

AMDS: A Multistage Framework Using Deep Learning Models for Early Diagnosis of Melanoma and Non-Melanoma Skin Lesions

Afnan M. Hassan^{1,2,*}, Khaled Mahar³ and Mohamed M. Fouad¹

¹College of Computing and Information Technology, Arab Academy for Science, Technology, and Maritime Transport, Cairo, Egypt

²Cairo Higher Institute for Engineering, Computer Science and Management, Cairo, Egypt

³College of Computing and Information Technology Arab Academy for Science, Technology, and Maritime Transport, Alexandria, Egypt

Received: 1 Mar. 2024, Revised: 23 Apr. 2024, Accepted: 7 May 2024.

Published online: 1 Jul. 2024.

Abstract: Skin cancer, an aggressive cancer with a global frequency, presents an evolving public health challenge requiring novel diagnostic approaches. The traditional method for diagnosing skin cancer involves a thorough examination of tissue samples obtained from skin lesions. Healthcare professionals must perform the complex task of identifying specific early symptoms to make an accurate diagnosis. Early detection of skin cancer is particularly challenging because of the tendency for misdiagnosis due to similarities with other dermatological conditions and variations in specialist expertise. Researchers have used machine learning algorithms to improve the performance of numerous medical applications in recent years to improve the dependability, productivity, efficiency, predictability, and precision of medical diagnostics. Current research presents a Multistage Deep Learning model for Skin Cancer Classifier (AMDS), a framework designed to improve the early detection of melanoma and non-melanoma skin lesions. The AMDS consists of several crucial phases, beginning with precise preprocessing techniques to remove extraneous components surrounding skin lesions. Given the inherent imbalances within most skin cancer datasets, the subsequent stage employs Generative Adversarial Networks (GANs) to generate synthetic images for enhancing dataset diversity and equip the classifier to handle a broad spectrum of skin lesions. In the subsequent stage, an attention-based U-Net model is introduced that is capable of generating masks for regions of interest while removing background noise. The process ended with the classification stage which uses distinct forms of the cutting-edge EfficientNet, ResNet, and DenseNet architectures, carefully trained using the segmented images, to find the best model for skin lesion classification. The proposed deep learning models are systematically evaluated by utilizing the International Skin Imaging Collaboration dataset (ISIC), a dermatology benchmark. The experimental results demonstrate that the proposed framework using a modified EfficientNetV2S with attention mechanism outperforms other tested architectures as well as most recent research. Notably, it achieves a 0.96 accuracy rate, 0.91 F1- score, 0.90 recall rate, and 0.93 precision rate on the test benchmarking datasets. These results highlight the importance of the proposed multistage framework as a potential transformative instrument for early skin cancer detection.

Keywords: Skin Cancer, Deep learning, Generative Adversarial Networks, Image Segmentation, Convolutional Neural Network.

1 Introduction

Skin cancer is classified into Melanoma and Nonmelanoma which are considered among the 10 most common cancer types. It is contributing to 7.2% of new cases across the world that are suffering from cancer [1]. Despite the low percentage of melanoma compared to nonmelanoma, it causes most skin cancer deaths. Early identification of skin cancer is crucial for increasing a patient's chances of being cured. For the last five years, the survivability rate had declined from 99% in cases of local disease to 63% in situations of regional disease and 20% in cases of disease disseminated across the body [2].

In the context of clinical lesion assessment, when relying just on visual inspection by the human eye, the detection rate for melanoma is approximately 54%, which can be considered low. However, when carried out by skilled dermoscopy practitioners, the use of dermoscopy—a non-invasive method that entails the microscopic examination of pigmented skin lesions—increases the diagnostic efficacy to a notable 79%. However, it is essential to acknowledge that the

*Corresponding author e-mail: A.Hassan3109@student.aast.edu

effectiveness of dermoscopy relies heavily on human judgment. The inherent possibility of human error resulting from the participation of either unskilled or inadequately educated examiners may significantly influence the outcome [3]. Researchers have used machine learning algorithms to improve the performance of numerous medical applications in recent years [4, 5, 6]. Machine learning approaches are distinguished by their capability to analyze large amounts of data to uncover interesting correlations, give insight, and detect patterns. For many diseases, machine learning improved the dependability, efficiency, predictability, and precision of medical diagnostics. In [7], the authors use wearable technology and machine learning to control hypovolemia disorders. By applying machine learning on electroencephalogram (EEG), a record of brain activity, Parkinson's disease can be diagnosed [8]. the convolutional neural network (CNN) is used to detect brain tumors using magnetic resonance imaging (MRI) [9, 10]. CNNs have also been utilized to develop mammogram-based breast cancer detection systems [11, 12]. Regarding skin cancer, Researchers have used CNNs to classify and extract the lesion boundary from this type of cancer [13, 14]. CNN can implicitly extract the ABCDE rule, which is commonly used to identify skin cancers. These features represent the asymmetry of two halves of skin lesion, smoothness of the border, color contrast, and lesion diameter size [15]. One of the most critical challenges in machine learning is an unbalanced dataset, which causes the model to be biased toward the dominant class. Over-sampling, under-sampling, data augmentation, and class weights modification on the loss function are all used to address this issue [16, 17, 18]. The generative adversarial network (GAN) is a potential solution for over-sampling since it generates synthetic pictures for classes with fewer samples and ensures the model generalization [19]. The current research endeavors to put forth a comprehensive framework aimed at the implementation of a sophisticated skin cancer classification system. This system is cleverly made up of a powerful combination of Generative Adversarial Network (GAN), attention U-Net, EfficientNet, ResNet, and DenseNet networks. All of these networks were carefully put together to deal with the complexity of the problem and improve the ability to find and classify melanoma and non-melanoma skin lesions [20, 21, 22, 23]. Noteworthy enhancements in image quality have been harnessed through judiciously applied techniques encompassing hair removal, median filtering, and image normalization. One of the notable aspects of this complicated framework involves the utilization of the GAN model, which is effectively used to generate synthetic images of skin cancer. This strategy serves an additional goal by addressing the issue of dataset imbalance while augmenting it with a diverse range of malignant skin representations. The main contributions can be summarized as follows:

- Preprocessing of skin cancer images to improve image content and quality.
- Balancing dataset based on GAN and augmentation techniques.
- Proposing AMDS, a highly accurate skin cancer classification system composed of different stages.

Experimental results demonstrated that the proposed framework had achieved high accuracy, recall, and precision compared with recent published results as will be shown next. This paper is organized as follows. Section 2 presents the recent studies that utilize machine learning for skin cancer classification. Section 3 illustrates the techniques used to implement AMDS from preprocessing to classification and also it presents the used evaluation criteria. This is followed by Section 4 to show the results of the proposed framework. After that, Section 5 compares AMDS' best results with the recent studies. Finally, Section 6 presents the conclusions and future work.

2 Literature Review

Researchers have used deep learning for skin cancer classification. In [24], the authors utilized 1000 images from the ISIC dataset to categorize melanoma. They used ensemble learning by merging Random Forest (RF) and Support Vector Machine (SVM) classifiers. To eliminate noise, a median filter was utilized, followed by mean shift segmentation to extract ROI. For feature extraction, moments invariant features, grey Level Co-Occurrence Matrix (GLCM), and Gray Level Run Length Matrix (GLRLM) were used. The accuracy, recall, and specificity of GLCM coupled with their classifier were 0.8931, 0.8856, and 0.8781 respectively.

The HAM10000 dataset, which comprises 10015 images obtained from Australian and Austrian patients, was used in [25]. This dataset contains 6705 non-melanoma, 1113 malignant, and 2197 unknown lesions images. A noise reduction technique based on thresholds was utilized. To make the classification task easier for the classifier, low-contrast and color-illumination images were excluded from the dataset. For classification, the authors proposed four stacked CNN layers followed by a fully connected layer. On the test dataset, their approach achieved 0.9657 on precision, 0.9366 on recall, 0.9509 on F1-score, and 0.9143 on accuracy.

In [26], the authors of this study used the ISIC 2019 [27] and ISIC 2020 [28] datasets for skin cancer classification, totaling 24,225 images. To balance the dataset, the malignant class has been over-sampled. On the combined dataset, pre-trained Inception V3 achieved 0.869, 0.8747, 0.8614, 0.8766, and 0.87 on the accuracy, precision, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC).

