Hindawi
Contrast Media & Molecular Imaging
Volume 2023, Article ID 5869513, 15 pages
https://doi.org/10.1155/2023/5869513

Research Article

An Ensemble of Transfer Learning Models for the Prediction of Skin Lesions with Conditional Generative Adversarial Networks

Amal Al-Rasheed,[1] **Amel Ksibi**,[1] **Manel Ayadi**,[1] **Abdullah I. A. Alzahrani**,[2] **and Mohammad Mamun Elahi**[3]

[1]*Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia*
[2]*Department of Computer Science, Shaqra University, AlQuwaiiyah, Saudi Arabia*
[3]*Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh*

Correspondence should be addressed to Mohammad Mamun Elahi; mmelahi@cse.uiu.ac.bd

Received 15 June 2022; Revised 13 July 2022; Accepted 22 July 2022; Published 8 April 2023

Academic Editor: Mohammad Farukh Hashmi

Skin cancer is one of the most serious forms of the disease, and it can spread to other parts of the body if not detected early. Therefore, it is crucial to diagnose and treat skin cancer patients at an early stage. Due to the fact that a manual diagnosis of skin cancer is both time-consuming and expensive, an incorrect diagnosis is made due to the high degree of similarity between the various skin lesions. Improved categorization of multi-class skin lesions requires the development of automated diagnostic systems. We offer a fully automated method for classifying several skin lesions by fine-tuning the deep learning models, namely VGG16, ResNet50, and ResNet101. Prior to model creation, the training dataset should undergo data augmentation using traditional image transformation techniques and generative adversarial networks (GANs) to prevent class imbalance issues that may lead to model overfitting. In this study, we investigate the feasibility of creating dermoscopic images that have a realistic appearance using conditional generative adversarial network (CGAN) techniques. Afterward, the traditional augmentation methods are used to augment our existing training set to improve the performance of pretrained deep models on the skin lesion classification task. This improved performance is then compared to the models developed using the unbalanced dataset. In addition, we formed an ensemble of finely tuned transfer learning models, which we trained on balanced and unbalanced datasets. These models were used to make predictions about the data. With appropriate data augmentation, the proposed models attained an accuracy of 92% for VGG16, 92% for ResNet50, and 92.25% for ResNet101. The ensemble of these models increased the accuracy to 93.5%. There was a comprehensive discussion on the performance of the models. It is possible to conclude that using such a method leads to enhanced performance in skin lesion categorization compared to the efforts made in the past.

1. Introduction

Skin cancer is a condition that develops when the DNA of healthy skin cells undergoes mutations that allow them to divide abnormally and turn malignant [1, 2]. Excessive ultraviolet (UV) radiation exposure, time spent in the sun, and usage of a solarium are all possible causes [3]. Derived from the perspective of histology, skin cancer has an uneven cell structure with varying degrees of chromatin, nucleus, and cytoplasm [4]. Worldwide, skin cancer is one of the leading causes of death [5]. Basal cell carcinoma (BCC), melanoma (MEL) and nonmelanoma skin cancer, and squamous cell carcinoma are the most common types of skin cancer (SCC). Infrequent skin cancers, such as Kaposi sarcoma (KS) and actinic keratosis (AK), include solar keratosis, lymphoma, and keratoacanthoma. Certain kinds of skin cancer are fatal and metastasis by nature. However, not all lesions are caused by malignant tumors. As the cancer of the skin begins in the *epidermis*, the outermost layer of skin, where it is visible to the human eye (National Cancer Institute, 2019), identifying a lesion as malignant (cancerous) or benign (non-cancerous) is frequently made based on a visual examination followed by a biopsy [6].

Melanoma is the most severe type of cancer since it is incurable. The majority of case turnover is death. Occasionally, melanoma develops from a lesion when its size, irritation, and hue alter. Typically, nonmelanomas are more prevalent than melanomas, yet melanoma is the leading cause of death from skin cancer. However, early skin cancer detection and diagnosis will enhance the likelihood of recovery and survival; failing to do so will result in dire circumstances [7, 8].

The pervasive and lethal character of the disease necessitates the development of an accurate, noninvasive diagnostic method. Most skin cancers are diagnosed using visual, clinical, and histological examinations. Frequently, medical diagnosis depends on the patient's past, ethnicity, social behaviours, and sun exposure [9]. Visual inspection with the naked eye is typically incapable of identifying and revealing the intricacies. As a solution, dermatoscopy, an imaging tool for skin lesion investigation, was developed. The optical dermatoscopy records dermatoscopic images using a high-resolution and magnifying camera lens. This recording technique eliminates the skin's surface reflection, allowing a real-time examination of the *epidermis* and dermis structures so that more visual data may be gathered from the deeper layers of skin, which will further aid in creating more precise computer-aided diagnostic (CAD) systems. With only a visual examination, a dermatologist's accuracy rate ranged between 65 and 80% [10]. However, dermatoscopy significantly improved the accuracy of early disease diagnosis [11]. A dermatologist's eye examination and dermoscopic images have a combined accuracy rate of 75% to 84% [12, 13].

Even though dermoscopic images have improved accuracy, it still relies on the clinician's expertise and subjective opinion to a large extent [14]. Color, dermal, contour, geometric, and texture features of lesions classify skin lesions. Skin lesions are difficult to classify visually. The degree of resemblance among the visual features of different lesion classes may lead to the incorrect recognition of lesions, especially when the cancer is in its early stages [15]. As a result, dermatologists frequently misclassify malignant and benign melanomas, which can devastate patients. It is more dangerous than the squamous and basal because melanoma spreads throughout the body much more quickly and attacks organs, including the brain and liver [3]. Dermatologists must develop new diagnostic techniques and methods to assist them in making early and accurate diagnoses of skin cancer to prevent or cure the disease due to the rapid development of skin cancer, the risk of metastasis, and the lack of therapeutic access [16]. New diagnostic instruments and methodologies are required for dermatologists and other medical professionals to accurately diagnose skin cancer.

Given the difficulty of diagnosing and treating skin cancer with the human eye, computer vision can be utilized for this purpose. To reduce the complexity of traditional machine learning techniques, a subject matter expert must first specify the features that will be employed. However, deep learning (DL) methods, a subfield of machine learning, can be trained on many benign and cancerous images. The DL model can determine if a picture is malignant or benign by learning nonlinear correlations. As a result, no domain expertise is required for feature extraction in DL. Using convolutional neural networks (CNNs) for deep learning is the topic of this study.

The current work attempted to develop a novel diagnosis solution for skin cancer that had an affordable computational cost and high accuracy as the early detection of cancer is vital for both treatment and a cure for cancer. We develop an ensemble-based architecture that can be successfully employed to improve the accuracy of individual CNNs. The fusion of CNNs is one possible way to address the issues connected to the applicability of a single CNN for a given job. This is accomplished by allowing additional classifiers, each of which is based on a distinct CNN, in order to compensate for each other's shortcomings. To be more specific, we demonstrate how we can build a CNN ensemble in order to outperform the accuracy of individual neural networks that are trained on the available dataset. In addition to this, an investigation into the impact that data augmentation has on the overall performance of ensemble models was carried out. This study is the first of its kind in the field of early identification of skin cancer. The deep learning models that are presented can also be scalable to many devices, platforms, and operating systems, thereby transforming these into contemporary medical instruments.

The contributions of the work are as follows:

(1) Exploring image augmentation methods such as flip, affine, linear contrast, multiply, and Gaussian blur (image transformation methods) to balance the dataset

(2) Exploring the conditional GAN architecture for generating skin lesion images

(3) Performance analysis of the fine-tuned pretrained models, namely VGG16, ResNet50, and ResNet101 on both balanced and unbalanced datasets.

(4) An ensemble algorithm by combining the predictions of the three fine-tuned models to improve the performance obtained by deep individual models.

The rest of this article is organized as follows: Section 2 discusses previous research undertaken on the topic. Section 3 provides a full mathematical explanation and visual results for the proposed methodology. In Sections 4 and 5, the experimental design and findings are discussed. Finally, the conclusion is presented in Section 6.

