generating the masks are much less and the boundaries are not complete. Lina Liu et al. [31], proposed a new convolutional neural network (CNN)-based method for segmentation of skin lesions with auxiliary task learning to identify area of infection. Jafari et al. [21], proposed an automated skin lesion segmentation model using deeper networks and a residual learning technique. Sohaib et al. [18], proposed a modified U-Net architecture with a 46-layered structure to obtain a good segmentation rate on the ISIC 2018 dataset after pre-processing the images to remove artifacts.

As CNN is a deep learning model, training these models requires more training images. This can be achieved by a data augmentation method using Generative Adversarial Network. Federico Pollastri et al. [39], proposed a modified Deep Conventional Generative Adversarial Network(DCGAN) model to generate both skin lesion images and their segmentation masks for dermoscopic images. Cristian Lazo et al. [29], proposed a Generative Adversarial Network model built using the pix2pix network architecture called PatchGAN. Umaseh Sivanesan et al. [43], proposed a conventional GAN model with pix2pixHD and U-Net + architecture, which proves that the GAN model works more efficiently with a discriminator to enhance mask generation by getting the feedback from the discriminator. Shubham et al. [19] proposed a unique adversarial learning-based approach namely Efficient-GAN (EGAN), that creates precise lesion masks using an unsupervised generative network.

Several proposals were made to overcome the challenges faced in using the dermoscopic images. Kassem et al. [27], proposed a model utilizing transfer learning and a pre-trained GoogleNet model by adding more filters to each layer, to enhance features and reduce noise. Gessert et al. [13] proposed a classification model for dermoscopic images using ensemble of deep learning methods which also predicts an unknown class with a data-driven approach and also addresses the class-imbalance problem using loss balancing approach. In [17] authors compared the diagnosis performance between dermatologists and a Inception-v4 model at two levels using only dermoscopic images and clinical information.

The authors in [35] proved that the accuracy of deep learning models is better after resolving the class imbalance problem. In [48], the authors studied the effect of using data augmentation and SMOTE for balancing the data set for classifying tuberculosis, pneumonia, and COVID-19.

The effect of image size for skin lesion classification was studied using multi-scale multi-CNN fusion approach, which boosts the classification performance compared to a single network or a single image scale [32]. A hyper-connected CNN (HcCNN) with a multi-scale attention block(MsA) was used to classify skin lesions by prioritizing semantically important subtle regions [5]. An ensemble network based on the major voting step from pre-trained models using ResNet50, MobileNetV3 and EfficientNet was proposed for skin classification process [47].

Zhou et al. [52], proposed a model which includes several deep CNN models such as densenet121, se-resnext50, se-resnext101, Efficientnet-B2, Efficientnet-B3, and Efficientnet-B4 and experimented by taking different combinations of the base models mentioned above and obtained the best values for precision and recall. Mingxing Tan et al. [46], demonstrated that a mobile size EfficientNet model can be scaled up very effectively, surpassing state-of-the-art accuracy with an order of magnitude fewer parameters and FLOPS, on both ImageNet and five commonly used transfer learning datasets. Hence, skin lesion segmentation can be performed before classification in order to get the boundary information or regions of interest (ROI).

There are only few studies conducted on experimenting the impact of skin lesion segmentation on the performance of the classifiers. Yu et al. [50], proposed a two staged lesion classification model with the help of deep residual network. This study showed that filtering out the useless background considerably increased the performance of the classifier. Zhang et al. [51], proposed an attention residual learning convolutional neural network that includes an attention method to obtain the essential features of the images and there is no explicit segmentation, and the proposed model was able to adaptively focus on discriminative parts and achieved the state-of-the-art performance. Al-masni et al. [2], proposed an integrated diagnostic framework that combines skin lesion segmentation using deep segmentator called FrCN and classification accuracy were compared with different deep learning models and best performance was achieved using Inception-ResNet-v2. Diaz [15], presented DermaKNet, a CAD system based on CNNs, which first segments the dermoscopic image by passing it to a Lesion Segmentation Network(LSN) for generating the binary mask which then agumented and passed to the the Diagnosis Network(DN), which provides the diagnosis. Mahbod et al. [33] investigated the impact of the skin lesion segmentation on the performance of dermoscopic image classification using ISIC 2017 image dataset. In [30], the authors proposed two deep learning frameworks, the Lesion Indexing Network (LIN) by adopting FCRN-88 and the Lesion Feature Network, which is a CNN-based framework to address lesion segmentation, classification and dermoscopic feature extraction.

From the literature, it is shown that DL-based techniques are more effective in obtaining accurate segmentation results, utilizing a generative adversarial network increased the accuracy of creating precise lesion masks. Classification results will also be improved when the precision lesion region is segmented and then classified.

2.1 Research Gap

This section summarizes some of the research gaps found in the literature. Visual inspection of skin lesions with the naked eye cannot easily distinguish between different types of skin cancer and may result in an incorrect diagnosis. The use of dermoscopy, a reliable medical imaging technology, has helped in the better visualization of the skin lesions. But still there remains some complexity in manual examination of the dermoscopic images and it is time consuming. The pre-processing of a dermoscopic image before classification is very much needed. Yu et al. [50] applied segmentation without preprocessing images, which resulted in the failure of some segmentation and classification cases. We have overcome this problem by preprocessing images prior to classification by removing unwanted artifacts using Contrast Limited Adaptive Histogram Equalization (CLAHE) and inpainting algorithms.

The dermoscopic images that we have used for the purpose of classification are heavily imbalanced, varying from 200 to 12000. As a result of this imbalanced dataset, there is a chance of bias towards classes having a greater number of samples. This problem was not handled by majority of the works. The authors of [27], [15], [13] and [52] have handled this problem by randomly adding images to the minority classes and by using simple image augmentation methods. These methods increase the risk of overfitting as they produce images similar to the existing ones. Hence, to overcome the problem of overfitting, SMOTE oversampling technique [41], [48] was used in this research work to increase the number of samples in the minority classes. SMOTE generates new synthetic samples from the existing images, thus increasing the generalization capabilities of the classifier and therefore increasing the performance.

The classification of the entire dermoscopic images may not be effective, as they may contain some background features that affect the performance of the identification process. Most of the papers [27], [13] have performed only classification on the entire dermoscopic images and have obtained lesser accuracy. But few papers [2],[15],[51],[18] have segmented the skin lesions images before classifying them and were unable to obtain the region of interest accurately. We proposed a conditional GAN based approach [29], [39], [43] to generate the masks with precise boundary with the help of discriminator feedback.

2.2 Key Contributions

The contributions we have made towards improving the performance of classification on dermoscopic images are as follows:

- Re-weighting and SMOTE techniques were used to mitigate the class imbalance problem and prevents over fitting of augmentation methods in the dataset which helps in enhancing the robustness of the classification model.
- Employed the BCDU-Net model for effective segmentation of regions of interest in dermoscopic images. Achieved improved segmentation performance through the integration of a Conditional GAN-based generator-discriminator training method.
- Generated and validated ground truth masks for the ISIC 2019 dataset with the input of three experts, contributing to the creation of a reliable and publicly available dataset for skin lesion segmentation.
- Utilized the EfficientNet architecture, starting with the B2 version as a baseline. Demonstrated performance enhancement through a systematic approach of compound scaling, progressively transitioning from B3 to B6 versions.

3 Proposed Method

In this work, two different deep-learning based automatic diagnostic systems are proposed. The first system performs classification on the entire dermoscopic images to identify the lesion types, while the second system involves a two-level process, where the region of interest (ROI) is obtained through deep-learning-based segmentation followed by the classification of the segmented region. In both approaches, the problem of unbalanced dataset is overcome using either the SMOTE or re-weighting technique. The results were compared and analyzed using an exhaustive experimental setup. The proposed architecture system is depicted in Figure 1.

3.1 Pre-Processing

Dermoscopic images are characterized by a range of artifacts, and eliminating them before the training phase can significantly boost the performance of the system. To this end, the following methods were applied on the dermoscopic images as preprocessing steps to reduce the artifacts present and to enhance the image quality for further processing.

3.1.1 Hair and Vein Removal

The dermoscopic images in the dataset have different artifacts that include hair and veins. Since these artifacts obscure the region of interest of the lesion images, learning discriminating features becomes difficult for the neural network. Hence, the images are pre-processed by detecting and deleting hair nodules using the Frangi vesselness algorithm, and the removed portions are refilled using the inpainting method based on the Fast March Method (FMM) [24].