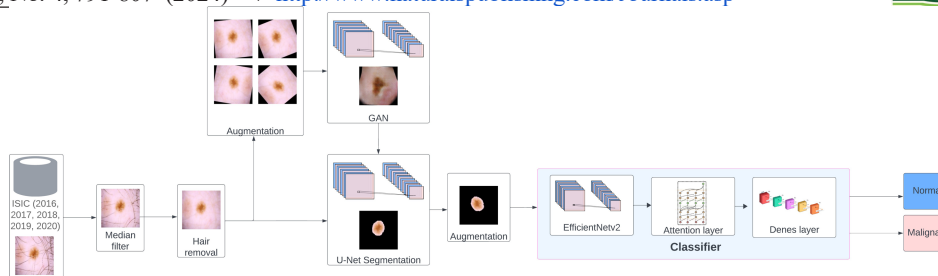


Fig. 1: Generalized block diagram of AMDS

Inception-V3 has been used for skin cancer classification achieving 0.8861, 0.86, and 0.8742 on non-melanoma accuracy, malignant accuracy, and F1-score using a balanced dataset from the ISIC archive [29]. In [30], the authors have proposed a two-staged classification method. This first stage has been implemented using R-CNN which is responsible for the segmentation stage to generate a bounding box around the skin lesion using the ISIC 2017 dataset [31]. The second stage used the ResNet152 model for classification achieving 0.925, 0.820, and 0.904 on specificity, sensitivity, and accuracy. ISIC 2019 dataset has been used to train AlexNet and GoogLeNet to classify skin cancer. GoogLeNet has outperformed AlexNet achieving an accuracy of 0.902 [32].

Inception Block Skin Network (InSiNet) and U-Net were trained using ISIC 2018 [33, 34], ISIC 2019, and ISIC 2020 datasets. Morphological techniques, as well as a Gaussian blur filter followed by the Otsu algorithm [35], were utilized in hair removal to reduce noisy regions. U-Net has achieved a dice coefficient of 0.9367, while InSiNet has achieved average accuracy, sensitivity, and specificity of 0.9234, 0.9361, and 0.913 [36].

Darkflow You Only Look Once (YOLO) network was used as a feature extractor integrated with GLCM, Gabor, and Color Level Co-occurrence Matrix (CLCM) features [37]. The concatenated features were then passed to a fully connected layer achieving an accuracy of 0.94, a precision of 0.85, a recall of 0.88, and an AUC of 0.95. They have used ISIC 2016 dataset [38] with an augmentation technique to perform oversampling.

In [39], the authors have used a dataset from ISIC archive. A hair removal algorithm is used to eliminate noise followed by Morphological Active Contours without Edges (MorphACWE) for lesion segmentation. After extracting the skin lesion, low-level features have been extracted to represent the texture, color, and lesion border. Ensemble learning using majority voting between K-nearest neighbors (kNN), SVM, and CNN has been implemented achieving an accuracy of 0.884.

GAN has been used to increase the size of the ISIC archive dataset in [40]. CNN was proposed to perform skin cancer classification with and without the proposed GAN preprocessing. The performance metrics have improved by applying oversampling using GAN achieving an accuracy of 0.71, sensitivity of 0.68, specificity of 0.74, and F1-score of 0.7.

Tailored deep neural network architecture has been proposed in [41]. The authors have implemented data balancing associated with data augmentation including random rotations, shifts, illumination correction, and contrast enhancement in order to enhance image quality and the generalization ability of their proposed model. They have designed several designs, the best model has achieved a recall of 0.928, a precision of 0.785, and an accuracy of 0.837. Table 1 shows a summary of recent related works for skin cancer classification.

3 Materials and Methods

This section presents the techniques used for implementing the multistage framework for skin cancer classification. As shown in Fig. 1, AMDS is implemented using four main stages, the first stage consists of a median filter to remove the noise from the image and a hair removal algorithm. After that, the GAN network was used to balance the dataset by generating malignant images. This stage is followed by the U-Net to extract ROI from the images. Finally, the classifier stage is implemented to classify the pre-processed image into two main classes.

Table 1: A summary of recent scholarly publications pertaining to the classification of skin cancer

Source	Model/Algorithm	Findings	Challenges/Shortcoming
[24]	Ensemble model between SVM and RF	Accuracy: 0.8931 Recall: 0.8856 Specificity: 0.8781	Classical machine learning could suffer from translation variance which highly impacts the performance
[25]	CNN	Accuracy: 0.9143 Recall: 0.9366	low-contrast and color-illumination images were excluded

		Precision: 0.9657 F1-score: 0.9509	which consequently ease the task on the CNN. Although the model could be subjected to the eliminated samples on deployment.
[26]	Inception-V3	Accuracy: 0.869 Sensitivity: 0.8614 Precision: 0.8747 Specificity: 0.8766 AUC: 0.87	The research has not used any segmentation algorithm in order to aid the model to concentrate on the tumor features which could negatively impact the performance.
[29]	Inception-V3	non-melanoma accuracy: 0.8861 Malignant accuracy: 0.86 F1-score: 0.8742	The research has not used any segmentation algorithm in order to aid the model to concentrate on the tumor features which could negatively impact the performance.
[30]	ResNet152	Accuracy: 0.904 Sensitivity: 0.82 Specificity: 0.925	The research has not used all the available ISIC dataset which could impact the performance and the generalization ability of their model.
[32]	GoogLeNet	Accuracy: 0.9202	The research has not provided performance metrics such as precession, recall, or throughput.
[36]	Inception blocks	Accuracy: 0.9234 Sensitivity: 0.9361 Specificity: 0.913	The research has not provided any complexity metrics such as throughput.
[37]	YOLO network	Accuracy: 0.94 Precision: 0.85 Recall: 0.88 AUC: 0.95	The research has not used all the available ISIC dataset which could impact the performance and the generalization ability of their model.
[39]	KNN, SVM, and CNN	Accuracy: 0.884	The research has not provided performance metrics such as precession, recall, or throughput.
[40]	GAN and CNN	Accuracy: 0.71 Sensitivity: 0.68 Specificity: 0.74 F1-score: 0.7	The research has achieved low performance compared to other approaches. The research has not provided any complexity metrics.
[41]	CNN	Accuracy: 0.837 Recall: 0.928 Precision: 0.785	The research has not provided any complexity metrics.

Table 2: Dataset distribution of non-melanoma and malignant

Dataset	non-melanoma	Malignant
ISIC 2016	727	173
ISIC 2017	1626	374
ISIC 2018	8061	1983
ISIC 2019	16619	8712
ISIC 2020	32542	584

**(a)** Image with hair**(b)** Image after hair removal**Fig. 2:** Comparison between the image before and after hair removal

3.1 Dataset

This section describes the datasets that have been used in this study which are a wide range of freely available online datasets such as the ISIC 2016, 2017, 2018, 2019, and 2020. The ISIC 2016 dataset, also known as the 2016 ISIC-ISBI challenge, contains 900 training images. This dataset is categorized into two classes which are melanoma with 173 images and non-melanoma with 727 images. The 2017 ISBI Challenge on Skin Lesion Analysis towards Melanoma Detection, is also known as the ISIC 2017 dataset. This dataset includes training data (2000 images), validation data (150 images), and a blind held-out test dataset (600 images). The training dataset is divided into three classes, 374, 254, and 1372 dermoscopic images for melanoma, seborrheic keratosis, and nevus respectively. The ISIC 2018 dataset, also known as the HAM10000 ("Human Against Machine with 10,000 training images") dataset, was divided into two parts: a training dataset of 10015 images and a test dataset of 1512 images. This dataset was driven from a retrospective sample of patients who had undergone skin cancer screening at multiple institutions using a variety of dermatoscopy techniques on all anatomic sites (except mucosa and nails). AKIEC, BCC, non-melanoma Keratosis (BKL), Dermatofibroma (DF), Melanocytic nevus (NV), Melanoma (MEL), and Vascular lesion comprise the training dataset (VASC). Each of these groups contains a different number of images. The MEL has 1113 people, the NV has 6705 people, the BCC has 514 people, and the AKIEC has 327 people, there are 1099 in the BKL, 115 in the DF, and 142 in the VASC. ISIC 2019 is made up of eight well-known classes and one for outlier images. MEL, NV, BCC, AKIEC, BKL, DF, VASC, and SCC are the classes involved. ISIC 2019 has 25,331 images, with AKIEC having 867, BKL having 2624, BCC having 3323, DF having 239, NV having 12,875, MEL having 4522, SCC having 628, and VASC having 253. Finally, ISIC 2020 came from the Hospital Clinic de Barcelona, the Medical University of Vienna, the Memorial Sloan Kettering Cancer Center, the Melanoma Institute Australia, the University of Queensland, and the University of Athens Medical School. More than 2,000 patients are included in the dataset generating 33,126 dermoscopic training images of unique non-melanoma and malignant skin lesions forms. Table 2 shows the different dataset distributions over non-melanoma and malignant classes.