2. Literature Review

Several studies have utilized databases of dermoscopic skin lesions to aid in diagnosing lesions. Early studies on skin cancer focused mostly on various algorithms for categorizing skin lesions using traditional AI approaches, which typically begin with a phase of manual feature extraction, followed by a distinct period of classifier training. Early attempts to distinguish between skin lesions that were either

MEL or nonmelanoma depended on low-level, manually-created characteristics [17]. Handcrafted features for dermoscopy images often have a low generalization power due to a lack of biological principles, understanding, and human intuition. Low-level handcrafted traits cannot distinguish complex skin lesions. In addition, there were considerable visual similarity challenges, high levels of intraclass disparity, and the appearance of artifacts in dermoscopic images that resulted in poor performance [18]. So deep learning and CNNs are unquestionably the preferred techniques in many computer vision applications [6, 19–21] trained on a dataset of over 100,000 clinical images annotated by experienced dermatologists using Inception-v3 architecture. Deep CNN was developed for two critical binary classification use cases: keratinocyte carcinomas versus benign seborrheic keratoses and malignant melanomas versus benign nevi. The first involves identifying the most common cancers, and the second consists in identifying the deadliest skin cancer. Differentiating benign nevi from malignant melanomas achieved 72.1 ± 0.9% accuracy, which was better than dermatologist discrimination rates.

AlexNet was used by [22] to classify three different lesions (melanoma, common nevus, and atypical nevus). The PH^2 dataset is used to train and test the proposed model [23]. The well-known quantitative measures of accuracy, sensitivity, specificity, and precision are used to evaluate the proposed method's performance, with 98.61%, 98.33%, 98.93%, and 97.73% obtained, respectively.

An ensemble strategy for CNNs has been suggested by [24], which incorporates both intra-architecture and inter-architecture network fusion in its design. Feature abstraction levels are represented by a variety of CNN architectures in the proposed method. For each network, the in-depth features were used to train different support vector machine (SVM) classifications. The proposed algorithm has an area under the receiver operating characteristic (ROC) curve for melanoma classification of 87.3% and an area under the ROC curve for seborrheic keratosis classification of 95.5% when tested on the 600 test images from the ISIC 2017 skin lesion classification challenge.

Patch-based attention architecture suggested by [25] in 2020 provides global context between high-resolution patches. Patch-based attention enhanced the mean sensitivity by 7% in the three pretrained architectures studied.

The two-phase strategy presented by [26] includes mid-level features. They first identified the region of interest using dermoscopic images and then used pretrained algorithms to extract information from the images. Their feature-based mid-level algorithm achieved a ROC of 0.87 for MEL and 0.97 for BKL. Aburaed et al. [27] explored how accurately skin cancer can be diagnosed because of the development of CNNs. This research demonstrates the skin cancer classification approach using the HAM 10000 dataset. Implementation, training, and evaluation of VGG16, VGG19, and a deep CNN are also proposed. Garg et al. [28] used dermoscopy images from the MNIST HAM-10,000 datasets in this study. Along with DL, image augmentation techniques also helped to boost the total number of images. They turned to the transfer learning approach for the last boost in image classification precision. CNN's weighted average precision of 0.88%, weighted recall average of 0.74%, and weighted F1-score of 0.77% were all achieved with our model. The ResNet model's transfer learning method produced an accuracy of 90.51%.

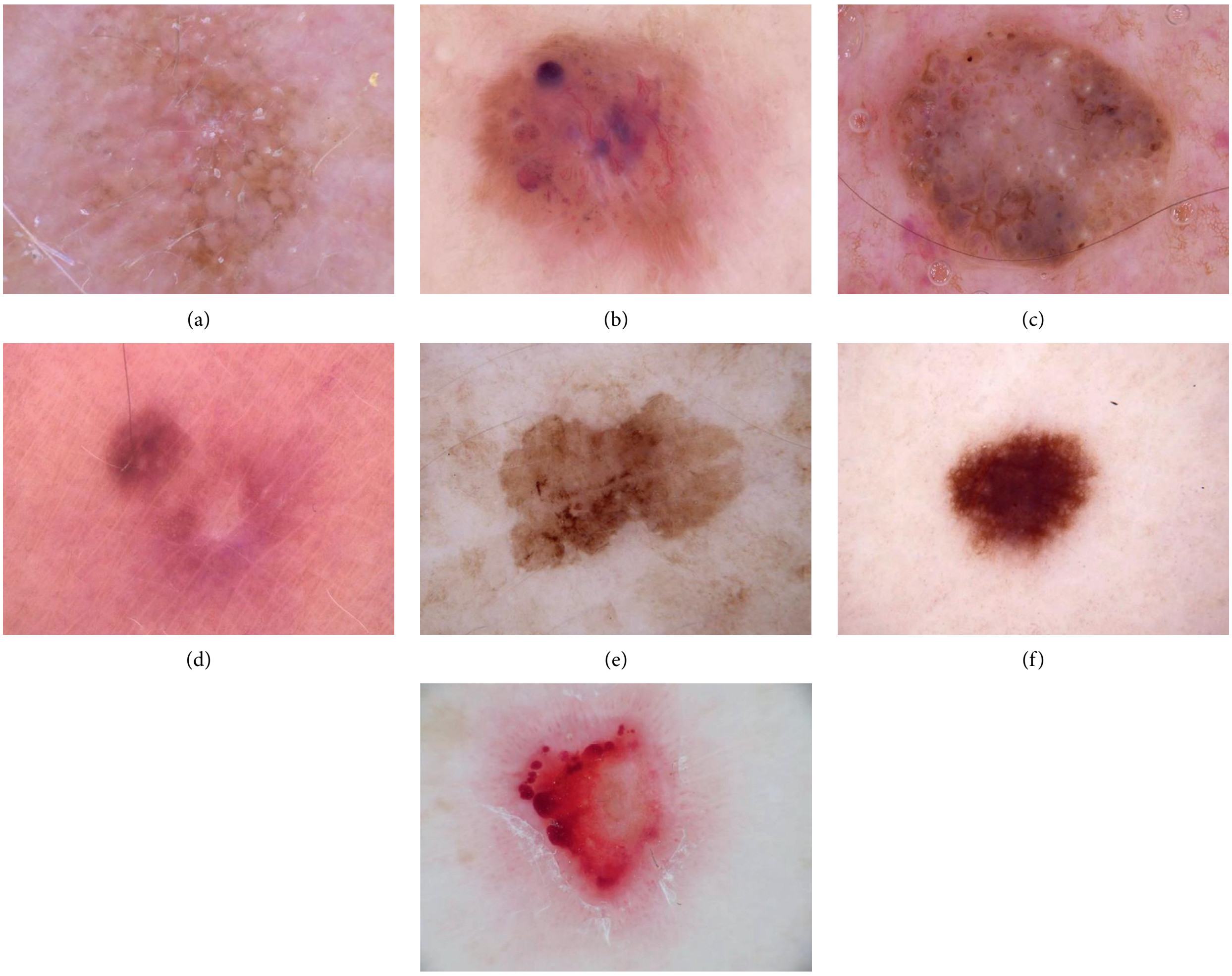
A pretrained DarkNet19 deep neural network model was utilized by [16] to generate image gradients by tweaking the parameters of the third convolutional layer. Next, high-frequency and multilayered feed-forward neural networks are used to merge all visual images (HFaFFNN). DarkNet53 and NASNet-Mobile are then used to train two deep models that can be finely tailored to the datasets that were chosen. Later, the idea of using transfer learning to train both models is investigated, with the input feed generating images of localized lesions. The collected characteristics are then combined using the parallel max entropy correlation (PMEC) method in the next stage. An approach called entropy-kurtosis controlled whale optimization (EKWO) is used to avoid overfitting and to pick the most discriminating feature information. Three datasets HAM10000, ISBI2018, and ISBI2019 were used in this study.

In the majority of instances, a lack of data or an imbalance of data between classes included in the dataset is the fundamental cause of poor performance. A recent study [29] created a deep generative adversarial network (DGAN) multi-class classifier capable of generating images of skin disorders by learning the distribution of authentic data from publicly available datasets. To handle the class-imbalanced dataset, they used images from several Internet databases. Improving the DGAN model's stability during training is a major task. To analyze GAN's performance, they created two CNN models based on ResNet50 and VGG16 and tested the models with labelled and unlabelled data. DGAN outperformed conventional data augmentation by 91.1% for unlabelled and 92.3% for labelled datasets. CNN models with data augmentation obtained 70.8% accuracy on unlabelled data.