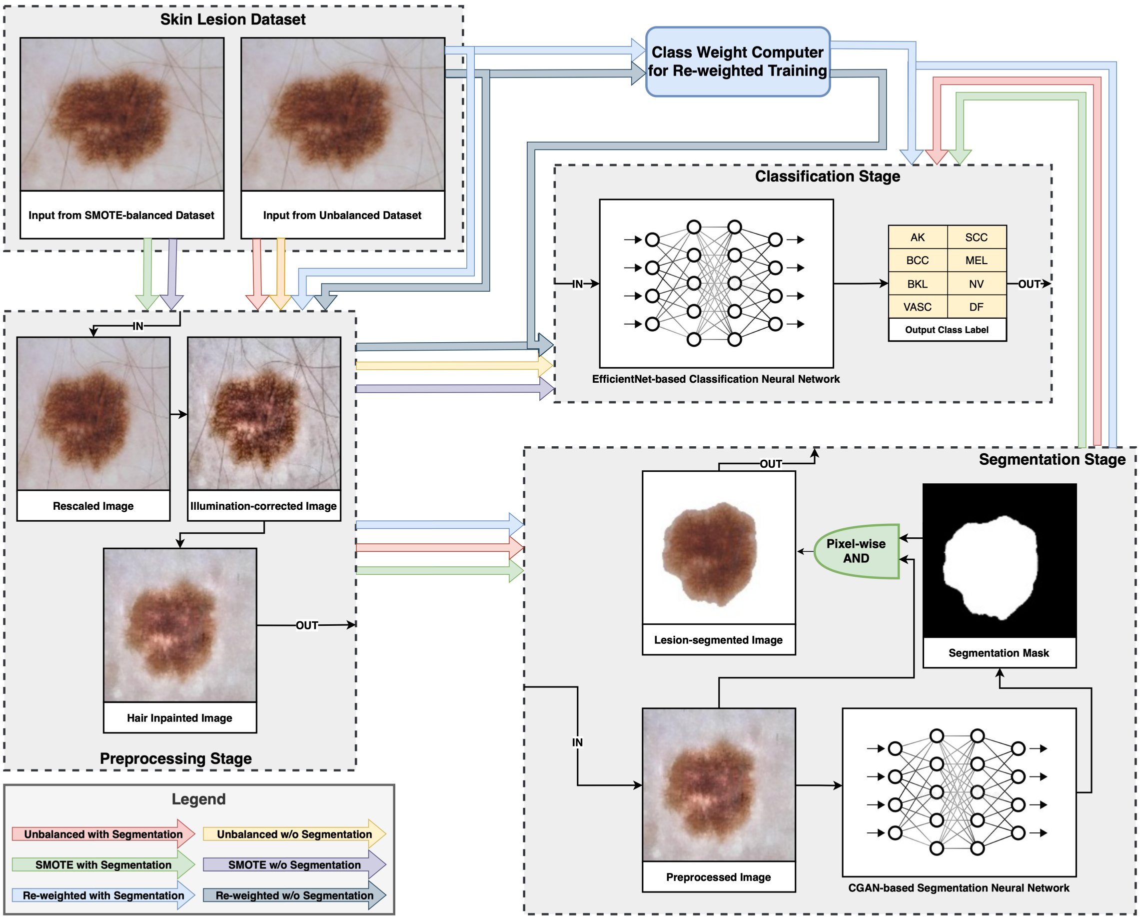

Fig. 1 Overall pipeline of the study, color-coded by the different experiments designed to analyze the effect of dataset balancing techniques and segmentation on lesion characterization.

3.1.2 Uneven Illumination Correction

Dermoscopic images are obtained using non-invasive microscopic technique usually suffers from uneven illumination during image acquisition [42]. This influences the automatic analysis system to suffer from inaccurate segmentation which results in decreased classification accuracy. So, CLAHE algorithm [38][53] was used as an illumination correction technique to overcome this problem. The CLAHE algorithm is applied to the L channel of the LAB color space converted image, which splits each image into smaller blocks and applies contrast limiting followed by histogram equalization for each block. Finally, the contrast-enhanced L channel is merged to form a LAB image and subsequently converted to a RGB image [20]. Thus, all the images are pre-processed using the aforementioned steps and passed on for subsequent analysis.

3.2 Classification of Dermoscopic Images

After pre-processing, the pre-processed images were classified into various categories of cancer types using different deep architectural models. In this proposed work, experiments were conducted to study the performance of the classifiers by training the architecture using, with and without segmenting the lesion region from the dermoscopic images. For segmenting the lesion region Bi-Directional ConvLSTM U-Net (BCD U-Net) [3] was applied and for classification EfficientNet was used. Additionally, the performance of the classifiers were analysed, by training the classifiers using balanced and unbalanced dataset. For balancing the dataset SMOTE and Re-Weighting technique were utilized. The performance of the above mentioned methods was investigated and the results were compared.

Several experiments were conducted to analyze the performance of the proposed methodology, by training the architecture with different strategies as listed :

- Classification of dermoscopic images with and without segmentation.
- Classification of dermoscopic images with balanced and unbalanced training set.

3.2.1 Classification of Dermoscopic Images with and without Segmentation

Two different methods were proposed to classify dermoscopic images into eight different categories of cancer types that are listed in Table 1. The first method uses entire dermoscopic images while the latter uses the segmented lesion region to train the EfficientNet architecture.

EfficientNets are a family of CNNs with increasing network complexities that uses a technique called compound scaling method. Rather than scaling randomly width, depth, or resolution, compound scaling method uniformly scales in all dimensions. This approach paves the way to achieve improved accuracy for fewer parameters and less computations than previous models. Specifically, the basic architecture version B0 is scaled up to obtain different variants of EfficientNets, named B1 through B7. In our work we have used B2 variant as a baseline to analyze the classification performance.

Although the model is capable of classifying the dermoscopic images, it also results in incorrect categorization because the model might not be able to distinguish between the lesion and the surrounding skin region. Therefore, segmenting the lesion region will help to separate the lesion region and also provides the classification model with the relevant portions of the image. This helps to improve the accuracy of the classifier and also improves the convergence rate of the model. The BCDU-Net architecture is used for segmentation and trained under two different settings, namely the baseline (simple supervised learning) and the generator-discriminator (CGAN method) settings.

Segmentation using Bi-Directional ConvLSTM U-Net — Baseline Method

An extended version of the U-Net with densely connected convolutions called BCDU-Net uses BConvLSTM layers for skip connections and reuses feature maps with densely connected convolutions. It has an encoding and a decoding path as depicted in Figure **??**. The encoding path of the architecture has four steps. It has two convolutional 3×3 filters followed by a 2×2 max-pooling function and ReLU. The feature maps are doubled at each step by collecting all the image representations, effecting an increase in dimension of these representations layer-by-layer. The high dimensional image representation of the final layer of the encoding path leads to rich feature representations.

Following the encoding path, there are three densely connected convolution layers. These densely connected convolution layers improve its performance through collective knowledge, and these feature maps are reused throughout the network. There are N consecutive blocks, and each block comprises two consecutive convolution layers. The decoding path performs up-sampling to the outputs of the previous layers. The feature maps from the contracting and expanding path are then processed using BConvLSTM. The output of the BConvLSTM is calculated as

$$Y_t = tanh(W_y^{\overrightarrow{H}} * \overrightarrow{H}_t + W_y^{\overleftarrow{H}} * \overleftarrow{H}_t + b)$$

where $\overrightarrow{H}_t$ and $\overleftarrow{H}_t$ denote the hidden state tensors for the forward and backward states, respectively, b is the bias term and Y_t indicates the final result. The upsampling function in each layer increases the size of each feature map and decreases the feature channels. At the end of the expanding path, the discriminative features learnt by the encoder are projected onto the pixel space.

Thus, the corresponding mask for the input image is generated using the baseline setup.

Segmentation using CGAN Method - Discriminator Generator Model

To achieve more precise segmentation results than the BCDU-Net, CGAN was used. The CGAN model has been trained with the aforementioned Bi-Directional ConvLSTM UNet (BCDU-Net) network, coupled with a pixel discriminator network under a Conditional Generative Adversarial Network (CGAN) setting, to perform skin lesion segmentation. In this setting, the generator is trained to generate masks of lesions, while simultaneously, a pixel discriminator is trained to identify ground truth masks as "real" and predicted masks as "fake". In effect, the generator eventually learns to produce masks that are close to the "real" image (the segmentation mask). The two simultaneous losses drive the generator to produce precise segmentation masks of lesions. The trained segmentation model was then applied to the entire classification data set to generate segmentation masks and validated the segmentation regions obtained with the help of medical experts. The overall architecture of the proposed segmentation network and adversarial training approach is depicted in figure 2.

The baseline BCDU-Net architecture was used as the generator of the CGAN model, an encoder-decoder neural network. This generator produces output images, which are conditional on the input image and the produced images were classified as real or fake by the pixel discriminator network. The discriminator is a fully convolutional neural network that evaluates the generated images at the pixel level. This fine-grained discrimination helps in capturing finer details in the generated output by comparing the generated target images and the ground truth images, pixel by pixel. Since the discriminator leverages both conditional information and pixel-wise evaluation, it is more appropriate for pixel-wise segmentation and generates accurate, well-defined lesion segmentation. The discriminator feedback is used to improve the efficiency of the generator in generating high quality masks (see figure 2).