3.2 Preprocessing

Various preprocessing techniques have been employed on the dataset in order to improve the quality of the images. Initially, the images were resized to a dimension of 128x128 using bilinear interpolation. This was done in order to accelerate the training process and decrease the computing cost. Subsequently, the images received a median filter application, wherein the pixel values were substituted with the median value derived from the surrounding pixels to remove the presence of outlier noise, including salt, Gaussian, and pepper noises [42].

The presence of hair in nearby areas of the lesions may potentially limit the visualization of colored lesions. Consequently, the image received a hair removal algorithm, which involved the initial conversion of the colored image to grayscale format. Subsequently, the morphological black-hat transformation was applied to the grayscale image, followed by the application of the inpainting algorithm to the original image. The mask utilized for the inpainting algorithm was created from the grayscale image [43]. Figure 2 illustrates the impact of the hair removal algorithm on a representative image sample. Image augmentation has also been utilized to enhance the generalization potential of deep learning models, resulting in improved performance [44]. Rotation transformations using randomly generated angles have been employed, including random picture flipping, encompassing both horizontal and vertical flipping.

Balancing the dataset is significant in data preprocessing if the dataset is heavily unbalanced, like in this case study. GAN is used to generate synthetic images of malignant cases which are the minority class to balance the dataset. GAN is made up of two primary networks: the generator, which generates a realistic image based on random inputs, and the discriminator, which judges the realistic probability of the input image. The training of generator and discriminant networks takes place concurrently as shown in Fig. 3. The discriminant network distinguishes whether the incoming image is fake or not. The generator network, on the other hand, attempts to produce more real images to mislead the discriminator network. The reliability of a trained GAN is determined by its ability to mislead the discriminator network with newly generated images from the generator network. The generator proposed in this study has four transposed convolutional layers with 128, 256, 512, and 512 as the number of filters respectively, that utilize the Leaky Rectified Linear Unit (LReLU) as the activation function. Leaky Rectified Linear Unit (LReLU) has been employed as a strategy to mitigate the issues of disappearing gradients and the occurrence of dying Rectified Linear Unit (ReLU) problems, as referenced in [45]. Subsequent to these layers, there is a single convolutional layer that employs the sigmoid activation function. In contrast, the discriminant network is comprised of five convolutional layers with 64, 128, 128, 256, and 512 as the number of filters respectively, utilizing the Leaky Rectified Linear Unit (LReLU) as an activation function. This is then followed by a dense layer for classification, employing the sigmoid function as its activation function.

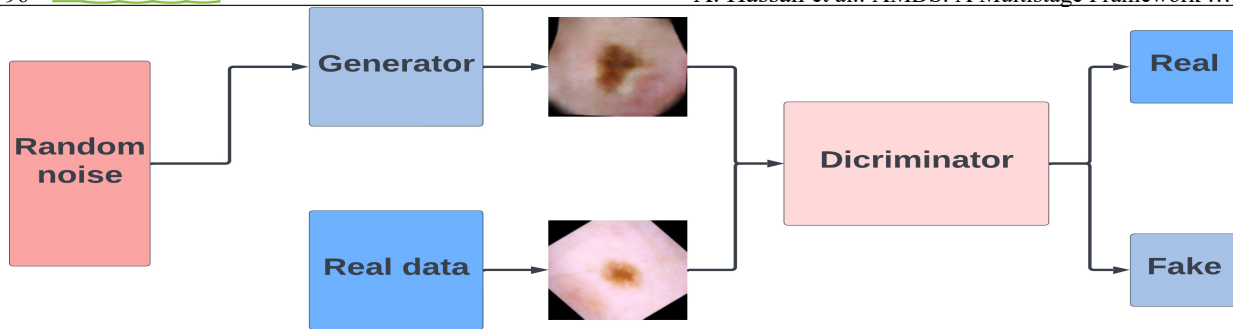


Fig. 3: GAN block diagram

Image segmentation can be implemented using several techniques such as thresholding, region growing, Markov random field models, and clustering using k-means [46]. The previously mentioned techniques can be sensitive to noise and intensity inhomogeneities which cannot generalize well on different patterns for the same object. CNN, on the other hand, utilizes weight sharing, memory savings, independence from local changes in the image, and performance improvement as data size increases [47]. As a consequence, the U-Net architecture, which consists of stacked layers of CNN, has been applied to image segmentation since it achieves high performance in a variety of biomedical applications [48, 49]. The U-Net architecture is composed of two main components: an encoder and a decoder. The encoder architecture is tasked with capturing important patterns found in the input images, while the decoder architecture is responsible for reconstructing the input from the encoder into a binary mask. This binary mask is then used to segment the original images through a multiplication operation. The fundamental component of the suggested encoder architecture consists of a convolutional layer, which is subsequently followed by a batch normalizing layer. The block completed a cascading process 18 times, wherein the number of filters in the convolution layer was incrementally increased every six blocks, ranging from 64 to 256 filters. The decoder architecture presented in this study comprises an upsampling layer followed by a convolution layer. A two-dimensional attention block has been incorporated between the up-sampling and matching encoding layers. After the upsampling layer, a sequence of six convolution layers has been incorporated, followed by a subsequent batch normalizing layer. The process of upsampling involved the repetition of the upsampling layer, along with the six convolution layers, until the original input dimensions were achieved. The ReLU activation function was employed for all layers except the output layer. In order to obtain a binary output mask, the sigmoid activation function was utilized. The attention mechanism was initially proposed as a means to enhance the efficiency of the encoder-decoder model in the context of machine translation [50]. The proposed model incorporates the attention mechanism to enhance the performance of the network since the structure of the machine translation encoder-decoder closely resembles the U-Net design at the highest level. Figure 4 illustrates the U-Net architecture.

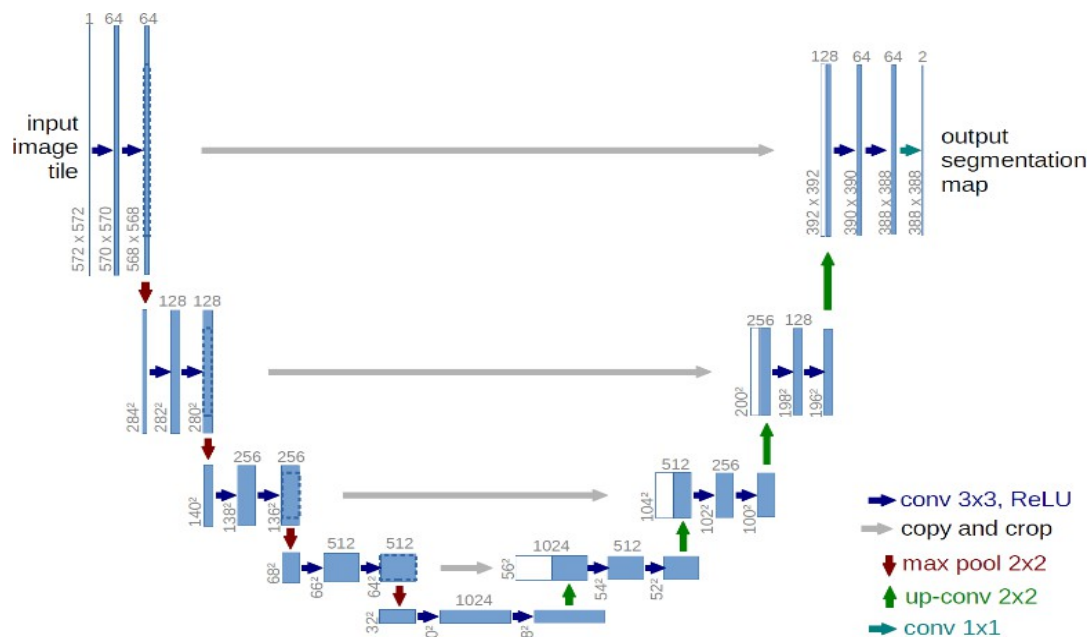


Fig. 4: The U-Net architecture [20]

3.3 Classification

The proposed classifiers are based on different versions of pre-trained EfficientNet, ResNet, and DenseNet networks [21, 22, 23]. ResNet has introduced the skip layer which is contributing to increasing network depth without suffering from vanishing gradient issues. Two different networks from ResNet architecture have been used as classifiers which are ResNet101v2 and ResNet152v2 having depth of 205 and 307 respectively. DenseNet architecture has introduced the principle of concatenating the feature maps of all previous layers allowing access to the features of all preceding levels. This principle allows the network to gain knowledge more effectively. Two different networks from DenseNet architecture have been used as classifiers which are DenseNet169 and DenseNet201 having depths of 338 and 402 respectively. EfficientNet depends on the principle of compound model scaling. It scales the width, depth, and resolution of the network with a certain fixed set of scaling coefficients using AutoML [51]. Two different networks from EfficientNet architecture have been used as classifiers which are EfficientNetV2B2 and EfficientNetV2S. EfficientNet architecture has been modified by adding an attention layer instead of receiving the output on global average pooling or flattening layer. The attention layer has been used to summarize the output of the EfficientNet giving large weights to the significant output filters.