3. Materials and Methods

3.1. Dataset. This study utilized the HAM10000 [30] dataset, which stands for "Human Against Machine with 10,000 training photos." This dataset was used as the ISIC 2018 challenge training set (Task 3) [31]. In order to compile the collection, dermatoscopic photographs from diverse communities around the world were used. Data collection was conducted to include all of the vital diagnostic categories linked with the field of pigmented lesions. Therefore, seven distinct types of skin lesions, namely actinic keratosis (AKIEC), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanoma (MEL), nevus (NV), and vascular lesion (VASC), were included. The whole data collection contains 10015 images, each with 600 × 450 pixels resolution. Figure 1 depicts a selection of images representative of all groups of lesions.

A metadata file including demographic information for each lesion in question was supplied as supplementary data. In the other instances, the gold standard is a follow-up examination, expert consensus (confocal), or confirmation



(a) (b) (c) (d) (e) (f) (g)

Figure 1: Sample skin lesion images from the dataset: (a) AKIEC, (b) BCC, (c) BKL, (d) DF, (e) MEL, (f) NV, and (g) VASC.

by in-vivo confocal microscopy, with histopathology (histo) accounting for more than half of them.

The objective of the current work was the classification of skin lesions. In order to expedite the development of the model, the images in the dataset were rescaled to 256 × 256 pixels. The three distinct datasets were generated by partitioning the original dataset into three sections. They were train, validation, and test sets each consisting of 70%, 10%, and 20% of the whole dataset's images, respectively. The statistical breakdown for the seven different categories is presented in Table 1.

3.2. Data Augmentation. From Table 1, it was evident that the number of images for the seven classes spans a range of 115 (DF) to 6705 (NV). The HAM10000 dataset was clearly unbalanced. The skewed dataset may cause overfitting while training the model [32]. To solve this data scarcity for some classes that will affect the classification model's efficiency, data augmentation method was employed. Data augmentation is a method that undergoes random transformations on the images to increase the count of images for underrepresented classes without the overhead of collecting more images [33]. The possibilities for image augmentation were enormous such as rotation, translation, and flipping.

In this study, we did a number of operations as shown in Table 2. All of them were available on the Python library imaging [34]. The images after augmentation are shown in Figure 2.

4. Image Generation Conditional Generative Adversarial Networks (CGANs)

In addition, we investigated the idea of generating synthetic data to solve the class imbalance issue. The well-known CGAN [35] architecture is used to generate the images. Figure 3 depicts the high-level design of the network.

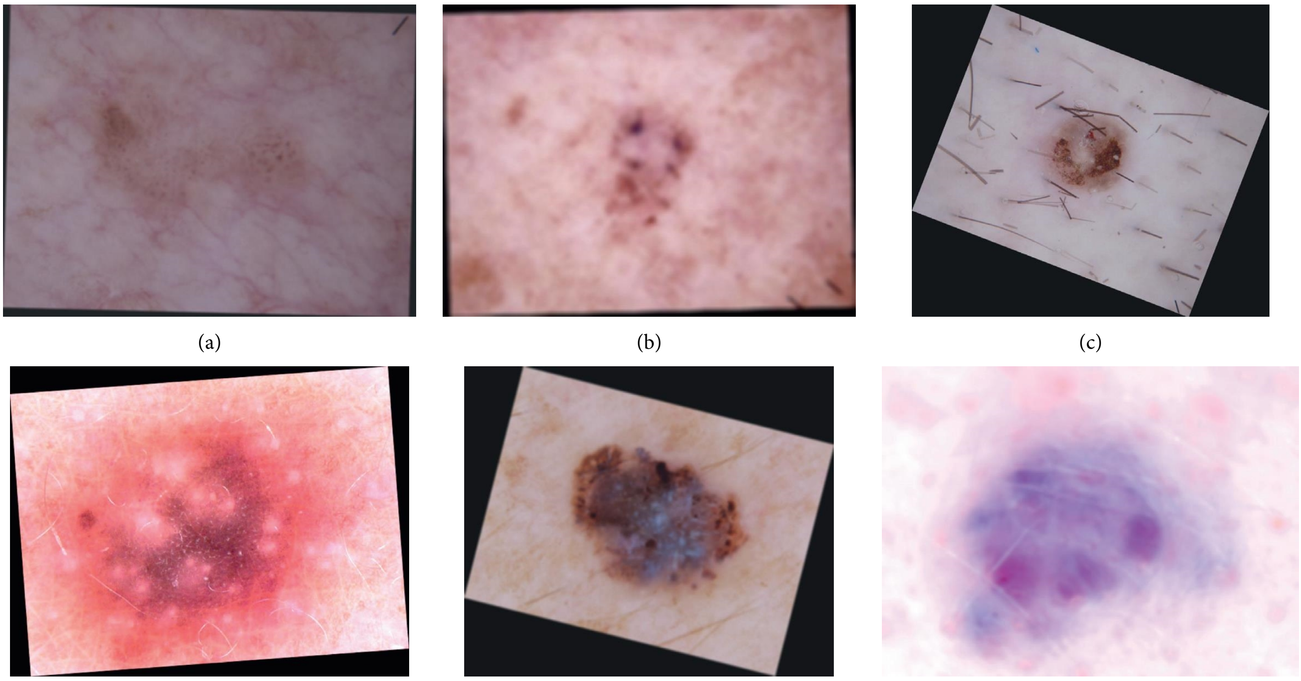
Generative adversarial networks (GANs) [36] normally use a generator to learn how to create new images and a discriminator to learn how to distinguish between artificial and genuine images. However, there was no mechanism to regulate the images generated, such as the development of multi-class data.

Table 1: Dataset statistics.

Class	Train	Validation	Test	Total	Benign/malignant
Actinic keratosis (AKIEC)	236	26	65	327	Benign or malignant
Basal cell carcinoma (BCC)	371	41	102	514	Malignant
Benign keratosis (BKL)	792	88	219	1099	Benign
Dermatofibroma (DF)	83	9	23	115	Benign
Melanoma (MEL)	802	89	222	1113	Malignant
Nevus (NV)	4828	536	1341	6705	Benign
Vascular lesion (VASC)	103	11	28	142	Benign or malignant

Table 2: Image augmentation techniques.

Flip	50% of horizontal and vertical flip on all images.
Affine	Translation: move each image −20 to +20% per axis Rotation: rotate each image by −30 to 30 degrees Scaling: zoom in each image by 0.5 to 1.5 times
Multiply	Multiplication of each image by a random value sampled from [0.8, 1.2].
Linear contrast	Change contrast by equation $127 + \text{alpha}^{*}$ (v-127) V: pixel value Alpha: samples from [0.6, 1.4]
Gaussian blur	Blur the images using Gaussian kernel with standard deviation sampled from the interval [0.0, 3.0].



(a) (b) (c)

(d) (e) (f)

Figure 2: Augmented images: (a) AKIEC, (b) BCC, (c) BKL, (d) DF, (e) MEL, and (f) VASC.

A conditional setting governs the training of the generator and discriminator in cGANs (such as class labels or data). Discrimination judgments will be based on both the generated images and their labels, with the former being more important to the discriminator. The ideal model can learn multimodal input-to-output mapping by being fed a range of contextual inputs. Following the creation of the model, we retained only the trained generator model that was utilized to generate the synthetic images. Some of the synthetic images generated by the trained CGAN model are displayed in Figure 4.

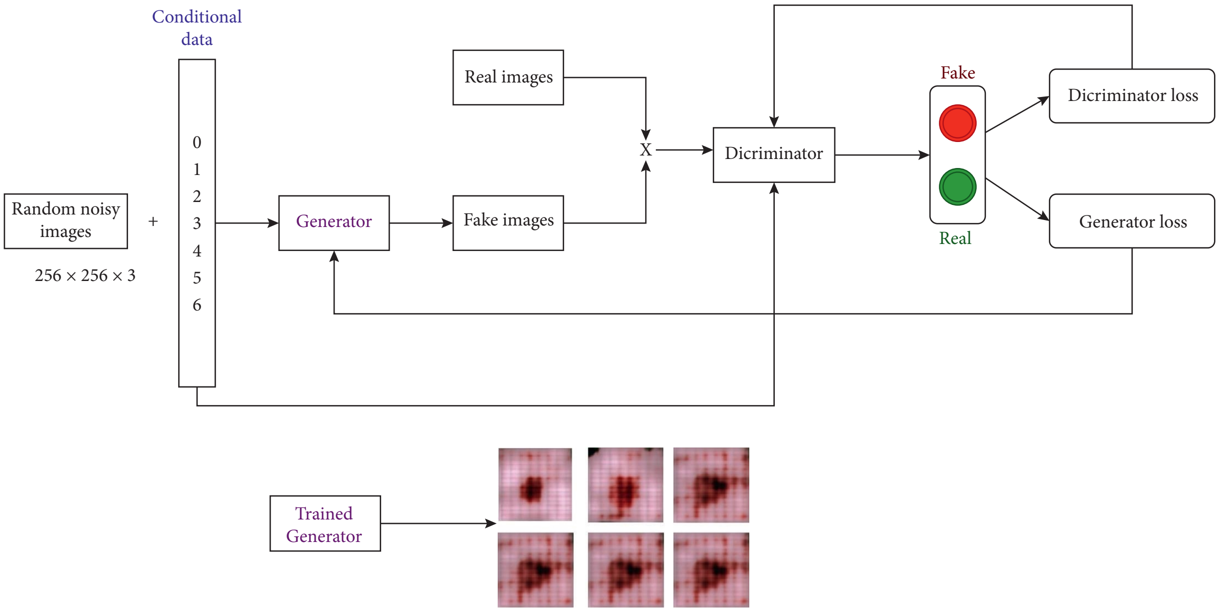


Figure 3: CGAN architecture.