The proposed CGAN approach can significantly increase the segmentation performance of a baseline segmentation network leveraging the conditioning imposed by the generator-discriminator learning setup. The proposed generator-discriminator model generates excellent quality segmentation masks for images with low contrast, unclear, and irregular boundaries, in comparison to the baseline segmentation network in this work. This performance improvement is analyzed in subsequent sections. EfficientNet models are then used to classify the segmented lesion images into different cancer subtypes.

Post-Processing

Post-processing involves the generation of the segmented region of the image by performing a logical AND operation of the binary mask with the original skin lesion image. The segmented skin lesion images obtained are then classified using the same EfficientNet variants, and the performance of the two approaches are compared.

3.2.2 Classification of Dermoscopic Images with Balanced and Unbalanced Training Set

The classification of the dermoscopic images was performed in two ways. Since the dataset is highly unbalanced with an uneven number of images in each class, there is a possibility of bias toward the class having a larger number of samples. Hence, to overcome the problem of imbalance, SMOTE and re-weighting techniques were applied to balance the dataset. Then, the EfficientNet architecture was used to classify both the entire dermoscopic images and segmented lesion regions. The performance of both classifiers was compared before and after balancing the dataset.

Dataset Balancing using SMOTE

Synthetic Minority Oversampling TEchnique (SMOTE) [8] generates synthetic samples from the minority class to balance the dataset. The synthetic instance is created by interpolating the positive class instance with one of the k-nearest neighbours at random in the feature space. The generated synthetic data point is neither the exact copy of the positive sample nor too different from the minority class. This algorithm helps to overcome the overfitting problem posed by random oversampling method.

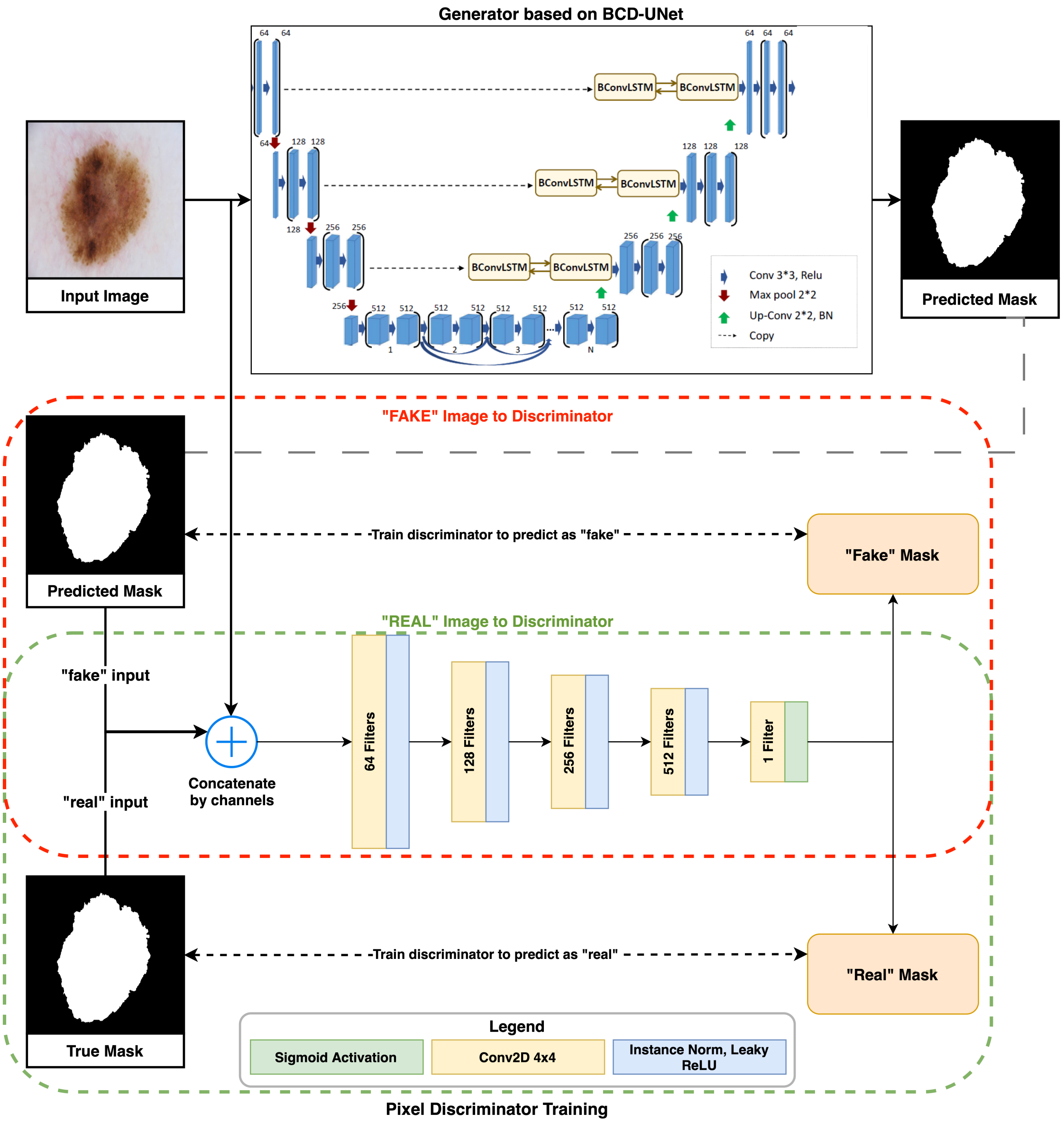

Fig. 2 Conditional adversarial training setup with BCD-UNet -based generator network and convolutional pixel discriminator to apply CGAN strategy for lesion segmentation training. *BCD-UNet architecture diagram adapted from [3].*

Dataset Balancing using Re-weighting

In re-weighting, higher weights are assigned to the minority classes and lower weights are assigned to the majority classes to influence the loss function. This method can be an alternative technique for balancing the imbalanced image dataset without the need for image resampling. The estimated class weights for unbalanced dataset were determined using equation 1.

$$weight_i = \frac{\text{total no. of samples}}{\text{no. of classes} \times \text{no. of samples in i}} \tag{1}$$

where, i denotes the class for which the weight is to be determined.

The proposed system was assessed using different experiments by training the entire dermoscopic images and segmented lesions, before and after balancing the imbalanced data set.

4 Experiments and Results

4.1 Skin Lesion Dataset

Two datasets from the ISIC Skin Lesions cohort were used as benchmark to evaluate the methods proposed in this work. The following sections describe the composition and use of these datasets in this study.

4.1.1 ISIC 2018 Segmentation Dataset

The 2018 segmentation dataset consists of 3694 dermoscopic images with 2594 training images, 100 validation images and 1000 testing images in the JPEG format. The binary mask ground truth for the training images are available as PNG images. A sample skin lesion image and its corresponding binary mask is depicted in figure 3.

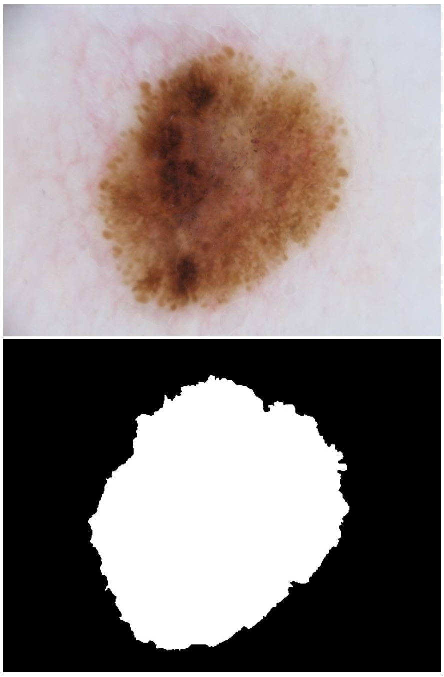

Fig. 3 ISIC 2018 segmentation dataset depicting a skin lesion image (Top) and its segmentation ground truth as a binary mask (Bottom)

4.1.2 ISIC 2019 Classification Dataset

The dataset for ISIC 2019 challenge contains 33569 dermoscopic images with 25331 images in 8 categories for training as shown in Table 1. The test set contains 8238 test images given in JPEG format. The classes melanoma, vascular lesion and squamous cell carcinoma are malignant type of skin cancer and the remaining classes fall under benign lesion types. In this dataset, the ground truth contains image ID and the class label of skin lesion for each images.

In this research work, we propose to classify the images by segmenting the lession region, but the dataset doesn't contain the segmentation mask, so we created the segmentation mask using CGAN and validated with the medical expert.