3.4 Evaluation

Evaluation metrics are employed in the machine learning pipeline to evaluate each machine learning algorithm's ability to model the provided data using accuracy and complexity. Accuracy metrics measure the ability of each model to classify data samples correctly. After model training, this model must be tested on unseen data to ensure its generalization ability. A confusion matrix is generated by the trained model on the test dataset which consists of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). TP was defined as the number of malignant samples that were correctly classified, while FP was defined as the number of normal samples that were incorrectly classified as malignant. TN was defined as the number of normal samples that were correctly classified, finally, the FN was defined as the number of malignant samples that were incorrectly classified as normal. The accuracy (AC), precision (P), recall (R), and F1-score (F) of each class were calculated from TP, FP, TN, and FN. The AC is defined as the ratio between correctly classified samples to the domain of all samples. The precision for a given class is defined as the ratio between the correctly classified samples of that class to the domain of the predicted samples for that class. Recall for a given class is defined as the ratio between the correctly classified samples of that class to the domain of actual samples for that class. F1-score is defined as the harmonic mean of P and R. Table 3 shows the mathematical equations of AC, P, R, F1-score. Another type of accuracy evaluation metric is AUC which gives us the accumulated performance over all the possible classification thresholds which makes it scale and threshold invariant since it is not a function of threshold.

Table 3: Mathematical equations of evaluation metrics

Metric	Equation
Accuracy	$TP + TN$
	$TP + TN + FP + FN$
Precision of normal class	TN
	$TN + FN$
Precision of malignant class	TP
	$TP + FP$
Recall of normal class	TN
	$TN + FP$
Recall of malignant class	TP
	$TP + FN$
F1-score	$\frac{2 * P * R}{P + R}$
	$P + R$

4 Results

The proposed GAN was trained and tested on a total number of 2,538 malignant images from ISIC 2016, ISIC 2019, and ISIC 2020 using binary cross entropy as a loss function along with adaptive moment estimation (Adam) optimizer with a learning rate of 0.001. Adam provides faster convergence and a smaller number of hyperparameters to tune compared to other optimizers [52]. The input images have been subjected to a median filter, hair removal algorithm, and augmentation using rotation transformations. The proposed GAN comprises 7,252,867 and 3,027,009 with respect to the number of parameters on the generator and discriminator respectively. Fig. 5 shows the images generated by the

proposed GAN and Fig. 6 shows the generator loss and discriminator loss across epochs achieving 0.677127 and 0.845434 on discriminator and generator losses respectively.

U-Net has also been evaluated using binary cross-entropy loss in addition to the dice coefficient. Fig. 7 shows the descent of training and validation loss across the number of epochs. The proposed U-Net has achieved 0.967 and 0.033 on dice coefficient and loss respectively. The proposed U-Net has been trained and tested on 5,494 images from ISIC 2016, ISIC 2017, and ISIC 2018 with a total number of parameters of 5,904,291. This dataset has been split to form the training dataset and test dataset with ratios of 0.9 and 0.1 respectively. Fig. 8 shows examples of generated masks by U-Net compared to their ground truth binary masks on the dataset.

Table 4 shows the details of each proposed model architecture. ISIC 2020, ISIC 2019, and ISIC 2016 datasets have been used to train the proposed classifiers using the concept of transfer learning [53] since these models are pre-trained on ImageNet [54]. The dataset has been split into training, validation, and test datasets with ratios of 0.8, 0.1, and 0.1 respectively. Two trials have been conducted for each proposed model to solve the unbalanced dataset problem. First, the non-melanoma class images are under-sampled to be the same size as the malignant class. The second approach is to perform over-sampling using the artificial images generated by the proposed GAN. Finally, each trial is subjected to post-training dynamic range quantization which converts weights to 8-bit precision to compress model size and decrease the inference time to fit a real-time system on edge devices [55]. Tables 8, 5, 6 and 7 show the results of the proposed classifiers trials with respect to the size of the model file in megabytes (MB), inference time in milliseconds (ms), AUC, the precision of non-melanoma (Norm) and malignant (Mal), recall of Norm and Mal, F1-score of Norm and Mal, and accuracy of train, validation and test data. EfficientNetV2S and EfficientNetV2B2 represent the models that are subjected to under-sampling, the same models with postfix GAN are the same architecture but the proposed GAN has been used to perform over-sampling to overcome the unbalanced dataset issue. Post-training dynamic range quantization has been expressed using postfix QUANT to reduce model size. This reduction is accompanied by a compression ratio, making the models more suitable for real-time edge device deployment. The models with postfix Attention are the same architecture but an attention layer has been used instead of a flattened layer.

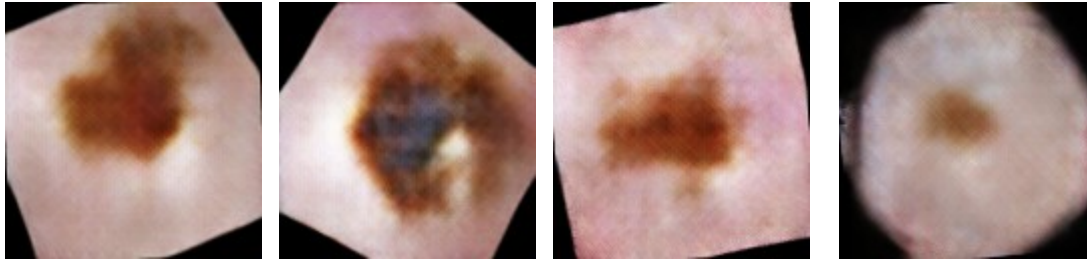


Fig. 5: Generated malignant images by the proposed GAN

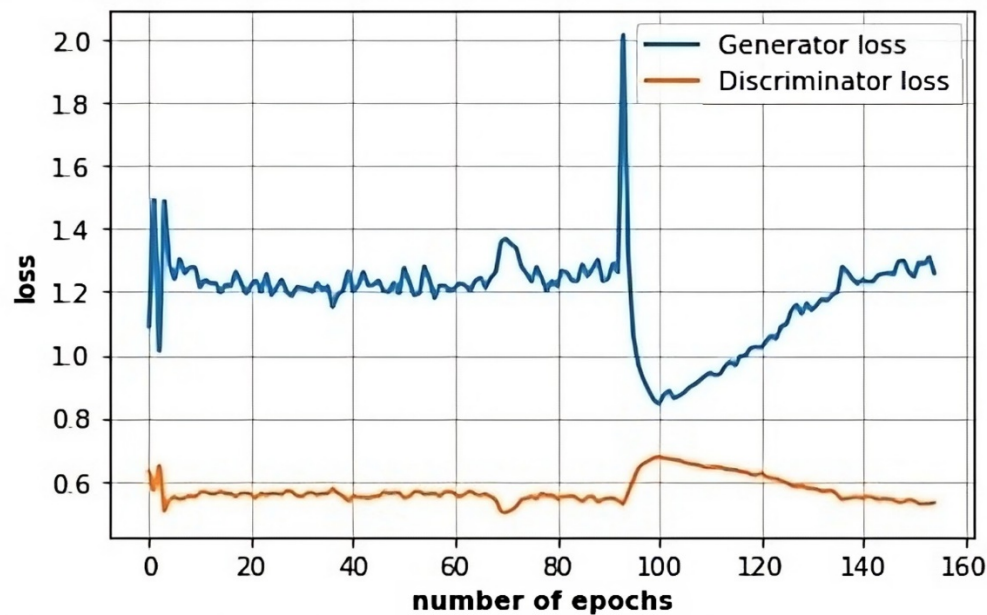


Fig. 6: GAN generator loss and discriminator loss across epochs

Table 4: Details of each model architecture

Name	Parameters	Layers count	Attention	Global average pooling	Dense count
EfficientNetV2S	20.333 M	513	0	1	1
EfficientNetV2S Attention	20.334 M	513	1	0	2
EfficientNetV2B2	8.772 M	349	0	1	1
EfficientNetV2S	8.772 M	513	1	0	2
ResNet101V2	42.630 M	377	0	1	1
Resnet152v2	58.335 M	564	0	1	1
DenseNet169	12.646 M	595	0	1	1
DenseNet201	18.325 M	707	0	1	1
U-Net	5.904 M	143	0	0	0
GAN generator	3.553 M	20	0	0	0
GAN discriminator	1.291 M	11	0	0	1