5. Classification Model Development

The majority of real-world datasets suffer from data insufficiency issues, and constructing the most effective deep learning model for computer vision applications necessitates plenty of data. In addition, there will be insufficient processing capability if the dataset is enormous. With the development of the transfer learning [37] approach, such issues were resolved. Transfer learning is the most extensively utilized method for categorization tasks. As an alternative to training from scratch, it is a popular strategy in deep learning where pretrained models are employed as the starting point. It is usual practice to utilize deep models such as VGGNet and ResNet, which have been pretrained for a large and challenging image classification task such as the ImageNet 1000-class. The feature extractors of such deep models will be crucial for capturing critical features for classification. Only the dense layers at the output must be modified in accordance with the number of classes we wish to create.

In this study, we used three deep learning models, namely VGG16 [38], ResNet50, and ResNet101 [39], all of them pretrained on the ImageNet dataset. The architecture of the designed architectures is shown in Figure 5.

5.1. VGG16. There are 16 layers in VGG16 CNN. The input dimension to the network is (224, 224, 3). What makes VGG16 stand out from other implementations is it uses a 2×2 stride 2 filter with the same padding and max pool layer, instead of having many hyper-parameters. Throughout the architecture, the convolution and max pool layers are arranged in the same manner. For output, there are two fully connected (FC) and a softmax layer.

5.2. ResNet50. Deep learning research has seen a widespread trend toward increasing the number of layers in CNN architecture in order to improve performance. However, there was a vanishing/exploding gradient problem as layers rose. Therefore, the concept of "residual network" was introduced in architecture. When the network uses the "skip connection" idea, some subsequent connections are skipped, and the output is directly connected. The ResNet variation that contains 50 layers is called ResNet50.

5.3. ResNet101. The 101 layered ResNet is ResNet101. The architecture is more complex than ResNet50 as it contains more trainable parameters.

6. Ensemble Algorithm

We have trained three deep learning models for skin lesion prediction. However, we knew that a single algorithm might not provide the most accurate forecast for a specific dataset. There were limitations to machine learning methods, and developing a model with great accuracy is difficult. By combining multiple models, overall accuracy could be improved. The combination can be done by averaging the output of each model with two goals in mind: minimizing model error and preserving its generalizability. Each model predicted the likelihood of each class's forecast given its class. Taking the average of the prediction probabilities by the three models may result in a performance increase. The architecture for the ensemble algorithm is shown in Figure 6.

6.1. Performance Evaluation. The suggested model architecture was evaluated for its performance in predicting skin lesions using many performance assessment indicators. Accuracy, recall, precision, and F1-score are the four measures. True positives (TP) and false negatives (FN) are the numbers of positive images accurately predicted. In contrast, the number of incorrectly anticipated negative images is known as false positives (FP), while the number of accurately predicted negative images is known as true negatives (TN) [40].

6.1.1. Accuracy. The ratio of the number of classes a model successfully predicts to the total number of predictions.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}. \tag{1}$$

6.1.2. Precision. Precision is defined as the proportion of the number of correct predictions divided by the total number of positive class predictions. Precision is calculated as

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FP})}. \tag{2}$$

6.1.3. Recall. Recall is defined as the proportion of correct predictions divided by the number of actual count of the positive class in the dataset. Recall is calculated as

$$\text{Recall} = \frac{(\text{TP})}{(\text{TP} + \text{FN})}. \tag{3}$$

6.1.4. F1-Score. The F1-score represents the balance between precision and recall.

$$\text{F1} - \text{score} = \frac{(2 * \text{Precision} * \text{Recall})}{(\text{Precision}\,\text{Recall})}. \tag{4}$$

7. Results

7.1. Transfer Learning Model with the Unbalanced Dataset. Following the conclusion of the training for the VGG16, ResNet50, and ResNet101 architectures, predictions were made on the test set to understand how well they performed. The final layer of our design is called the softmax, and it was this layer that was responsible for producing the prediction probability of each of the seven classes. It is necessary to be aware of each model's performance before selecting the most appropriate model for the classification of skin lesions.

An analysis was carried out on the data obtained from the VGG16, ResNet50, and ResNet101 models' respective experiments. Figure 7 depicts the validation accuracy, error rate, and loss plots for each of the three models. It is possible to see that the accuracy has improved while the loss has decreased. Therefore, there was no evidence suggesting that any models would overfit.

The confusion matrices on the test predictions are shown in Figure 8, indicating the number of correct and incorrect predictions based on each class in the test dataset (AKIEC, BCC, BKL, DF, MEL, NV, and VASC) on the test set.

The performance evaluation metrics of the three models are shown in Table 3.

7.1.1. The Effect of the Ensemble Algorithm. The ensemble model combines the mean of the predictions by all three models. The confusion matrix of the ensemble model is shown in Figure 9. The performance analysis of the ensemble model on the unbalanced dataset is shown in Table 3. Finally, the class-wise performance is shown in Table 4.

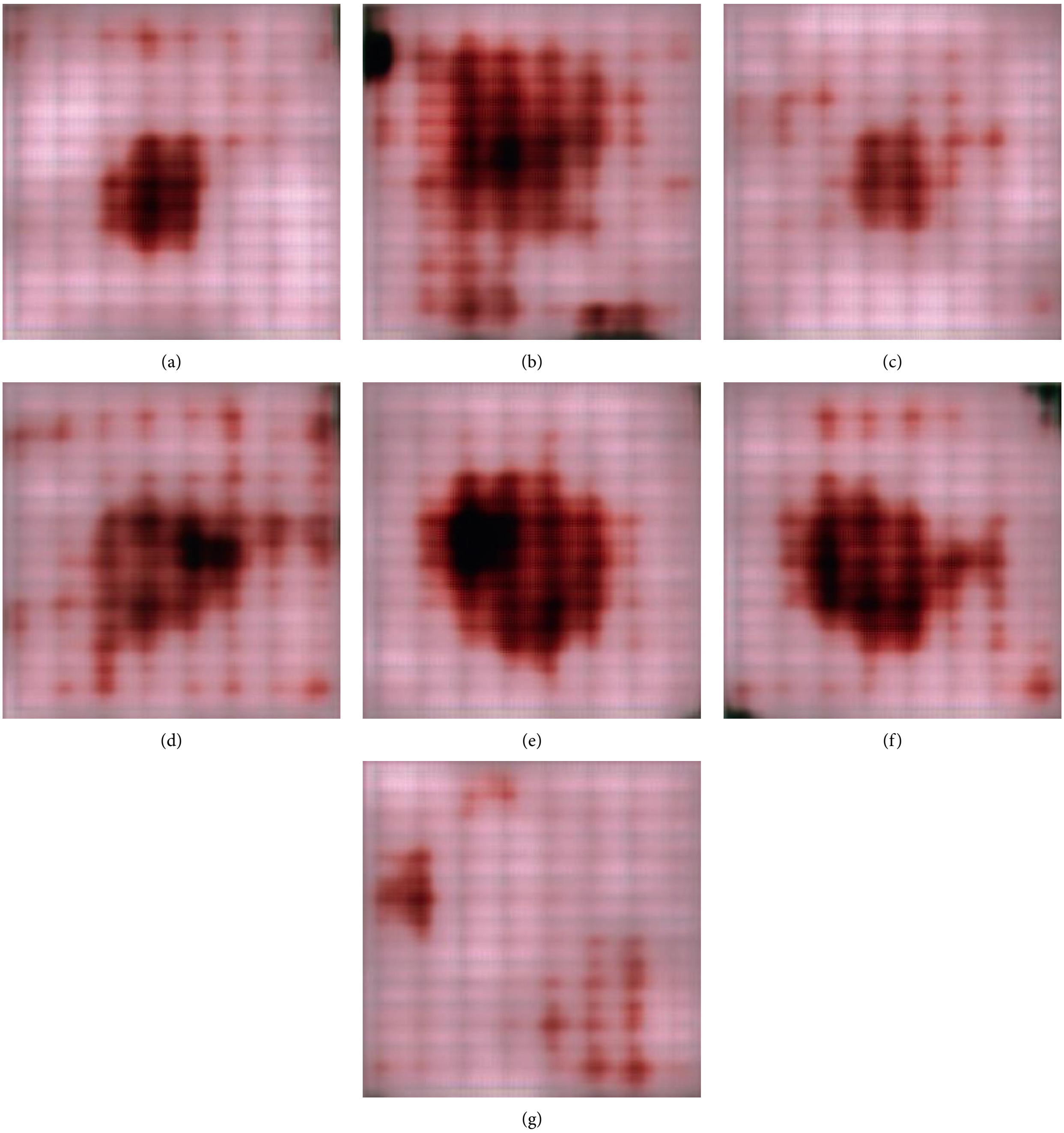
7.2. Transfer Learning Model on the Balanced Data Obtained by Data Augmentation. Several image augmentation techniques were used to increase the number of images within the six lesions. The skin lesion type "NV" images were excluded from the data augmentation process since it was an overrepresented class. Only underrepresented groups underwent the image augmentation procedure. In addition, we attempted to create synthetic images for each class. However, as shown in Figure 4, the produced images did not appear to possess the characteristics that differentiate the seven types of skin lesions. We anticipate that such data will not be suitable for training the model for optimal performance. Therefore, the resulting CGAN data were excluded from the training data with augmentation. The performance of the developed models VGG16, ResNet50, and ResNet101 is shown in Table 5.