Table 1 Skin lesion classes and their corresponding sample sizes in the ISIC 2019 dataset.

Lesion Type	No. of Samples
Melanoma	4522
Melanocytic Nevus	12875
Basal Cell Carcinoma	3323
Actinic Keratosis	867
Benign Keratosis	2624
Deramatofibroma	239
Vascular Lesion	253
Squamous Cell Carcinoma	628

4.2 Dataset Partitioning

Both the datasets used as source data for this research work, the training subset is provided with corresponding ground truth. However, no validation subset is demarcated for the classification dataset, and the segmentation ground truth is not publicly available for the designated testing subset. Hence, the training subsets of both the datasets were first divided into 80% and 20% splits, randomly and universally for all experiments. During classifier training, 20% of the first split of the classifier data was used for epoch-wise validation. Whereas for training the segmentation network, the ISIC-designated validation set of 100 samples in the segmentation dataset was used for epoch-wise validation.

4.2.1 Unbalanced Classification Dataset

After the aforementioned split, the 25331 training images of the ISIC 2019 classification dataset were divided into 20265 training images and 5066 testing images. The split was performed class-wise, so that the original proportion of samples per class is conserved in each subset of the split. The long-tailed distribution of samples in the dataset is evident in Table 1. The number of samples present in each class is not equally distributed, so they are termed as unbalanced datasets. In this work, re-weighting and SMOTE analysis were performed to balance the dataset.

4.2.2 Classification Dataset Balanced by Re-weighting

The re-weighting technique for balancing assigns weights to the classes during training based on their relative proportions in the dataset, according to equation 1. The class weights computed is listed in Table 2

Table 2 Skin lesion classes and their corresponding Weights Computed using Re-weighting Technique

Lesion Type	Computed Weights
AK	3.5
BCC	0.9
BKL	1.2
DF	12.9
MEL	0.6
NV	0.2
VC	12.8
SCC	4.9

Classification Dataset Balanced using SMOTE

To obtain this version of the dataset, the SMOTE technique was directly applied on the original training set of 25331 images in the ISIC 2019 dataset. SMOTE oversamples all minority classes to a count of 7000 images, resulting in a total of 61875 images. These images were split into a training set comprising 49500 images, and a testing set with 12375 images. The split was performed class-wise to conserve the relative proportions of samples per class.

Segmentation Dataset

The 2594 training samples in the ISIC 2018 segmentation dataset were split in 80-20 ratio to obtain 2075 training and 519 testing images. The ISIC-designated validation set of 100 samples were used for epoch-wise validation during training experiments on the segmentation networks.

4.3 Experiments for Lesion Segmentation

This research work compares the classification performance on lesion-segmented and entire dermoscopic images. However, the classification datasets (unbalanced and balanced, both) do not contain segmentation masks to obtain the lesion regions. Hence, the segmentation dataset, that contains 2594 dermoscopic images with lesion segmentation masks, is used to train a segmentation model. This trained segmentation model is later used to predict lesion masks for all the images in the classification datasets .

To validate the segmentation masks predicted by the trained model, 50 representative images were chosen from each class of the ISIC 2019 classification dataset. Segmentation masks for these 400 samples were then prepared by three medical experts independently, to obtain three *annotated segmentation masks* for each sample. An inter-rater agreement was then computed based on regional overlap using the Generalized Conformity Index (GCI) metric [28] as presented in equation 2,

$$GCI = \frac{\sum_{pairs(i>j)} X_i \cap X_j}{\sum_{pairs(i>j)} X_i \cup X_j} \tag{2}$$

where X_i represents the set of nuclear region pixels of the segmented ground truth obtained from rater i and $\sum_{pairs(i>j)}$ ... represents the summation over every unique pair of raters.

A GCI score of 0.93 was obtained across the three raters. A bitwise OR operation was performed, thereupon, across the three *annotated segmentation masks* per sample to obtain a single *ground truth mask*. The OR operation ensures that any pixel that is marked as a lesion by at least one expert is treated as a lesion in the mask of *ground truth*. These 400 sets, each containing a dermoscopic image and its ground truth mask, is used as the testing set for the segmentation model.

After mask generation, the BCDU-Net architecture is used for segmentation and trained under two different settings, namely the baseline (simple supervised learning) and the generator-discriminator (CGAN method) settings. In the baseline setting, the model converged after 120 epochs with a validation F1-Score of 0.87 for segmentation. Whereas in the CGAN setting, that uses a pixel discriminator, the model converged in 100 epochs and obtained a validation F1-Score of 0.91. The segmentation performances, evaluated on the expert annotated test set of 400 annotated samples, are compared in Table 3. A qualitative comparison of the performance of the two models is presented in Figure 4. The CGAN setting clearly outperforms the baseline due to the discriminator's feedback to the generator during each cycle of model training propagation. Hence, for all classification experiments, the CGAN generator model was used to predict the lesion masks because the Conditional Adversarial Segmentation approach combines the power of deep learning for image segmentation with the contextual awareness provided by conditional adversarial networks. This integration aims to enhance the precision and context-awareness of segmentation models, making them more suitable for complex segmentation tasks.

4.4 Experiments for Lesion Classification

The performance of the classification architecture trained using Entire Dermoscopic Images (EDI), and Lesion Segmented Dermoscopic Images (LSDI) were evaluated using different experimental setup to achieve the following objectives:

- To ascertain the classifier performance trained on the unbalanced dataset
- To ascertain the classifier performance trained on the dataset balanced by re-weighting
- To ascertain the classifier performance trained on the dataset balanced using SMOTE

4.4.1 Classification of Entire Dermoscopic Images (EDI)

Experiments were performed to ascertain the classifier's performance in classifying entire dermoscopic images. Since some of the classes in the dataset have a disproportionate sample size, the performances of the model before balancing the dataset, after balancing using SMOTE and re-weighting techniques are compared.

Classifier Trained on Unbalanced Data

An EfficientNetB2 classifier model was trained for 50 epochs using dataset. A learning rate of $1e-03$ was used and a validation accuracy of 98.4%, and testing accuracy of 80.3% was obtained.

Classifier Trained on Balanced Data using SMOTE

Class imbalanced dataset is a major hurdle which leads to biased model performance, incorrect classification of minority classes and this challenges can be addressed with SMOTE method that synthetically generates samples for minority class. In this work, the dataset was first balanced using SMOTE and then classified using EfficientNetB2 classifier model. The model was trained for 50 epochs at a learning rate of $1e^{-}03$, using the balanced dataset. A validation accuracy of 97.4%, and testing accuracy of 89.1% were obtained and the results were compared.

Classifier Trained on Balanced Data using Re-weighting

Re-weighting is a popular method used for class balancing which ensures better accuracy. This method effectively re-balances the loss by assigning weights for different classes. After balancing, an EfficientNetB2 model was trained with the class weights described in Section 4.2 for 50 epochs at a learning rate of $1e-03$, to obtain a validation accuracy of 95.4%, and a testing accuracy of 79.9%.

The class-wise comparison between unbalanced and balanced using the SMOTE and re-weighting techniques is presented in Table 4.

From Table 4, it is apparent that the NV class has the highest number of samples, whereas VASC has the lowest number of samples. Therefore, NV attains the highest accuracy of 0.91% without balancing and VASC obtained a lesser accuracy comparatively. When comparing the SMOTE and re-weighting methods, SMOTE accomplished better performance than re-weighting.