Table 5: Results of the proposed models with respect to trained data

Model	AUC	P(Norm)	R(Norm)	F(Norm)	P(Mal)	R(Mal)	F(Mal)	AC
EfficientNetV2S	0.939	0.978	0.958	0.968	0.854	0.919	0.885	0.95
EfficientNetV2S Attention	0.939	0.972	0.986	0.979	0.944	0.892	0.917	0.966
EfficientNetV2S QUANT	0.931	0.97	0.958	0.966	0.85	0.903	0.875	0.947
EfficientNetV2S GAN	0.939	0.98	0.953	0.966	0.839	0.926	0.88	0.947
EfficientNetV2S GAN QUANT	0.947	0.981	0.964	0.972	0.873	0.93	0.9	0.957
EfficientNetV2B2	0.94	0.974	0.977	0.976	0.915	0.904	0.909	0.962
EfficientNetV2B2 Attention	0.944	0.985	0.942	0.963	0.812	0.946	0.874	0.943
EfficientNetV2B2 QUANT	0.936	0.971	0.984	0.977	0.936	0.888	0.912	0.964
EfficientNetV2B2 GAN	0.923	0.962	0.992	0.977	0.966	0.854	0.906	0.963
EfficientNetV2B2 GAN QUANT	0.921	0.96	0.99	0.975	0.962	0.851	0.903	0.96
ResNet101v2	0.927	0.982	0.916	0.9484	0.7481	0.9381	0.8324	0.921
ResNet101v2 GAN	0.92	0.977	0.921	0.948	0.76	0.92	0.83	0.92
ResNet152v2	0.914	0.97	0.936	0.953	0.788	0.8932	0.8375	0.9275
ResNet152v2 GAN	0.65	0.856	0.819	0.837	0.42	0.486	0.45	0.749
DenseNet201	0.907	0.96	0.9667	0.963	0.87	0.847	0.859	0.942
DenseNet201 GAN	0.871	0.958	0.887	0.922	0.667	0.855	0.749	0.88
DenseNet169	0.81	0.907	0.9994	0.948	0.963	0.627	0.759	0.915
DenseNet169 GAN	0.8	0.904	0.993	0.95	0.96	0.607	0.744	0.916

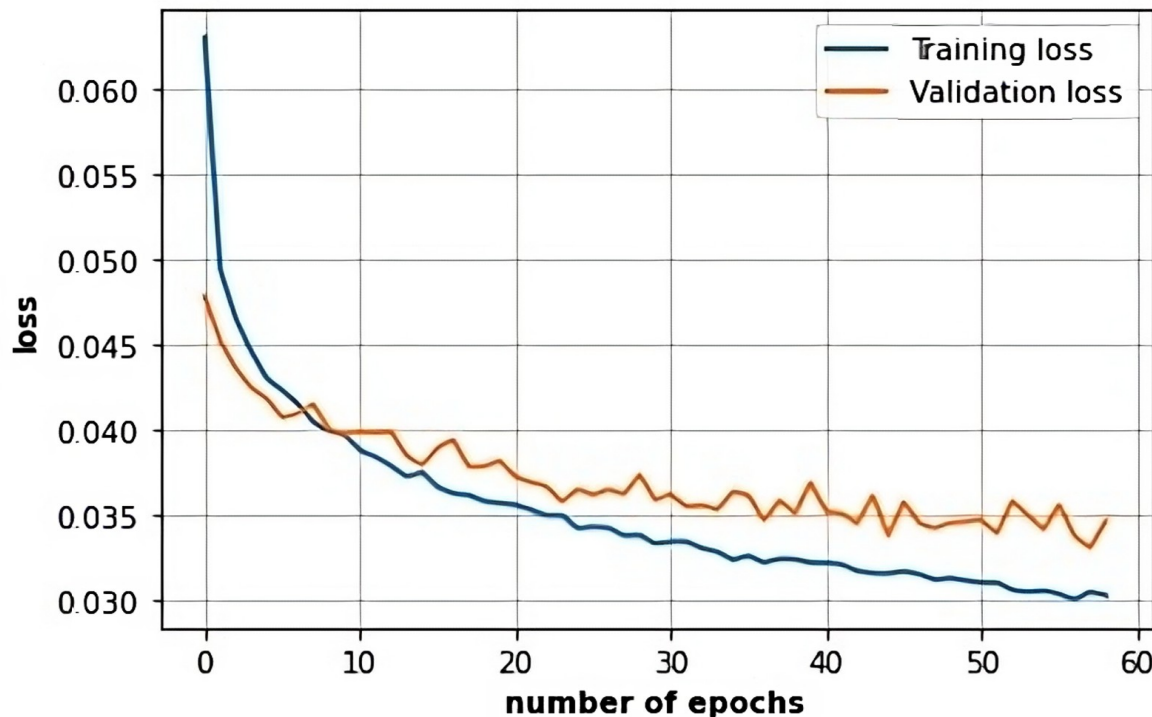


Fig. 7: Comparison between U-Net training loss and validation loss

5 Discussion

The results of EfficientNet outperform all previous CNN architectures on most benchmarking datasets. EfficientNetV2S Attention has outperformed other models on malignant F1-score and accuracy achieving 0.9168 and 0.965 respectively. Using GAN on preprocessing has improved the recall of the normal class in addition to the precision of the malignant but it has a negative impact on the precision of the normal class and recall of the malignant class. Quantization has reduced the EfficientNetV2S model size from 76.9 to 21 MB with a compression ratio of 3.66 and the EfficientNetV2B2 model size from 33.1 to 9.3 MB with a compression ratio of 3.55. Although the positive impact of quantization on model size and inference time, it has a relatively small negative effect on the other performance metrics. The proposed framework has outperformed state-of-the-art studies on the recall of normal and malignant classes and accuracy. EfficientNetV2S Attention has achieved an improvement ratio of 2.7 over the best-achieved accuracy by the state-of-the-art studies. The EfficientNetV2S model Attention might have been subjected to extensive hyperparameter tuning in order to maximize its learning rates, regularization approaches, and other crucial configurations. Optimizing these parameters has the potential to enhance the performance of the model. In addition, the attention processes enable the model to selectively concentrate on particular regions of interest within the input data, specifically in the context of skin lesion images. This technique enhances the model's ability to effectively collect important features relevant to the classification task. Fig. 9 shows the Receiver operating characteristic curves (ROC) of the best five models on AUC, EfficientNetV2B2 has achieved the highest AUC with a value of 0.96. Table 9 shows the proposed framework results compared to state-of-the-art models.

6 Conclusions

In this study, a framework for the classification of skin lesion images is proposed, combining several techniques, beginning with image preprocessing, including the implementation of a median filter, followed by the execution of a hair removal algorithm, segmentation of ROI, and ending with the generation of melanoma lesions images in the training set to solve the unbalanced dataset problem. Finally, transfer learning was applied to six different pre-trained CNNs from the EfficientNet, ResNet, and DenseNet families. These techniques enabled CNN to outperform in the evaluation metrics. In contrast to prior research, the findings showed an enhancement in the accuracy of classification. The selection of EfficientNetV2S attention model as a classifier is based on its superior performance in terms of malignancy F1-score and accuracy, surpassing other models. AMDS has not only achieved low latency but also has outperformed state-of-the-art studies in recall, accuracy, and F1-score. This finding is of tremendous significance since it can aid in the early detection of skin cancer, which is critical for increasing patient survival rates and providing a