7.2.1. The Effect of the Ensemble Algorithm. The performance analysis of the ensemble model on the balanced dataset is shown in Table 5. The class-wise performance is shown in Table 4.

8. Discussion

There was a wide variety of categorizations for skin lesions, some of which were cancerous while others were benign. It is essential to determine the specific type of skin lesion to determine whether the condition may progress to cancer and to ensure that the appropriate therapy is administered. Obtaining a cancer diagnosis at an earlier stage is essential if one wants to experience a full recovery from the disease. A delayed detection could cause the problem to become more complicated, putting a person's life in danger.

It has been discovered that deep learning is the most effective way of determining the different types of skin lesions. However, to extract the appropriate features for identifying the various classes without the participation of a human, they need a massive amount of the sampled data acquired from patients in each class [41]. However, the collection of such a massive volume of labelled data is nearly impossible, particularly in the field of medicine. As a consequence of this, the majority of the medical datasets that are accessible to the public have a problem with data imbalance [42]. As can be observed in Table 1, the dataset that was used for this investigation has some serious imbalances. Nevus, a type of benign lesion, accounts for the majority of the images and has an abundance of them in comparison to other categories. If you train the model using this kind of dataset, there is a chance that it will have a bias for the category that has the most images. The classification of skin lesions with a relatively low number of available examples for training



(a)

(b)

(c)

(d)

(e)

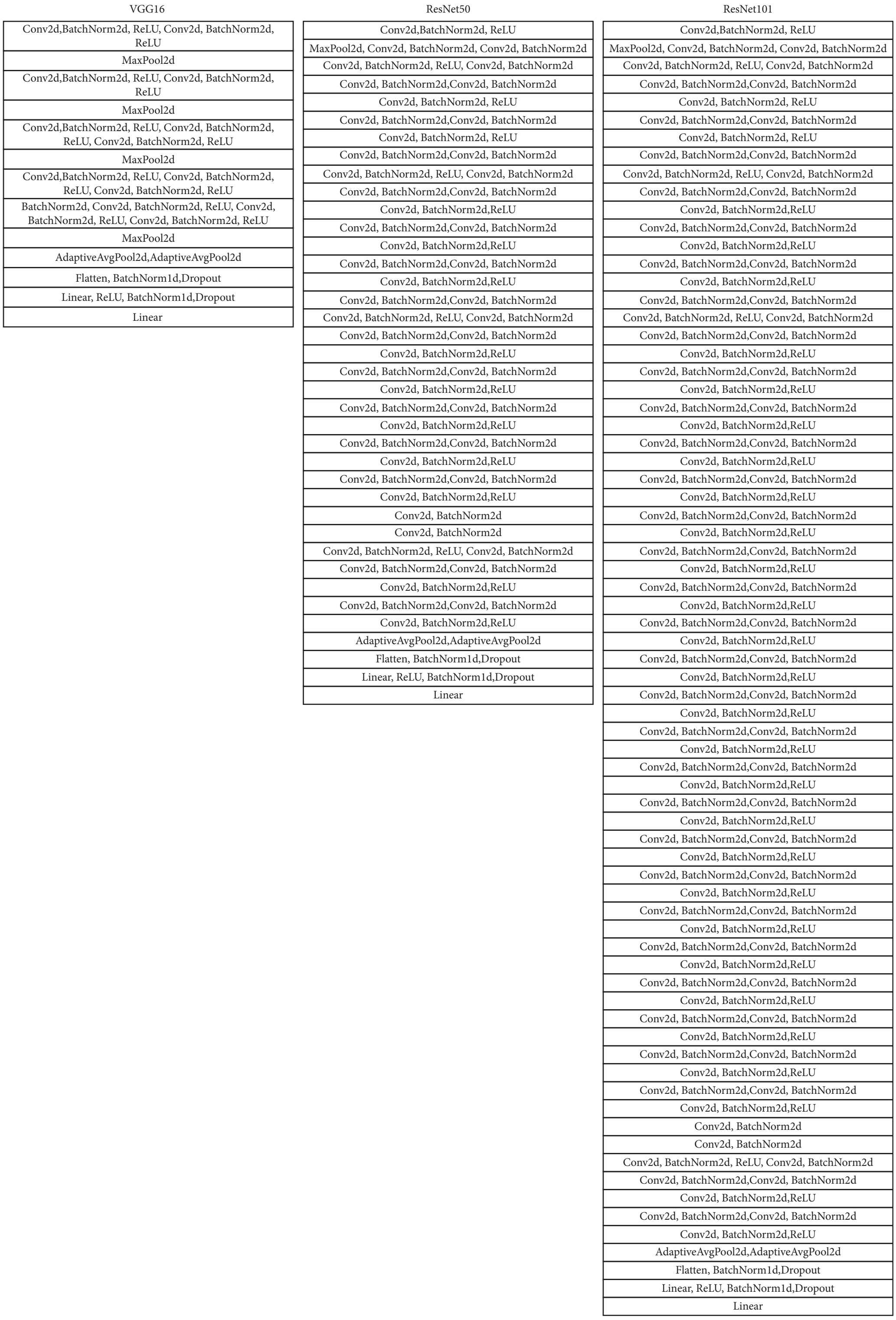
(f)

(g)

Figure 4: Some of the synthetic images from CGAN: (a) AKIEC, (b) BCC, (c) BKL, (d) DF, (e) MEL, (f) NV, and (g) VASC.

ended up being incorrectly identified as the NV class. A similar thing can be seen from the confusion matrices in Figure 8, which show that the class dermatofibroma (DF), which has the fewest number of images, is most commonly misdiagnosed as nevus.