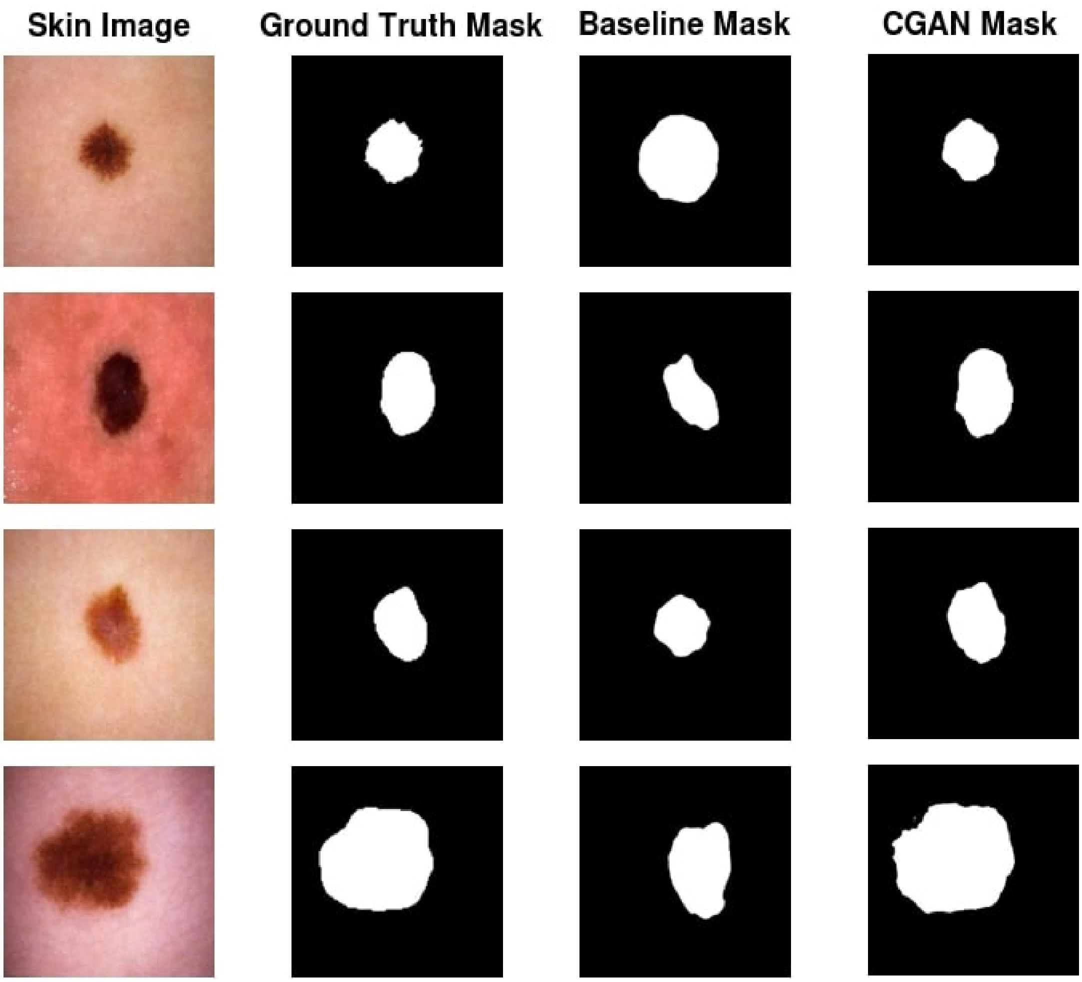

Fig. 4 Qualitative comparison of segmentation performance trained under the baseline and CGAN settings.

Table 3 Quantitative comparison of segmentation performance trained under the baseline and CGAN settings. Evaluated on expert-annotated testing samples.

Method	Accuracy	Specificity	Sensitivity	Precision	f1-score
BCDU-Net: Baseline Setting	0.92	0.95	0.85	0.86	0.86
BCDU-Net + Pixel Discriminator: CGAN Setting	0.95	0.96	0.89	0.89	0.89

4.4.2 Classification of Lesion Segmented Dermoscopic Images

Experiments were performed to ascertain the classifier's performance in classifying lesion segmented dermoscopic images. Segmented skin lesions are obtained using segmentation masks predicted by a previously trained BCDU-Net based segmentation model and the CGAN setting. Since the segmentation model achieved better results when trained under the CGAN setting, this version of the model was used to produce the segmentation masks for all classification experiments. The performance of the model before balancing the dataset and after balancing using SMOTE and re-weighting technique is compared and presented. Figures 5 and

Table 4 Comparison between unbalanced and balanced dataset using SMOTE and Re-Weighting Technique

Class -es	Unbalanced Classification				SMOTE				Re-weighting			
	Acc	Pre	Rec	F1-S	Acc	Pre	Rec	F1-S	Acc	Pre	Rec	F1-S
AK	0.57	0.61	0.57	0.59	0.90	0.91	0.90	0.90	0.51	0.50	0.51	0.51
BCC	0.86	0.73	0.86	0.79	0.83	0.85	0.83	0.84	0.70	0.84	0.70	0.71
BKL	0.68	0.62	0.68	0.65	0.81	0.85	0.81	0.83	0.68	0.69	0.68	0.69
VASC	0.68	0.88	0.68	0.77	1.00	0.98	1.00	0.99	0.82	0.88	0.82	0.85
SCC	0.56	0.75	0.56	0.64	0.96	0.89	0.96	0.92	0.61	0.75	0.61	0.67
MEL	0.60	0.79	0.60	0.68	0.70	0.84	0.70	0.76	0.71	0.71	0.71	0.71
NV	0.91	0.87	0.91	0.89	0.93	0.87	0.93	0.90	0.91	0.86	0.91	0.89
DF	0.57	0.86	0.57	0.68	0.99	0.95	0.99	0.97	0.64	0.78	0.64	0.70

6 show the sample predictions from the ISIC2019 dataset.

Classifier Trained on Unbalanced Data

An EfficientNetB2 classifier model was trained using dataset. It converged after 50 epochs at a learning rate of $1e - 03$, and attained a validation accuracy of 99. 3%, and a testing accuracy of 79.6%.

Classifier Trained on Balanced Data using Re-weighting

An EfficientNetB2 model was trained with the class weights described in Section 4.2 for 50 epochs at a learning rate of $1e-03$, to obtain a validation accuracy of 88.7%, and a testing accuracy of 74.3%.

Classifier Trained on Balanced Data using SMOTE

An EfficientNetB2 model was trained for 50 epochs at a learning rate of $1e-03$, using dataset. A validation accuracy of 98. 6% and a testing accuracy of 91. 1% were obtained.

Tables 5 and 6 present class based results of the classifier model trained with segmented and unsegmented images before and after balancing the dataset. Table 5 shows that VASC, SCC, MEL and DF produced accuracy of 0.71, 0.58, 0.67 and 0.59 respectively which is a improved accuracy for segmented classification over unsegmented classification. This improvement in accuracy is due to the fact that the lesion and the background region are highly similar. So segmenting the lesion region and classifying based on the ROI showed good accuracy for the corresponding classes.

From Table 6 it is evident that segmenting the lesion region and balancing the dataset using SMOTE, produced very high accuracy when compared to unbalanced dataset present in Table 5. For the classes AK, VKL, SCC, MEL and DF showed a very high improvement in terms of accuracy from 0.54 ,0.64, 0.58, 0.67 and 0.59 to 0.91, 0.83, 0.91, 0.84 and 0.99 respectively because segmenting and balancing the dataset.

The comparison between segmented, unsegmented and balancing using SMOTE and reweighting technique is shown in Table 7. From the results it is visible segmenting the lesion from the images and balancing the data set using SMOTE technique outperforms for all the classes.

5 Comparison with Existing Work

We have compared our results in terms of precision, recall, F1_score and class average and contrasted them with the results of previous investigations in the literature. The comparison results are presented in the following Tables 8, 9, 10 and 11.

Precision estimates the reliability of correct predictions. From our experiments it is proved that our classifiers produced improved precision for most of the classes when compared with [14], [52] and [40]. This proves our system produces high diagnostic accuracy which helps in critical decision making. A high recall rate helps in minimizing the number of false negative which could

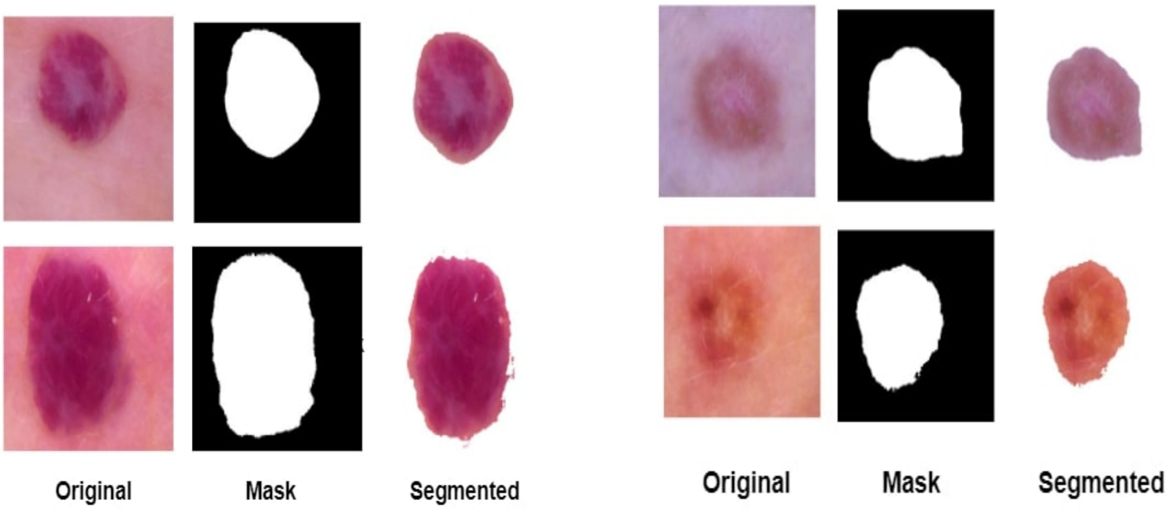

Fig. 5 **Predictions for (Left) VASC and (Right) DF**

Table 5 **Comparison between unsegmented and segmented classification (unbalanced)**

classes	Unsegmented Classification				Segmented Classification			
	Acc	Precision	Recall	F1-Score	Acc	Precision	Recall	F1-Score
AK	**0.57**	0.61	0.57	0.59	0.54	0.62	0.54	0.58
BCC	**0.86**	0.73	0.86	0.79	0.76	0.76	0.76	0.76
BKL	**0.68**	0.62	0.68	0.65	0.60	0.66	60	0.63
VASC	0.68	0.88	0.68	0.77	**0.71**	0.82	0.71	0.76
SCC	0.56	0.75	0.56	0.64	**0.58**	0.68	0.58	0.63
MEL	0.60	0.79	0.60	0.68	**0.67**	0.73	0.67	0.70
NV	0.91	0.87	0.91	0.89	0.91	0.86	0.91	0.88
DF	0.57	0.86	0.57	0.68	**0.59**	0.87	0.59	0.70