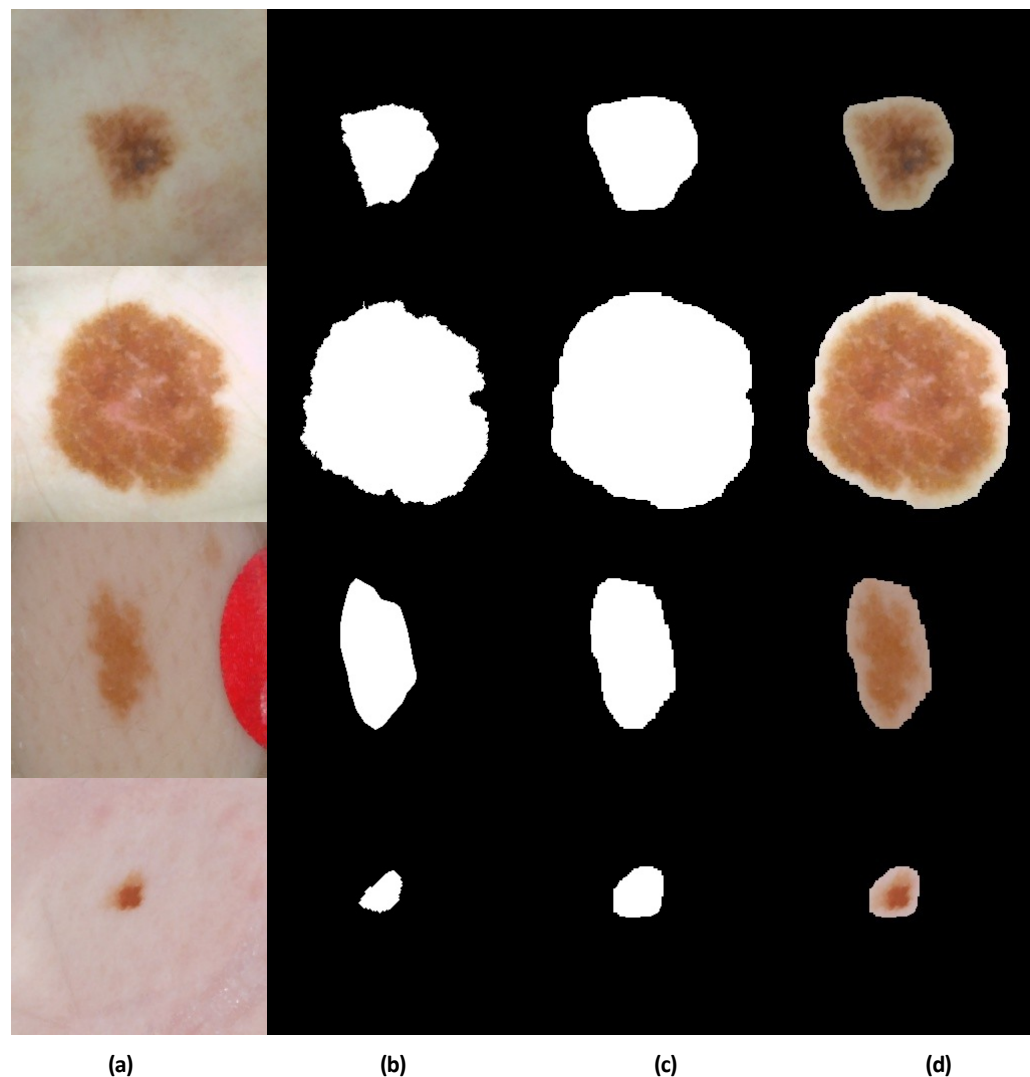


Fig. 8: Group (a) represents the original image, Group (b) represents the ground truth, Group (c) represents the predicted mask, Group (d) represents the ROI output of U-Net

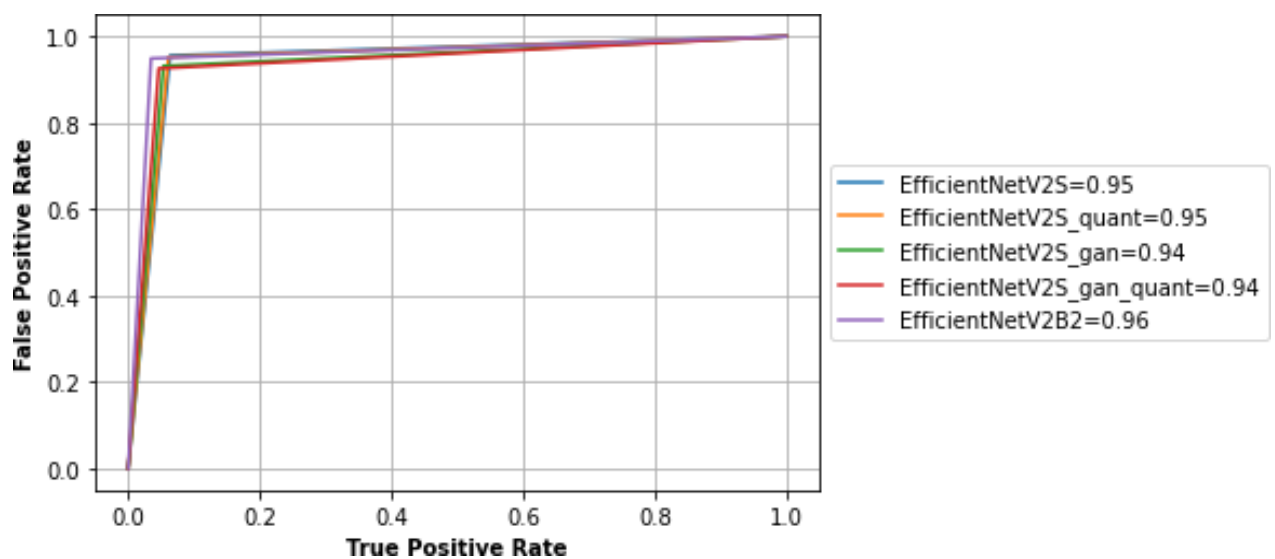


Fig. 9: ROC curves of the best five models

Table 6: Results of the proposed models with respect to validation data

Model	AUC	P(Norm)	R(Norm)	F(Norm)	P(Mal)	R(Mal)	F(Mal)	AC
EfficientNetV2S	0.936	0.9761	0.9614	0.9687	0.862	0.911	0.885	0.95
EfficientNetV2S Attention	0.942	0.973	0.987	0.98	0.948	0.897	0.922	0.968
EfficientNetV2S QUANT	0.938	0.97	0.963	0.97	0.86	0.91	0.89	0.95
EfficientNetV2S GAN	0.89	0.947	0.99	0.968	0.956	0.79	0.86	0.948
EfficientNetV2S GAN QUANT	0.93	0.977	0.964	0.97	0.866	0.91	0.89	0.953
EfficientNetV2B2	0.93	0.973	0.9788	0.976	0.918	0.898	0.9	0.961
EfficientNetV2B2 Attention	0.946	0.985	0.947	0.965	0.82	0.945	0.88	0.946
EfficientNetV2B2 QUANT	0.934	0.971	0.979	0.975	0.92	0.89	0.9	0.96
EfficientNetV2B2 GAN	0.91	0.956	0.989	0.973	0.955	0.831	0.88	0.956
EfficientNetV2B2 GAN QUANT	0.906	0.954	0.99	0.972	0.956	0.82	0.884	0.95
ResNet101v2	0.922	0.977	0.923	0.95	0.76	0.92	0.83	0.923
ResNet101v2 GAN	0.919	0.974	0.929	0.95	0.77	0.9	0.83	0.92
ResNet152v2	0.916	0.97	0.94	0.955	0.798	0.891	0.84	0.93
ResNet152v2 GAN	0.658	0.861	0.81	0.85	0.41	0.5	0.45	0.746
DenseNet201	0.907	0.95	0.97	0.96	0.88	0.841	0.86	0.943
DenseNet201 GAN	0.88	0.963	0.89	0.926	0.68	0.87	0.76	0.88
DenseNet169	0.78	0.89	0.99	0.943	0.957	0.579	0.721	0.906
DenseNet169 GAN	0.786	0.89	0.993	0.943	0.957	0.57	0.72	0.906

Table 7: Results of the proposed models with respect to test data

Model	AUC	P(Norm)	R(Norm)	F(Norm)	P(Mal)	R(Mal)	F(Mal)	AC
EfficientNetV2S	0.946	0.99	0.94	0.96	0.8	0.96	0.87	0.94
EfficientNetV2S Attention	0.942	0.974	0.982	0.978	0.9317	0.9024	0.9168	0.9657
EfficientNetV2S QUANT	0.946	0.99	0.94	0.96	0.8	0.95	0.87	0.94
EfficientNetV2S GAN	0.94	0.98	0.95	0.96	0.82	0.93	0.87	0.94
EfficientNetV	0.94	0.98	0.95	0.97	0.84	0.93	0.88	0.95