Data augmentation was the most straightforward response to this problem. However, data augmentation might be one of two different types: picture transformations using the conventional methods or the development of synthetic images using the capabilities of GAN architecture. Several of the earlier researchers successfully implemented the GAN algorithm as a means of image augmentation, and they obtained performance improvements as a result [29, 43]. However, the synthetic images that were produced by our CGAN generator and displayed in Figure 4 did not appear to be effective enough to contribute to improved classification performance even though a prior work successfully applied the CGAN to produce synthetic skin lesion images using the HAM10000 dataset. Training with such a significantly better dataset has the potential to assist improve the proposed model's overall performance. However, the CGAN architecture that we designed was unable to generate images that were similar to genuine images; in fact, even humans could tell that these images had been artificially created by looking



VGG16

Conv2d,BatchNorm2d, ReLU, Conv2d, BatchNorm2d, ReLU
MaxPool2d
Conv2d,BatchNorm2d, ReLU, Conv2d, BatchNorm2d, ReLU
MaxPool2d
Conv2d,BatchNorm2d, ReLU, Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d, ReLU
MaxPool2d
Conv2d,BatchNorm2d, ReLU, Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d, ReLU
BatchNorm2d, Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d, ReLU
MaxPool2d
AdaptiveAvgPool2d,AdaptiveAvgPool2d
Flatten, BatchNorm1d,Dropout
Linear, ReLU, BatchNorm1d,Dropout
Linear

ResNet50

Conv2d,BatchNorm2d, ReLU
MaxPool2d, Conv2d, BatchNorm2d, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d
Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
AdaptiveAvgPool2d,AdaptiveAvgPool2d
Flatten, BatchNorm1d,Dropout
Linear, ReLU, BatchNorm1d,Dropout
Linear

ResNet101

Conv2d,BatchNorm2d, ReLU
MaxPool2d, Conv2d, BatchNorm2d, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d
Conv2d, BatchNorm2d
Conv2d, BatchNorm2d, ReLU, Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
Conv2d, BatchNorm2d,Conv2d, BatchNorm2d
Conv2d, BatchNorm2d,ReLU
AdaptiveAvgPool2d,AdaptiveAvgPool2d
Flatten, BatchNorm1d,Dropout
Linear, ReLU, BatchNorm1d,Dropout
Linear

Figure 5: Transfer learning architectures.

at them. The model could only produce the pink hue presented in the actual image and it was evident that they were unable to generate skin lesion images that had any distinctive characteristics for each class.

However, the conventional image enhancement was successful in accomplishing the goal of achieving the performance increase of all three models that were built. When using the balanced dataset, all of the performance evaluation

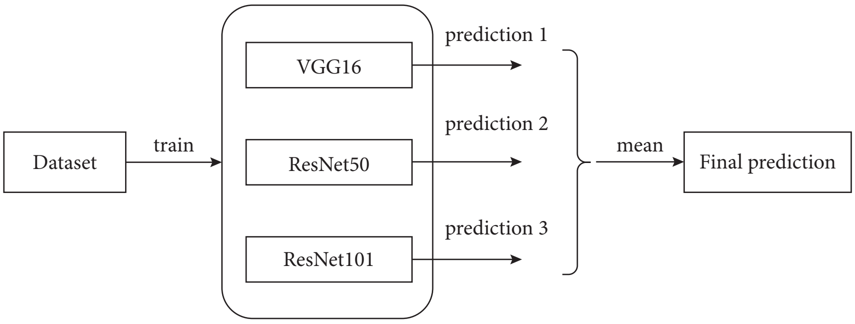
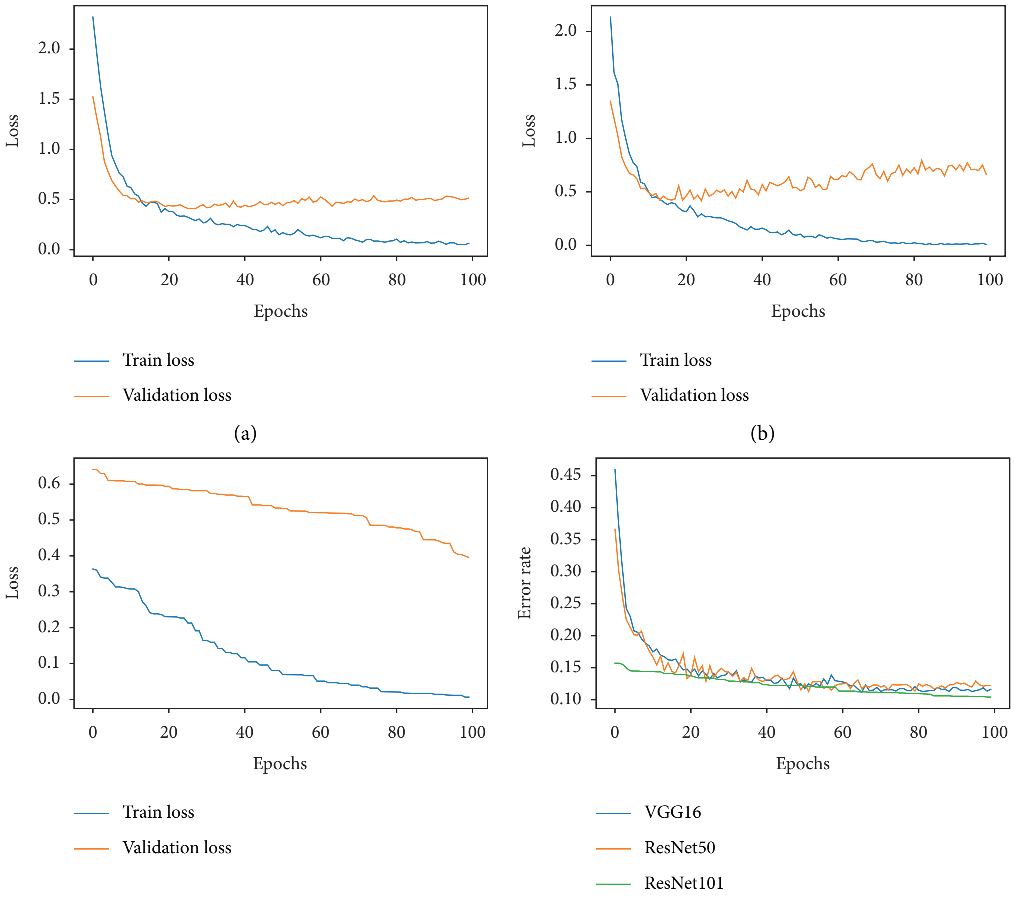


Figure 6: Ensemble algorithm.



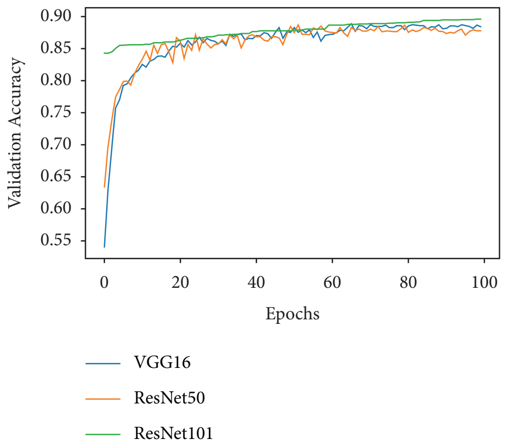
(a)

(b)

(c)

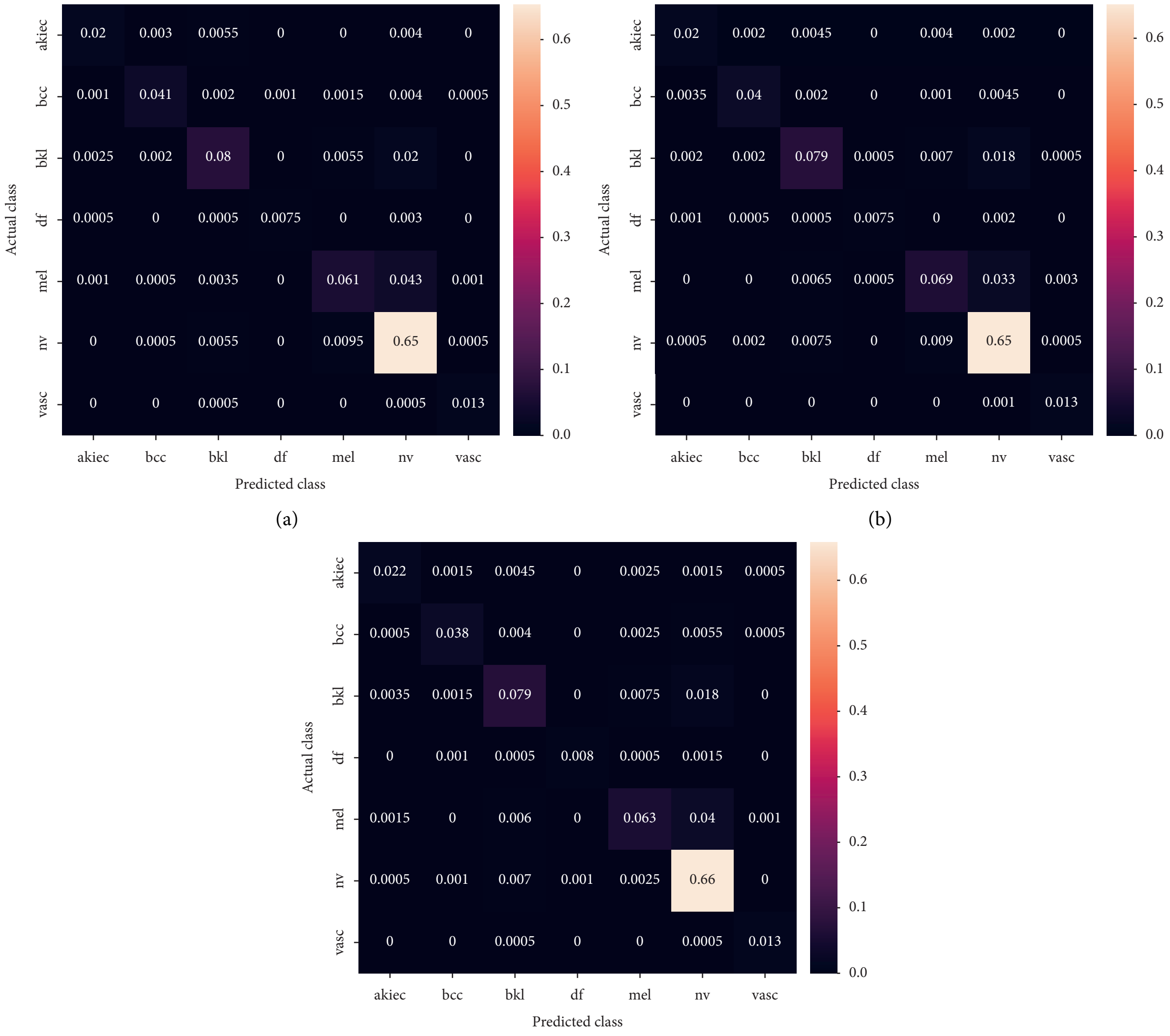
(d)