Table 6 **Comparison between unsegmented and segmented classification (balanced)**

classes	Unsegmented Classification				Segmented Classification			
	Acc	Precision	Recall	F1-Score	Acc	Precision	Recall	F1-Score
AK	0.90	0.91	0.90	0.90	**0.91**	0.86	0.91	0.89
BCC	**0.83**	0.85	0.83	0.84	0.76	0.87	0.76	0.81
BKL	0.81	0.85	0.81	0.83	**0.83**	0.77	0.83	0.80
VASC	**1.00**	0.99	1.00	0.99	0.99	0.98	0.99	0.99
SCC	**0.96**	0.89	0.96	0.92	0.91	0.95	0.91	0.93
MEL	0.70	0.84	0.70	0.76	**0.84**	0.83	0.84	0.83
NV	0.93	0.87	0.93	0.90	**1.00**	1.00	1.00	1.00
DF	0.99	0.95	0.99	0.97	0.99	0.97	0.99	0.98

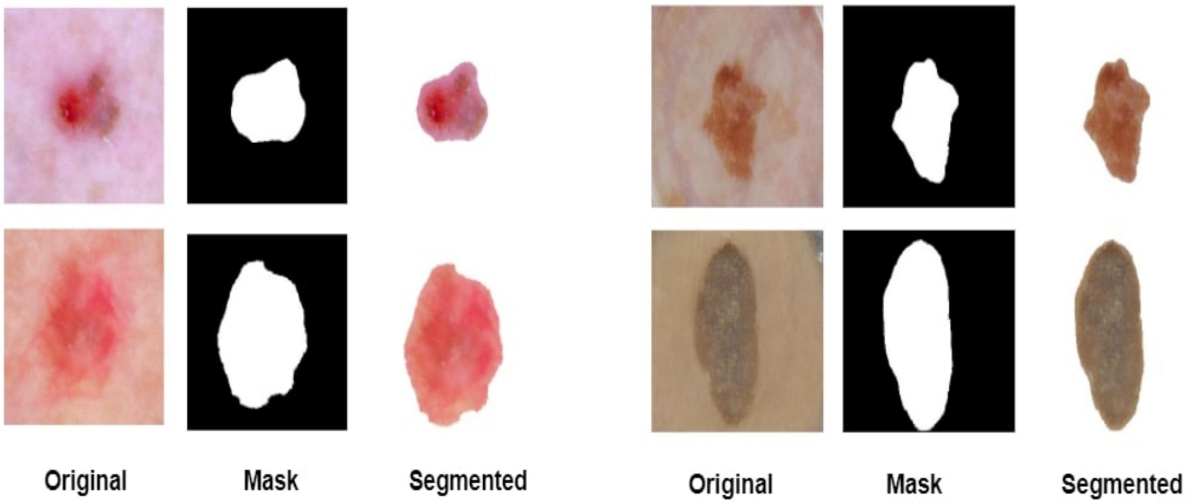

Fig. 6 Predictions for (Left) BCC and (Right) BKL

Table 7 Comparison between SMOTE and re-weighting for unsegmented and segmented classification

Classes	Unsegmented		Segmented	
	SMOTE	Re-weights	SMOTE	Re-weights
AK	**0.9**	0.51	**0.91**	0.51
BCC	**0.83**	0.70	**0.76**	0.70
BKL	**0.81**	0.68	**0.83**	0.58
VASC	**1.00**	0.82	**0.99**	0.82
SCC	**0.96**	0.61	**0.91**	0.62
MEL	0.7	**0.71**	**0.84**	0.69
NV	**0.93**	0.91	**1.00**	0.81
DF	**0.99**	0.64	**0.99**	0.57

avoid in delayed treatment or an incorrect prognosis. And our experiments have proved to produce good results in real time systems with high recall rate. From all the experiments conducted in this research work it is demonstrated that segmenting and balancing the dataset produced improved precision and recall which results in high F1-score. From Table 11 it is apparent that after taking the average of the performance metrics the proposed system produces 69% increase in average F1score.

6 Discussion

6.1 Effects of Balancing the Dataset

This section examines the effects of balancing the dataset before training the model for classification. From Table 1, it can be seen that the distribution of samples is highly disproportionate, producing a long-tailed distribution. Since deep learning methods cannot handle such heavily imbalanced datasets due to consequent bias towards classes with more samples, there is a need

Table 8 Comparison of Methods - Precision.

Method	Precision							
	AK	BCC	BKL	DF	MEL	NV	SCC	VASC
Ensemble of Multi-Res EfficientNets + SEN154 2 [14]	0.43	0.65	0.71	0.47	0.76	0.93	0.42	0.50
Ensemble of EfficienetB3-B4-Seresnext101 [52]	0.33	0.55	0.58	0.37	0.70	0.87	0.29	0.44
Ensemble, ood threshold 100% [40]	0.38	0.53	0.57	0.50	0.76	0.83	0.42	0.61
Our Proposed Methods								
Unsegmented, Unbalanced	0.61	0.73	0.62	0.86	0.79	0.87	0.75	0.88
Unsegmented, Balanced using SMOTE	**0.91**	0.85	**0.85**	0.95	**0.84**	0.87	0.89	0.98
Segmented, Unbalanced	0.62	0.76	0.66	0.87	0.73	0.86	0.68	0.82
Segmented, Balanced using SMOTE	0.86	**0.87**	0.77	**0.97**	0.83	**1.00**	**0.95**	**0.98**

to balance the dataset. This is supported by the empirical results obtained in this work, shown in Table 4. Classifiers were trained on both unbalanced and balanced datasets, and the accuracy of classification is significantly better when using the balanced datasets.

Furthermore, the method used to balance the dataset plays an important role in determining the model performance. It is clear from Table 7 that the re-weighting technique does little to improve the performance. In fact, the classifier performance is worse than that obtained using the unbalanced dataset for some of the classes. On the contrary, the SMOTE technique has been very effective in improving classifier performance. This is because the re-weighting technique only changes the priority towards the classes by assigning balanced class weights for each class, following equation 1. Clearly, the bias towards high sample size classes is not eliminated by such a simplistic relation to repeatedly propagate samples from the underrepresented classes. The SMOTE technique that generates new synthetic samples by oversampling minority classes, provides a complex enough balancing strategy to notably diminish the inter-class bias.

SMOTE stands out as a computationally less intensive alternative compared to reweighting methods. Reweighting methods involve adjusting the weights assigned to instances within the dataset to balance the influence of the minority and majority classes during model training resulting in high computational cost. Recalculating probabilities and updating model parameters for each instance can be resource-intensive, especially when dealing with large datasets. The efficiency of SMOTE lies in its ability to augment the dataset with synthetic samples, allowing the model to learn the underlying patterns of the minority class without a substantial increase in computational demands. SMOTE offers a practical and computationally efficient solution for handling imbalanced datasets and considered to be a valuable tool for addressing class imbalance without compromising efficiency.

The other most popular methods for balancing datasets are Random Under-sampling and Random Over-sampling. Random Under-sampling discards instances from the majority class, potentially leading to loss of valuable information while Random Over-sampling duplicates instances, which may not contribute new insights

Table 9 Comparison of Methods - Recall.

Method	Recall							
	AK	BCC	BKL	DF	MEL	NV	SCC	VASC
Ensemble of Multi-Res EfficientNets + SEN154 2 [14]	0.48	0.72	0.39	0.58	0.59	0.71	0.44	0.64
Ensemble of EfficienetB3-B4-Seresnext101 [52]	0.60	0.84	0.56	0.68	0.68	0.81	0.52	0.60
Ensemble, ood threshold 100% [40]	0.32	0.85	0.55	0.57	0.68	0.87	0.45	0.56
Our Proposed Methods								
Unsegmented, Unbalanced	0.57	**0.86**	0.68	0.57	0.60	0.91	0.56	0.68
Unsegmented, Balanced using SMOTE	0.90	0.83	0.81	0.99	0.70	0.93	**0.96**	1.00
Segmented, Unbalanced	0.54	0.76	0.60	0.59	0.67	0.91	0.58	0.71
Segmented, Balanced using SMOTE	**0.91**	0.76	**0.83**	**0.99**	**0.84**	**1.00**	0.91	**1.00**

and can lead to overfitting. Hence these methods may not contribute to improved generalization because Random Under-sampling may removes instances without considering their importance, and Random Over-sampling duplicates existing information. Also hese methods may not be as computationally efficient, especially in the case of Random Over-sampling, which increases the dataset size.