2S GAN QUANT								
EfficientNetV 2B2	0.957	0.99	0.97	0.98	0.88	0.95	0.91	0.96
EfficientNetV 2B2 Attention	0.94	0.982	0.947	0.964	0.824	0.936	0.876	0.944
EfficientNetV 2B2 QUANT	0.921	0.97	0.98	0.97	0.9	0.87	0.88	0.95
EfficientNetV 2B2 GAN	0.921	0.96	0.99	0.98	0.95	0.85	0.9	0.96
EfficientNetV 2B2 GAN QUANT	0.92	0.96	0.99	0.97	0.95	0.85	0.9	0.96
ResNet101v2	0.9	0.9677	0.9210	0.9438	0.7475	0.8839	0.8100	0.9133
ResNet101v2 GAN	0.689	0.8597	0.9905	0.9205	0.9153	0.3891	0.5460	0.8647
ResNet152v2	0.87	0.9419	0.9754	0.95838	0.8926	0.7724	0.82810	0.9329
ResNet152v2 GAN	0.8	0.9236	0.8759	0.8991	0.6074	0.7259	0.6614	0.8445
DenseNet201	0.9	0.9694	0.9244	0.9464	0.7569	0.8897	0.8179	0.9172
DenseNet201 GAN	0.8	0.9840	0.6442	0.7787	0.4166	0.9605	0.5812	0.7104
DenseNet169	0.82	0.9184	0.9791	0.9478	0.8947	0.6713	0.7671	0.9147
DenseNet169 GAN	0.833	0.9592	0.7951	0.8695	0.5296	0.8722	0.6591	0.8112

Table 8: Results of the proposed models

Model	Size(MB)	Inference(ms)
EfficientNetV2S	76.9	106.11
EfficientNetV2S Attention	233	348.6
EfficientNetV2S QUANT	21	105.93
EfficientNetV2S GAN	76.9	105.28
EfficientNetV2S GAN QUANT	21	105.77
EfficientNetV2B2	33.1	57.38
EfficientNetV2B2 Attention	100	61.3
EfficientNetV2B2 QUANT	9.3	52.55
EfficientNetV2B2 GAN	33.1	56.57
EfficientNetV2B2 GAN QUANT	9.3	52.84
ResNet101v2	488	79.3
ResNet101v2 GAN	488	76.7
ResNet152v2	688	94
ResNet152v2 GAN	688	125.2
DenseNet201	210	106.7

DenseNet201 GAN	210	123.7
DenseNet169	145	93.45
DenseNet169 GAN	145	91.8

Table 9: Results of the proposed model compared to the state of the art studies

name	AUC	R (Norm)	P (Mal)	R (Mal)	F (Mal)	AC
[24]	-	0.878	-	0.885	-	0.893
[25]	-	-	0.965	0.936	0.95	0.914
[26]	0.87	0.876	0.874	0.861	-	0.869
[29]	-	-	-	-	0.874	0.873
[30]	-	0.925	-	0.82	-	0.904
[32]	-	-	-	-	-	0.902
[36]	-	0.913	-	0.936	-	0.923
[37]	0.95	-	0.85	0.88	-	0.94
[39]	-	-	-	-	-	0.884
[40]	-	0.74	-	0.68	0.7	0.71
[41]	-	-	0.785	0.928	-	0.837

Acknowledgement

The authors are grateful to the anonymous referee for careful checking of the details and for helpful comments that improved this paper.

Conflicts of Interest Statement

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

- [1] Jacques Ferlay et al. "Cancer statistics for the year 2020: An overview". In: International Journal of Cancer 149.4 (Apr. 2021), pp. 778–789. DOI: 10.1002/ijc.33588. URL: <https://doi.org/10.1002/ijc.33588>.
- [2] Miguel A. Linares, Alan Zakaria, and Parminder Nizran. "Skin Cancer". In: Primary Care: Clinics in Office Practice 42.4 (2015). Primary Care Dermatology, pp. 645–659. ISSN: 0095-4543. DOI: <https://doi.org/10.1016/j.pop.2015.07.006>. URL: <https://www.sciencedirect.com/science/article/pii/S0095454315000512>.
- [3] M.E. Vestergaard et al. "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting". In: British Journal of Dermatology (June 2008). DOI: 10.1111/j.1365-2133.2008.08713.x. URL: <https://doi.org/10.1111%5C%2Fj.1365-2133.2008.08713.x>.
- [4] Dr Tribhuwan Kumar et al. "An evaluation on speech recognition technology based on machine learning". In: Webology 19.1 (Jan. 2022), pp. 646–663.
- [5] Shanshan Hua et al. "Research on 3D medical image surface reconstruction based on data mining and machine learning". en. In: Int. J. Intell. Syst. 37.8 (Aug. 2022), pp. 4654–4669.
- [6] Alexis W Mills et al. "Exploring potential energy surfaces using reinforcement machine learning". en. In: J. Chem. Inf. Model. 62.13 (July 2022), pp. 3169–3179.
- [7] Jacob P Kimball et al. "Wearable sensors and machine learning for hypovolemia problems in occupational, military and sports medicine: Physiological basis, hardware and algorithms". en. In: Sensors (Basel) 22.2 (Jan. 2022), p. 442.
- [8] Ana M Maitin, Juan Pablo Romero Muñoz, and A' lvaro Jose' Garc' ia-Tejedor. "Survey of Machine Learning

- techniques in the analysis of EEG signals for Parkinson's disease: A systematic review". en. In: Appl. Sci. (Basel) 12.14 (July 2022), p. 6967.
- [9] Pallavi Tiwari et al. "CNN Based Multiclass Brain Tumor Detection Using Medical Imaging". In: Computational Intelligence and Neuroscience 2022 (June 2022), p. 1830010. ISSN: 1687-5265. DOI: 10.1155/2022/1830010. URL: <https://doi.org/10.1155/2022/1830010>.
- [10] S Deepak and P M Ameer. "Brain tumor classification using deep CNN features via transfer learning". en. In: Comput. Biol. Med. 111.103345 (Aug. 2019), p. 103345.
- [11] Meha Desai and Manan Shah. "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)". en. In: Clin. eHealth 4 (2021), pp. 1–11.
- [12] Nada M. Hassan, Safwat Hamad, and Khaled Mahar. "A Deep Learning Model for Mammography Mass Detection Using Mosaic and Reconstructed Multichannel Images". In: (2022), pp. 544–559. DOI: 10.1007/978-3-031-10522-7_37. URL: https://doi.org/10.1007/978-3-031-10522-7_37.
- [13] Tanzila Saba et al. "Region Extraction and Classification of Skin Cancer: A Heterogeneous framework of Deep CNN Features Fusion and Reduction". In: Journal of Medical Systems 43.9 (July 2019). DOI: 10.1007/s10916-019-1413-3. URL: <https://doi.org/10.1007/s10916-019-1413-3>.
- [14] Cahyo Adhi Hartanto and Adi Wibowo. "Development of Mobile Skin Cancer Detection using Faster R-CNN and MobileNet v2 Model". In: 2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE). 2020, pp. 58–63. DOI: 10.1109/ICITACEE50144.2020.9239197.
- [15] Jwan Saeed and Subhi Zeebaree. "Skin Lesion Classification Based on Deep Convolutional Neural Networks Architectures". In: Journal of Applied Science and Technology Trends 2.01 (Mar. 2021), pp. 41–51. DOI: 10.38094/jastt20189. URL: <https://doi.org/10.38094/jastt20189>.
- [16] Roweida Mohammed, Jumanah Rawashdeh, and Malak Abdullah. "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results". In: 2020 11th International Conference on Information and Communication Systems (ICICS). 2020, pp. 243–248. DOI: 10.1109/ICICS49469.2020.239556.
- [17] Zahra Mungloo-Dilmohamud et al. "Balancing Data through Data Augmentation Improves the Generality of Transfer Learning for Diabetic Retinopathy Classification". In: Applied Sciences 12.11 (May 2022), p. 5363. DOI: 10.3390/app12115363. URL: <https://doi.org/10.3390/app12115363>.
- [18] Yin Cui et al. "Class-Balanced Loss Based on Effective Number of Samples". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 2019.
- [19] Ian Goodfellow et al. "Generative Adversarial Networks". In: Commun. ACM 63.11 (Oct. 2020), pp. 139–144. ISSN: 0001-0782. DOI: 10.1145/3422622. URL: <https://doi.org/10.1145/3422622>.
- [20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. DOI: 10.48550/ARXIV.1505.04597. URL: <https://arxiv.org/abs/1505.04597>.
- [21] Mingxing Tan and Quoc V. Le. "EfficientNetV2: Smaller Models and Faster Training". In: (2021). DOI: 10.48550/ARXIV.2104.00298. URL: <https://arxiv.org/abs/2104.00298>.
- [22] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [23] G. Huang et al. "Densely Connected Convolutional Networks". In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society, July 2017, pp. 2261–2269. DOI: 10.1109/CVPR.2017.243. URL: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.243>.
- [24] A.Murugan et al. "Diagnosis of skin cancer using machine learning techniques". In: Microprocessors and Microsystems 81 (Mar. 2021), p. 103727. DOI: 10.1016/j.micpro.2020.103727. URL: <https://doi.org/10.1016/j.micpro.2020.103727>.
- [25] Md Shahin Ali et al. "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models". In: Machine Learning with Applications 5 (Sept. 2021), p. 100036. DOI: 10.1016/j.mlwa.2021.100036. URL: <https://doi.org/10.1016/j.mlwa.2021.100036>.
- [26] Maad M. Mijwil. "Skin cancer disease images classification using deep learning solutions". In: Multimedia Tools