Figure 7: Continued.



(e)

FIGURE 7: (a) Loss plot for VGG16, (b) loss plot for ResNet50, (c) loss plot for ResNet101, (d) error rate, and (e) validation accuracy of the three models.



(a)

(b)

(c)

FIGURE 8: Confusion matrices: (a) VGG16, (b) ResNet50, and (c) ResNet101.

Table 3: Performance evaluation metrics of the three models.

Model	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
VGG16	87.7	75.08	84.66	79.06
ResNet50	87.9	75.57	81.69	78.01
ResNet101	88.15	75.96	84.48	79.57
Ensemble model	90	80.66	88.06	83.77

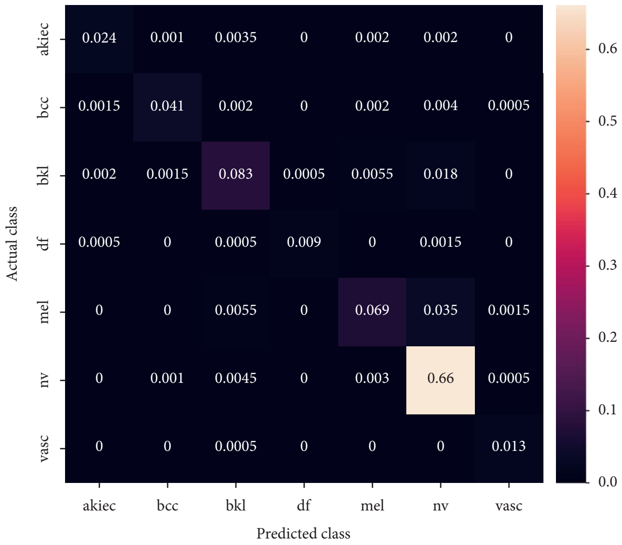


Figure 9: Confusion matrices of the ensemble model.

measures show a higher value, which was achieved by having less instances of incorrect classifications. The ensemble model was another clever strategy that was utilized to obtain a higher level of performance. If the predictions from all three models are combined, there is a chance that the overall performance will be improved. It is possible that some models will not be able to identify certain classes, and the probability of prediction of those classes might be low. However, taking the average of the three models' predictions might help increase the probability of correctly classifying those classes, which would help reduce the number of false negatives and false positives. Within the scope of this study, the combination of all three models unequivocally demonstrated an increase in performance (Tables 3 and 5). Incorporating the ensemble of models that were trained on the balanced dataset may also improve the performance of virtually all of the classes (Table 4).

8.1. Performance Comparison with Previous Works. A lot of previous works played on the HAM10000 dataset as a classification of skin lesions. Reference [44] developed a MobileNet model for classifying skin lesions. But the overall achieved accuracy of the classification was 83.1%. Reference [45] followed the transfer learning approach for the efficient feature extraction; they used ResNet50 and ResNet101 pretrained models. The feature selection process was done on the huge amount of extracted deep features. The selected features were given to the SVM and radial basis function (RBF) for the classification. But the developed model could achieve a performance of 89.8% even with the deep feature extraction and feature selection. Reference [46] also used MobileNet for the same task. The model's performance was tried to improve by the upscaling and augmentation methods, and the researchers succeeded in achieving the task. But the improved accuracy was limited to 83.23%. Also, [47] achieved 85.8% on the HAM10000 dataset. Reference [48] enhanced the images by local color-controlled histogram intensity values before training the CNN model. The developed model could achieve an accuracy of 90.67%. As shown in Table 6, all the comparisons are done with respect to our work.

The currently developed model could achieve greater results, but the dataset suffered from an imbalanced data problem, which had a direct impact on the performance of the model developed from the dataset. With the typical data augmentation technique, performance has been enhanced, but certain of the lesion classes in the dataset continues to suffer from poor detection, particularly melanoma—a particularly serious form of skin cancer. In order to obtain high performance for the low-performing classes in this study, a future study will integrate a more extensive dataset derived from various publicly available skin disease datasets. In addition, a study of computation time vs accuracy will be conducted to see how the model could be implemented in real-time, low-power medical devices.

Table 4: Class-wise performance of the ensemble models on balanced and unbalanced datasets.

Ensemble models	Unbalanced dataset			Balanced dataset		
Class of skin lesion	Recall (%)	Precision (%)	F1-score (%)	Recall (%)	Precision (%)	F1-score (%)
AKIEC	73.84	85.71	79.33	84.61	94.82	89.43
BCC	80.39	92.13	85.86	90.19	94.84	92.46
BKL	75.34	83.33	79.13	84.93	90.73	87.73
DF	78.26	94.73	85.71	95.65	95.65	95.65
MEL	61.71	84.56	71.35	72.07	92.48	81.01
NV	98.65	91.62	95.00	99.03	93.91	96.40
VASC	96.42	84.37	90	96.42	90	93.10

Table 5: Performance evaluation metrics of the three models on augmented dataset.

Model	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
VGG16	92	85.07	91.84	88.07
ResNet50	92.1	84.65	88.65	86.26
ResNet101	92.25	85.40	90.63	87.79
Ensemble model	93.5	88.98	93.20	90.82

Table 6: Comparative analysis.

Previous works	Accuracy (%)
[44]	83.1
[45]	89.8
[46]	83.23
[47]	85.8
[48]	90.67
Proposed work (augmentation + ensemble model)	93.5

9. Conclusions

The cancer that affects the skin is one of the deadliest forms of the disease, and identifying it using methods such as dermatoscopy and the naked eye was time-consuming. Identifying skin cancers at an early stage may allow for treatment that prevents them from progressing to more fatal forms. This work aimed to develop an effective deep learning model that could be used for the early diagnosis of seven different skin lesions. However, the process was not as easy as it could have been because the HAM10000 that was being used for this work appeared to be unbalanced. In order to find a solution to this problem, we investigated data augmentation methods, conventional image translations, and image generation possibilities. However, the work showed that the GAN architecture was not properly trained to generate the appropriate authentic-looking skin lesion images that qualify as the model training input. This was revealed by the GAN architecture failing to generate these images. On the other hand, the image transformations might be able to produce a magnificent dataset to solve the problem of image imbalance. In addition, an ensemble model that consisted of the VGG16, ResNet50, and ResNet101 models that had been trained on both balanced and unbalanced datasets was developed, and its performance was analyzed. According to the findings of the study, an ensemble of models that had been trained on a balanced dataset was able to produce the best results for skin lesion classification while also displaying less bias toward the category that contained the most significant number of examples (nevus). The accuracy of the model, which was obtained to be 93.5%, was significantly better than many of the previous efforts that had been made on the same dataset.