6.2 Effects of Segmenting the Skin Lesions

The original unsegmented dataset comprises irrelevant information that does not help in discerning between lesion types. The nonlesion portions are predominantly just regions of clear skin, but on occasion, they also include other objects such as reference stickers and marker annotations. When comparing the classification results using entire lesion segmented dermoscopic images (Table 6), it is evident that classifiers trained using entire lesion segmented images consistently outperform their entire image counterparts. And this observation holds for both balanced, as well as unbalanced datasets. Hence, the non-lesion regions act as artifacts that negatively impact classification accuracy. The classifier performance significantly improves when only relevant portions of the image are supplied, since the classifier trains to learn representative features exclusively from the portions relevant for classification.

6.3 Scaling the Classification Model

Compound scaling was performed using EfficientNet variants to analyze the improvement in classifier performance at the expense of increased usage of compute resources and training time. The scaling was performed on the best experimental setting for classification with SMOTE balanced dataset and lesion segmentation. It is evident from Table 12 that the B6 and B7 configurations produce significantly better classifiers that their lower versions, suggesting that the proposed skin lesion classification pipeline scales well to improve performance. Furthermore, since B6 performs almost as well as B7, lesser resource usage is required to tune to best performance.

7 Conclusion

Accurately classifying the skin cancer type is essential to provide timely treatment, as manual identification of the lesions is highly complex. Performing a classification of extracted skin lesions

Table 10 Comparison of Methods - F1-score.

Method	F1-score							
	AK	BCC	BKL	DF	MEL	NV	SCC	VASC
Ensemble of Multi-Res EfficientNets + SEN154 2 [14]	0.45	0.68	0.51	0.52	0.67	0.86	0.43	0.56
Ensemble of EfficienetB3-B4-Seresnext101 [52]	0.42	0.66	0.57	0.48	0.69	0.84	0.37	0.51
Ensemble, ood threshold 100% [40]	0.35	0.66	0.56	0.53	0.72	0.85	0.43	0.59
Our Proposed Methods								
Unsegmented, Unbalanced	0.59	0.79	0.65	0.68	0.68	0.89	0.64	0.77
Unsegmented, Balanced using SMOTE	**0.90**	**0.84**	**0.83**	0.97	0.76	0.90	0.92	0.99
Segmented, Unbalanced	0.58	0.76	0.63	0.70	0.70	0.88	0.63	0.76
Segmented, Balanced using SMOTE	0.89	0.81	0.80	**0.98**	**0.83**	**1.00**	**0.93**	**0.99**

Table 11 Comparison of Methods - Evaluation Metrics (Class Average).

Methods	**Avg. Acc.**	**Avg. Pre**	**Avg. Recall**	**Avg. F1-score**
Ensemble of Multi-Res EfficientNets + SEN154 2 [14]	**0.92**	0.60	0.51	0.52
Ensemble of EfficienetB3-B4-Seresnext101 [52]	0.91	0.51	0.61	0.53
Ensemble, ood threshold 100% [40]	0.92	0.58	0.54	0.52
Unsegmented, Unbalanced	0.80	0.76	0.68	0.71
Unsegmented, Balanced using SMOTE	0.89	0.89	0.89	0.89
Segmented, Unbalanced	0.80	0.75	0.67	0.71
Segmented, Balanced using SMOTE	0.91	**0.90**	**0.90**	**0.90**

helps to correctly classify the type of skin cancer much easier, as it considers only the exact regions of the lesion, ignoring unwanted artifacts. So in this research work, we have proposed two different methods, the first one uses a novel pipelined approach which delineate the lesion region present in the dermoscopic images and this delineated region is further classified into different categories. The later one is a classical approach where the entire lesion image is trained for classification. In the pipeline approach the lesion regions are extracted using BCDU-Net with CGAN Setting as pixel discriminator was used and this model achieved better accuracy than the baseline setting. The dataset under the consideration

Table 12 Comparison of performances of different EfficientNet variants to compound scale the model for best classification performance.

EfficientNet Variant	**Avg. Acc.**	**Avg. Pre.**	**Avg. Recall**	**Avg. F1-score**
B2	91.03	0.90	0.90	0.90
B3	91.44	0.89	0.93	0.90
B4	92.19	0.91	0.92	0.87
B5	94.88	0.95	0.94	0.94
B6	**96.87**	**0.96**	**0.97**	**0.96**
B7	**96.68**	**0.96**	**0.97**	**0.95**

for this research work is highly imbalanced and hence the accuracy was very less. To improve the accuracy we have used the balancing techniques using SMOTE and Re-Weighting. To analyze the efficacy of these two algorithms extensive experiments were conducted and found that the classification performance was high using pipelined approach with SMOTE balancing method when compared to classical approach.

Declarations

- Ethical Approval: The data used in this study is publicly available in https://challenge.isic-archive.com/data/#2018 and https://challenge.isic-archive.com/data/#2019.
- Competing interests: The authors have no competing interests to declare that are relevant to the content of this article.
- Funding: No funding was received for this research work, or for the preparation of this manuscript.
- Data Availability: Data supporting the findings of this study are available from the ISIC challenge websites (2018 and 2019) at https://challenge.isic-archive.com/data/#2018 and https://challenge.isic-archive.com/data/#2019.
- Code Availability: The implementation of this research work will be made available at https://github.com/karthik-d/lesion-characterization-using-cgan.
- Compute Resources Used: The implementation of these experiments was carried out using the specifications as follows: a CPU of Intel(R) Xeon(R) 2.20GHz, a Cuda enabled Tesla T4 and Tesla P100 GPU with a RAM size of 25GB and a TPU with a RAM size of 35GB of NVIDIA-SMI 460.32.03 for effective processing of a larger image dataset.
- Authors' contributions: Experiments and Implementation: Karthik Desingu, Aswatha S, Deepika R, and Deepika V; Validation: Mirunalini P, Jaisakthi S M, and Karthik Desingu; Manuscript preparation, review, and editing: Mirunalini P, Jaisakthi S M, and Karthik Desingu. All authors have read and agreed to the published version of the manuscript.

References

[1] Abbasi NR, Shaw HM, Rigel DS, et al (2004) Early diagnosis of cutaneous melanoma: revisiting the abcd criteria. JAMA 292 22:2771–6. URL https://api.semanticscholar.org/CorpusID:3894203

[2] Al-Masni MA, Kim DH, Kim TS (2020) Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. Computer methods and programs in biomedicine 190:105351

[3] Azad R, Asadi-Aghbolaghi M, Fathy M, et al (2019) Bi-directional convlstm u-net with densley connected convolutions. In: Proceedings of the IEEE/CVF international conference on computer vision workshops, pp 0–0

[4] Barata C, Celebi ME, Marques JS (2019) A survey of feature extraction in dermoscopy image analysis of skin cancer. IEEE Journal of Biomedical and Health Informatics 23(3):1096–1109. https://doi.org/10.