- and Applications (Apr. 2021). DOI: 10.1007/s11042-021-10952-7. URL: <https://doi.org/10.1007/s11042-021-10952-7>.
- [27] Marc Combalia et al. BCN20000: Dermoscopic Lesions in the Wild. 2019. DOI: 10.48550/ARXIV.1908.02288. URL: <https://arxiv.org/abs/1908.02288>.
- [28] Veronica Rotemberg et al. “A patient-centric dataset of images and metadata for identifying melanomas using clinical context”. In: Scientific Data 8.1 (Jan. 2021). DOI: 10.1038/s41597-021-00815-z. URL: <https://doi.org/10.1038/s41597-021-00815-z>.
- [29] Ahmet Demir, Feyza Yilmaz, and Onur Kose. “Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3”. In: (Oct. 2019). DOI: 10.1109/tiptekno47231.2019.8972045. URL: <https://doi.org/10.1109/tiptekno47231.2019.8972045>.
- [30] Mario Fernando Jojoa Acosta et al. “Melanoma diagnosis using deep learning techniques on dermatoscopic images”. In: BMC Medical Imaging 21.1 (Jan. 2021). DOI: 10.1186/s12880-020-00534-8. URL: <https://doi.org/10.1186/s12880-020-00534-8>.
- [31] Noel C. F. Codella et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). 2017. DOI: 10.48550/ARXIV.1710.05006. URL: <https://arxiv.org/abs/1710.05006>.
- [32] Amer Sallam, Abdulfattah E. Ba Alawi, and Ahmed Y. A. Saeed. “A CNN-Based Model for Early Melanoma Detection”. In: Innovative Systems for Intelligent Health Informatics. Ed. by Faisal Saeed, Fathey Mohammed, and Abdulaziz Al-Nahari. Cham: Springer International Publishing, 2021, pp. 41–51. ISBN: 978-3-030-70713-2.
- [33] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions”. In: Scientific Data 5.1 (Aug. 2018). DOI: 10.1038/sdata.2018.161. URL: <https://doi.org/10.1038/sdata.2018.161>.
- [34] Noel Codella et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). 2019. DOI: 10.48550/ARXIV.1902.03368. URL: <https://arxiv.org/abs/1902.03368>.
- [35] Ch Hima Bindu and K Satya Prasad. “An efficient medical image segmentation using conventional OTSU method”. In: International Journal of Advanced Science and Technology 38.1 (2012), pp. 67–74.
- [36] Hatice Catal Reis et al. “InSiNet: a deep convolutional approach to skin cancer detection and segmentation”. In: Medical & Biological Engineering & Computing 60.3 (Jan. 2022), pp. 643–662. DOI: 10.1007/s11517-021-02473-0. URL: <https://doi.org/10.1007/s11517-021-02473-0>.
- [37] Ruban Nersisson et al. “A Dermoscopic Skin Lesion Classification Technique Using YOLO-CNN and Traditional Feature Model”. In: Arabian Journal for Science and Engineering 46.10 (Apr. 2021), pp. 9797–9808. DOI: 10.1007/s13369-021-05571-1. URL: <https://doi.org/10.1007/s13369-021-05571-1>.
- [38] David Gutman et al. Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). 2016. DOI: 10.48550/ARXIV.1605.01397. URL: <https://arxiv.org/abs/1605.01397>.
- [39] Jinen Daghrir et al. “Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach”. In: 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). 2020, pp. 1–5. DOI: 10.1109/ATSIP49331.2020.9231544.
- [40] Pooyan Sedigh, Rasoul Sadeghian, and Mehdi Tale Masouleh. “Generating Synthetic Medical Images by Using GAN to Improve CNN Performance in Skin Cancer Classification”. In: 2019 7th International Conference on Robotics and Mechatronics (ICRoM). 2019, pp. 497–502. DOI: 10.1109/ICRoM48714.2019.9071823.
- [41] James Ren Hou Lee et al. “Cancer-Net SCA: tailored deep neural network designs for detection of skin cancer from dermoscopy images”. In: BMC Medical Imaging 22.1 (Aug. 2022). DOI: 10.1186/s12880-022-00871-w. URL: <https://doi.org/10.1186/s12880-022-00871-w>.
- [42] Ivan Culjak et al. “A brief introduction to OpenCV”. In: 2012 Proceedings of the 35th International Convention MIPRO. 2012, pp. 1725–1730.
- [43] Joost Koehoorn et al. “Automated Digital Hair Removal by Threshold Decomposition and Morphological Analysis”. In: (2015), pp. 15–26. DOI: 10.1007/978-3-319-18720-4_2. URL: https://doi.org/10.1007/978-3-319-18720-4_2.

- [44] Agnieszka Mikołajczyk and Michał Grochowski. “Data augmentation for improving deep learning in image classification problem”. In: 2018 International Interdisciplinary PhD Workshop (IIPHDW). 2018, pp. 117–122. DOI: 10.1109/IIPHDW.2018.8388338.
- [45] Andrew L. Maas. “Rectifier Nonlinearities Improve Neural Network Acoustic Models”. In: 2013.
- [46] Dzung L Pham, Chenyang Xu, and Jerry L Prince. “A survey of current methods in medical image segmentation”. In: Annual review of biomedical engineering 2.3 (2000), pp. 315–337.
- [47] Tianyi Liu et al. “Implementation of Training Convolutional Neural Networks”. In: CoRR abs/1506.01195 (2015). arXiv: 1506.01195. URL: <http://arxiv.org/abs/1506.01195>.
- [48] Xiao-Xia Yin et al. “U-Net-Based Medical Image Segmentation”. In: Journal of Healthcare Engineering 2022 (Apr.2022). Ed. by Hangjun Che, pp. 1–16. DOI: 10.1155/2022/4189781. URL: <https://doi.org/10.1155/2022/4189781>.
- [49] Pius Kwao Gadosey et al. “SD-UNet: Stripping down U-Net for Segmentation of Biomedical Images on Platforms with Low Computational Budgets”. In: Diagnostics 10.2 (Feb. 2020), p. 110. DOI: 10.3390/diagnostics10020110. URL: <https://doi.org/10.3390/diagnostics10020110>.
- [50] Ashish Vaswani et al. Attention Is All You Need. 2017. DOI: 10.48550/ARXIV.1706.03762. URL: <https://arxiv.org/abs/1706.03762>.
- [51] Xin He, Kaiyong Zhao, and Xiaowen Chu. “AutoML: A Survey of the State-of-the-Art”. In: (2019). DOI: 10.48550/ARXIV.1908.00709. URL: <https://arxiv.org/abs/1908.00709>.
- [52] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. 2014. DOI: 10.48550/ARXIV.1412.6980. URL: <https://arxiv.org/abs/1412.6980>.
- [53] Lisa Torrey and Jude Shavlik. “Transfer Learning”. In: (2010), pp. 242–264. DOI: 10.4018/978-1-60566-766-9.ch011. URL: <https://doi.org/10.4018/978-1-60566-766-9.ch011>.
- [54] Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee. 2009, pp. 248–255.
- [55] Jinjie Zhang, Yixuan Zhou, and Rayan Saab. Post-training Quantization for Neural Networks with Provable Guarantees. 2022. DOI: 10.48550/ARXIV.2201.11113. URL: <https://arxiv.org/abs/2201.11113>.