Data Availability

The dataset is taken from "P. Tschandl, C. Rosendahl, and H. Kittler, 'Data descriptor: The HAM10000 dataset, a large collection of multisource dermatoscopic images of common pigmented skin lesions', *Sci. Data*, 2018.

Conflicts of Interest

"The authors declare that there are no conflicts of interest regarding the publication of this article."

Acknowledgments

This study was supported by Princess Nourah Bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R235), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

References

[1] B. K. Armstrong and A. Kricker, "Skin cancer," *Dermatologic Clinics*, vol. 13, no. 3, pp. 583–594, 1995.

[2] M. C. F. Simões, J. J. S. Sousa, and A. A. C. C. Pais, "Skin cancer and new treatment perspectives: a review," *Cancer Letters*, vol. 357, no. 1, pp. 8–42, 2015.

[3] R. Cassano, M. Cuconato, G. Calviello, S. Serini, and S. Trombino, "Recent advances in nanotechnology for the treatment of melanoma," *Molecules*, vol. 26, no. 4, p. 785, 2021.

[4] M. Buljan, V. Bulat, M. Šitum, L. L. Mihić, and S. Stanić-Duktaj, "Variations in clinical presentation of basal cell carcinoma," *Acta Clinica Croatica*, vol. 47, no. 1, p. 25, 2008.

[5] H. E. Kanavy and M. R. Gerstenblith, "Ultraviolet radiation and melanoma," *Seminars in Cutaneous Medicine and Surgery*, vol. 30, no. 4, pp. 222–228, 2011.

[6] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[7] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN

estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[8] I. U. Khan, N. Aslam, T. Anwar et al., "Remote diagnosis and triaging model for skin cancer using EfficientNet and extreme gradient boosting," *Complexity*, vol. 2021, p. 13, 2021.

[9] K. Wolff, R. C. Johnson, A. Saavedra, and E. K. Roh, *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, McGraw Hill Professional, Taiwan, 2017.

[10] G. Argenziano and H. P. Soyer, "Dermoscopy of pigmented skin lesions--a valuable tool for early," *The Lancet Oncology*, vol. 2, no. 7, pp. 443–449, 2001.

[11] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *The Lancet Oncology*, vol. 3, no. 3, pp. 159–165, 2002.

[12] G. Fabbrocini, V. De Vita, F. Pastore et al., "Teledermatology: from prevention to diagnosis of nonmelanoma and melanoma skin cancer," *International Journal of Telemedicine and Applications*, vol. 2011, Article ID 125762, 5 pages, 2011.

[13] A.-R. A. Ali and T. M. Deserno, "A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data," *Med. Imaging 2012 Image Perception, Obs. Performance, Technol. Assess*, vol. 8318, pp. 421–431, 2012.

[14] C. Sinz, P. Tschandl, C. Rosendahl et al., "Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin," *Journal of the American Academy of Dermatology*, vol. 77, no. 6, pp. 1100–1109, 2017.

[15] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," *Machine Learning In Medical Imaging*, vol. 2015, pp. 118–126, 2015.

[16] N. H. Khan, M. Mir, L. Qian et al., "Skin cancer biology and barriers to treatment: recent applications of polymeric micro/nanostructures," *Journal of Advanced Research*, vol. 36, pp. 223–247, 2022.

[17] C. Barata, M. E. Celebi, and J. S. Marques, "A survey of feature extraction in dermoscopy image analysis of skin cancer," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1096–1109, 2019.

[18] M. E. Celebi, H. A. Kingravi, B. Uddin et al., "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 362–373, 2007.

[19] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 994–1004, 2017.

[20] M. Havaei, A. Davy, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[21] F. M. Osman, R. Marti, R. Zwiggelaar et al., "End-to-end breast ultrasound lesions recognition with a deep learning approach," *Medical imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10578, Article ID 1057819, 2018.

[22] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin cancer classification using deep learning and transfer learning," in *Proceedings of the 2018 9th Cairo international biomedical engineering conference (CIBEC)*, pp. 90–93, IEEE, Cairo, Egypt, December 2018.

[23] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2-A dermoscopic image database for research and benchmarking," in *Proceedings of the 2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 5437–5440, IEEE, Osaka, Japan, July 2013.

[24] A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang, "Fusing fine-tuned deep features for skin lesion classification," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 19–29, 2019.

[25] N. Gessert, T. Sentker, F. Madesta et al., "Skin lesion classification using cnns with patch-based attention and diagnosis-guided loss weighting," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 495–503, 2020.

[26] L. Liu, L. Mou, X. X. Zhu, and M. Mandal, "Automatic skin lesion classification based on mid-level feature learning," *Computerized Medical Imaging and Graphics*, vol. 84, Article ID 101765, 2020.

[27] N. Aburaed, A. Panthakkan, M. Al-Saad, S. A. Amin, and W. Mansoor, "Deep convolutional neural network (DCNN) for skin cancer classification," in *Proceedings of the 2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pp. 1–4, IEEE, Glasgow, UK, December 2020.

[28] R. Garg, S. Maheshwari, and A. Shukla, "Decision support system for detection and classification of skin cancer using CNN," in *Innovations in Computational Intelligence and Computer Vision*, pp. 578–586, Springer, Singapore, 2021.

[29] M. Heenaye-Mamode Khan, N. Gooda Sahib-Kaudeer, M. Dayalen et al., "Multi-class skin problem classification using deep generative adversarial network (DGAN)," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1797471, 2022.

[30] P. Tschandl, C. Rosendahl, and H. Kittler, "Data descriptor: the HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, Article ID 180161, 2018.

[31] N. C. F. Codella in *Proceedings of the 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)*, Washington, DC, USA, April 2018.

[32] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, and J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," *Computerized Medical Imaging and Graphics*, vol. 88, Article ID 101843, 2021.

[33] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, Article ID 100036, 2021.

[34] A. B. Jung et al., *Imgaug*, 2020.

[35] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv Prepr. arXiv1411.1784*, 2014.

[36] G. Ian, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Morgan Kaufmann Publishers Inc, San Francisco; CA, USA, 2014.

[37] T. Lisa and S. Jude, *Transfer Learning Handbook of Research on Machine Learning Applications*, IGI Glob, 2009.

[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prepr. arXiv1409.1556*, 2014.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Computer Vision and pattern Recognition*, vol. 2016, p. 90, 2016.

[40] Y. Liu, J. Ma, J. Niu, Y. Zhang, and W. Wang, “Roadside units deployment for content downloading in vehicular networks,” in *Proceedings of the 2013 International ConferenceOn Communication(ICC)*, June 2013.

[41] H. Rashid, M. A. Tanveer, and H. A. Khan, “Skin lesion classification using GAN based data augmentation,” in *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 916–919, IEEE, Berlin, Germany, July 2019.

[42] J. Diz, G. Marreiros, and A. Freitas, “Applying data mining techniques to improve breast cancer diagnosis,” *Journal of Medical Systems*, vol. 40, no. 9, pp. 203–207, 2016.

[43] Z. Qin, Z. Liu, P. Zhu, and Y. Xue, “A GAN-based image synthesis method for skin lesion classification,” *Computer Methods and Programs in Biomedicine*, vol. 195, Article ID 105568, 2020.

[44] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, “Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet,” in *Proceedings of the International Conference on Advanced Machine Learning Technologies and Applications*, pp. 165–176, Springer, Jaipur, India, May 2020.

[45] M. A. Khan, M. Y. Javed, M. Sharif, T. Saba, and A. Rehman, “Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification,” in *Proceedings of the 2019 international conference on computer and information sciences (ICCIS)*, pp. 1–7, IEEE, Saudi Arabia, April 2019.

[46] W. Sae-Lim, W. Wettayaprasit, and P. Aiyarak, “Convolutional neural networks using MobileNet for skin lesion classification,” in *Proceedings of the 2019 16th international joint conference on computer science and software engineering (JCSSE)*, pp. 242–247, IEEE, Chonburi, Thailand, July 2019.

[47] H.-W. Huang, B. W.-Y. Hsu, C.-H. Lee, and V. S. Tseng, “Development of a light-weight deep learning model for cloud applications and remote diagnosis of skin cancers,” *The Journal of Dermatology*, vol. 48, no. 3, pp. 310–316, 2021.

[48] M. A. Khan, T. Akram, M. Sharif, S. Kadry, and Y. Nam, “Computer decision support system for skin cancer localization and classification,” *Web Of science*, vol. 68, no. 1, pp. 1041–1064, 2021.