1109/JBHI.2018.2845939

[5] Bi L, Feng DD, Fulham M, et al (2020) Multi-label classification of multi-modality skin lesion via hyper-connected convolutional neural network. Pattern Recognition 107:107502

[6] Celebi ME, Aslandogan YA, Stoecker WV, et al (2007) Unsupervised border detection in dermoscopy images. Skin Research and Technology 13. URL https://api.semanticscholar.org/CorpusID:12772019

[7] Celebi ME, Wen Q, Hwang S, et al (2013) Lesion border detection in dermoscopy images using ensembles of thresholding methods. Skin Research and Technology 19. URL https://api.semanticscholar.org/CorpusID:414881

[8] Chawla NV, Bowyer KW, Hall LO, et al (2002) Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research 16:321–357

[9] Delgado-Gómez D, Butakoff C, Ersbøll BK, et al (2008) Independent histogram pursuit for segmentation of skin lesions. IEEE Transactions on Biomedical Engineering 55:157–161. URL https://api.semanticscholar.org/CorpusID:5564489

[10] Desingu K, Palaniappan M, Kumar J (2021) Snake species classification using transfer learning. Conference and Labs of the Evaluation Forum 2936:1571–1578. URL https://ceur-ws.org/Vol-2936/paper-133.pdf

[11] Desingu K, Bhaskar A, Palaniappan M, et al (2022) Classification of fungi species: A deep learning based image feature extraction and gradient boosting ensemble approach. Conference and Labs of the Evaluation Forum URL https://ceur-ws.org/Vol-3180/paper-162.pdf

[12] Erkol B, Moss RH, Stanley RJ, et al (2005) Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. Skin Research and Technology 11. URL https://api.semanticscholar.org/CorpusID:24884818

[13] Gessert N, Nielsen M, Shaikh M, et al (2020) Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. MethodsX 7:100864. https://doi.org/https://doi.org/10.1016/j.mex.2020.100864, URL https://www.sciencedirect.com/science/article/pii/S2215016120300832

[14] Gessert N, Nielsen M, Shaikh M, et al (2020) Skin lesion classification using ensembles of multi-resolution efficientnets with meta data. MethodsX 7:100864

[15] Gonzalez-Diaz I (2018) Dermaknet: Incorporating the knowledge of dermatologists to convolutional neural networks for skin lesion diagnosis. IEEE journal of biomedical and health informatics 23(2):547–559

[16] Goyal M, Oakley A, Bansal P, et al (2020) Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. IEEE Access 8:4171–4181. https://doi.org/10.1109/ACCESS.2019.2960504

[17] Haenssle H, Fink C, Schneiderbauer R, et al (2018) Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of oncology : official journal of the European Society for Medical Oncology 29

[18] Hasan SN, Gezer M, Azeez RA, et al (2019) Skin lesion segmentation by using deep learning techniques. In: 2019 Medical Technologies Congress (TIPTEKNO), IEEE, pp 1–4

[19] Innani S, Dutande P, Baid U, et al (2023) Generative adversarial networks based skin lesion segmentation. Scientific Reports 13(1):13467

[20] Izadi S, Mirikharaji Z, Kawahara J, et al (2018) Generative adversarial networks to segment skin lesions. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp 881–884, https://doi.org/10.1109/ISBI.2018.8363712

[21] Jafari MH, Karimi N, Nasr-Esfahani E, et al (2016) Skin lesion segmentation in clinical images using deep learning. In: 2016 23rd International conference on pattern recognition (ICPR), IEEE, pp 337–342

[22] Jaisakthi S, Desingu K, Mirunalini P, et al (2023) A deep learning approach for nucleus segmentation and tumor classification from lung histopathological images. Network Modeling Analysis in Health Informatics and Bioinformatics 12(1):22

[23] Jaisakthi SM, Mirunalini P, Aravindan C (2018) Automated skin lesion segmentation of dermoscopic images using grabcut and k-means algorithms. IET Computer Vision 12(8):1088–1095

[24] Jaisakthi SM, Mirunalini P, Aravindan C (2018) Automated skin lesion segmentation of dermoscopic images using grabcut and k-means algorithms. IET Computer Vision 12(8):1088–1095. https://doi.org/https://doi.org/10.1049/iet-cvi.2018.5289, URL https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-cvi.2018.5289, https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-cvi.2018.5289

[25] Jensen JD, Elewski B (2015) The abcdef rule: Combining the "abcde rule" and the "ugly duckling sign" in an effort to improve patient self-screening examinations. The Journal of clinical and aesthetic dermatology 8 2:15. URL https://api.semanticscholar.org/CorpusID:7871524

[26] Kassem MA, Hosny KM, Fouad MM (2020) Skin lesions classification into eight classes for isic 2019 using deep convolutional neural network and transfer learning. IEEE Access 8:114822–114832. https://doi.org/10.1109/ACCESS.2020.3003890

[27] Kassem MA, Hosny KM, Fouad MM (2020) Skin lesions classification into eight classes for isic 2019 using deep convolutional neural network and transfer learning. IEEE Access 8:114822–114832

[28] Kouwenhoven E, Giezen M, Struikmans H (2009) Measuring the similarity of target volume delineations independent of the number of observers. Physics in Medicine & Biology 54(9):2863. https://doi.org/10.1088/0031-9155/54/9/018, URL https://dx.doi.org/10.1088/0031-9155/54/9/018

[29] Lazo C (2021) Segmentation of skin lesions and their attributes using generative adversarial networks. arXiv preprint arXiv:210200169

[30] Li Y, Shen L (2018) Skin lesion analysis towards melanoma detection using deep learning network. Sensors 18(2):556

[31] Liu L, Tsui YY, Mandal M (2021) Skin lesion segmentation using deep learning with auxiliary task. Journal of Imaging 7(4):67

[32] Mahbod A, Schaefer G, Wang C, et al (2020) Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. Computer Methods and Programs in Biomedicine 193:105475

[33] Mahbod A, Tschandl P, Langs G, et al (2020) The effects of skin lesion segmentation on the performance of dermatoscopic image classification. Computer Methods and Programs in Biomedicine 197:105725

[34] Mirikharaji Z, Abhishek K, Bissoto A, et al (2023) A survey on deep learning for skin lesion segmentation. Medical Image Analysis 88:102863. https://doi.org/https://doi.org/10.1016/j.media.2023.102863, URL https://www.sciencedirect.com/science/article/pii/S1361841523001238

[35] Mohamed EH, El-Behaidy WH (2019) Enhanced skin lesions classification using deep convolutional networks. In: 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS), pp 180–188, https://doi.org/10.1109/ICICIS46948.2019.9014823

[36] Nachbar F, Stolz W, Merkle T, et al (1994) The abcd rule of dermatoscopy. high

prospective value in the diagnosis of doubtful melanocytic skin lesions. Journal of the American Academy of Dermatology 30 4:551–9. URL https://api.semanticscholar.org/CorpusID:4860343

[37] Palaniappan M, Desingu K, Bharathi H, et al (2022) Deep learning and gradient boosting ensembles for classification of snake species. Conference and Labs of the Evaluation Forum URL https://ceur-ws.org/Vol-3180/paper-173.pdf

[38] Pizer SM, Amburn EP, Austin JD, et al (1987) Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing 39(3):355–368. https://doi.org/https://doi.org/10.1016/S0734-189X(87)80186-X, URL https://www.sciencedirect.com/science/article/pii/S0734189X8780186X

[39] Pollastri F, Bolelli F, Palacios RP, et al (2018) Improving skin lesion segmentation with generative adversarial networks. In: 2018 IEEE 31St International symposium on computer-based medical systems (CBMS), IEEE, pp 442–443

[40] Pollastri F, Maroñas J, et al (2019) Aimagelab-prhlt at isic challenge 2019. AImageLab, Tech Rep

[41] Shah V, Autee P, Sonawane P (2020) Detection of melanoma from skin lesion images using deep learning techniques. In: 2020 International Conference on Data Science and Engineering (ICDSE), IEEE, pp 1–8

[42] Shah V, Autee P, Sonawane P (2020) Detection of melanoma from skin lesion images using deep learning techniques. In: 2020 International Conference on Data Science and Engineering (ICDSE), pp 1–8, https://doi.org/10.1109/ICDSE50459.2020.9310131

[43] Sivanesan U, Braga LH, Sonnadara RR, et al (2019) Unsupervised medical image segmentation with adversarial networks: From edge diagrams to segmentation maps. arXiv preprint arXiv:191105140

[44] Song Y, Cai W, Huang H, et al (2015) Large margin local estimate with applications to medical image classification. IEEE Transactions on Medical Imaging 34(6):1362–1377. https://doi.org/10.1109/TMI.2015.2393954

[45] Suganyadevi S, Seethalakshmi V, Balasamy K (2022) A review on deep learning in medical image analysis. International Journal of Multimedia Information Retrieval 11(1):19–38

[46] Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning, PMLR, pp 6105–6114

[47] Trang K, Nguyen HA, TonThat L, et al (2022) An ensemble voting method of pre-trained deep learning models for skin disease identification. In: 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), pp 445–450, https://doi.org/10.1109/CyberneticsCom55287.2022.9865634

[48] Venkataramana L, Prasad DVV, Saraswathi S, et al (2022) Classification of covid-19 from tuberculosis and pneumonia using deep learning techniques. Medical & Biological Engineering & Computing 60(9):2681–2691

[49] Vestergaard M, Macaskill P, Holt P, et al (2008) Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. British Journal of Dermatology 159(3):669–676

[50] Yu L, Chen H, Dou Q, et al (2016) Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE transactions on medical imaging 36(4):994–1004

[51] Zhang J, Xie Y, Xia Y, et al (2019) Attention residual learning for skin lesion classification. IEEE transactions on medical imaging 38(9):2092–2103

[52] Zhou S, Zhuang Y, Meng R (2019) Multi-category skin lesion diagnosis using dermoscopy images and deep cnn ensembles. línea], ISIC Chellange

[53] Zuiderveld K (1994) Contrast Limited Adaptive Histogram Equalization, Academic Press Professional, Inc., USA, p 474–485