# Two-Stages Input Space Image Augmentation and Interpretable Technique for Accurate and Explainable Skin Cancer Diagnosis

Catur Supriyanto *, Abu Salam , Junta Zeniarja , Adi Wijaya

Posted Date: 2 November 2023

doi: 10.20944/preprints202311.0161.v1

Keywords: deep learning; skin cancer; image augmentation; GAN; geometric augmentation; image classification; interpretable technique

*Article*

# Two-Stages Input Space Image Augmentation and Interpretable Technique for Accurate and Explainable Skin Cancer Diagnosis

**Catur Supriyanto [1,2,*], Abu Salam [1,2], Junta Zeniarja [1,2] and Adi Wijaya [3]**

[1] Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia
[2] DREAMS-Dinus Research Group for AI in Medical Science, Universitas Dian Nuswantoro, Semarang, Indonesia
[3] Department of Health Information Management, Universitas Indonesia Maju, Jakarta, Indonesia
* Correspondence: catur.supriyanto@dsn.dinus.ac.id

**Abstract:** This research paper presents a deep learning approach to early detection of skin cancer using image augmentation techniques. The authors propose a two-stage image augmentation technique that involves the use of geometric augmentation and generative adversarial network (GAN) to classify skin lesions as either benign or malignant. This research utilized the public HAM10000 dataset to test the proposed model. Several pre-trained models of CNN were employed, namely Xception, Inceptionv3, Resnet152v2, EfficientnetB7, InceptionresnetV2, and VGG19. Our approach achieved accuracy, precision, recall, and F1-score of 96.90%, 97.07%, 96.87%, 96.97%, respectively, which is higher than the performance achieved by other state-of-the-art methods. The paper also discusses the use of SHapley Additive exPlanations (SHAP), an interpretable technique for skin cancer diagnosis, which can help clinicians understand the reasoning behind the diagnosis and improve trust in the system. Overall, the proposed method presents a promising approach to automated skin cancer detection that could improve patient outcomes and reduce healthcare costs.

**Keywords:** deep learning; skin cancer; image augmentation; GAN; geometric augmentation; image classification; interpretable technique

## 1. Introduction

Skin cancer is one of the most prevalent and potentially life-threatening forms of cancer worldwide. Early detection and diagnosis are crucial for effective treatment and improved patient recovery [1,2]. In recent years, convolutional neural networks (CNN) have revolutionized the field of medical image analysis compared to other advanced machine learning models, supervised and unsupervised, such as -nearest neighbor (KNN) and support vector machine (SVM), offering a promising approach for the automated detection of skin cancer [3]. CNNs have proven to be highly effective in extracting intricate patterns and features from medical images, making them an ideal tool for automating the process of skin cancer detection. This technology has the potential to assist dermatologists and healthcare professionals in identifying skin lesions, distinguishing between benign and malignant tumors.

Two datasets that are commonly used for skin cancer detection research are HAM10000 and International Skin Imaging Collaboration (ISIC). HAM10000 is a comprehensive dataset containing diverse dermoscopic images of pigmented skin lesions, a common category of skin cancer [4]. An advantage of the HAM10000 dataset lies in its relatively smaller size compared to the expansive ISIC dataset. This may be beneficial for researchers facing limited computational resources or who want to focus on a specific subset of skin lesions. However, the ISIC dataset has unique advantages, including larger scale and inclusion of additional metadata such as lesion location and patient age. The ISIC datasets have been used for segmentation tasks, but the availability of delineated segmentation masks is limited compared to the classification tasks [5]. The choice of data set often depends on the specific research investigation and the resources available for the study.

 doi:10.20944/preprints202311.0161.v1



The issue with skin cancer detection datasets is the imbalance in the number of data samples across different classes. This imbalance is observed in both the HAM10000 and ISIC 2017-2020 datasets. In the HAM10000 dataset, which includes a total of 10,015 images, the highest number of data samples can be seen in the melanoma category, with 6,705 images, while the lowest number of samples is present in the dermatofibroma category, consisting of 115 images [6]. Meanwhile, in the ISIC 2020 dataset, encompassing a total of 33,126 images, the most abundant data samples are found within the unknown (benign) category, which comprises 27,126 images, whereas the solar lentigo category contains the fewest data samples, with only 7 images [7]. Data imbalance can lead to biased results in classification because the model may be more likely to predict the over-represented class.

Several approaches can be employed to address the imbalance in the number of data, such as geometric transformation-based augmentation, feature space augmentation, and GAN-based augmentation [8]. Geometric data augmentation is a technique used in machine learning and computer vision to increase the diversity of a dataset by applying geometric transformations to the existing data. The approach changes the geometrical structure of images by shifting image pixels from their original placements to new positions without changing the pixel values. These transformations involve altering the position, orientation, or scale of the data while preserving their inherent characteristics. Geometric data augmentation is particularly useful for image data and is often applied to improve the performance of deep learning models. Some common geometric augmentations include: rotation, scaling, translation, shearing, flipping, cropping, and zooming.

In feature space data augmentation, there are two approaches, namely the undersampling and oversampling approach. In undersampling approach, the number of samples from the majority class is reduced to create a more balanced distribution between the classes. By reducing the number of majority class samples, undersampling can help prevent the model from being biased toward the majority class and can improve its ability to recognize the minority class. However, undersampling may result in a loss of potentially valuable information, so it should be applied carefully. In oversampling approach, additional samples from the minority class are generated to create a more balanced distribution between the classes. The goal is to increase the representation of the minority class to match the number of samples in the majority class, making the dataset more balanced. There are several oversampling methods, with one of the most commonly used techniques being SMOTE (Synthetic Minority Over-sampling Technique). SMOTE generates synthetic samples for the minority class by interpolating between existing data points. This helps to improve the model's ability to learn from the minority class and can lead to better classification results. However, it's important to be cautious with oversampling, as generating too many synthetic samples can lead to overfitting and reduced model generalization.

GAN-based augmentation refers to the use of generative adversarial networks (GANs) to generate synthetic data samples that can be used to augment an existing dataset [9]. This technique is particularly useful in cases where the original dataset is small or imbalanced, as it can help to increase the size of the dataset and balance the class distribution. GAN augmentation has been successfully applied in various domains, including medical imaging, natural language processing, and computer vision. Some examples of GAN augmentation in medical imaging include the generation of synthetic CT scans, MRI images, and X-ray images to aid in disease diagnosis and treatment.

In general, resampling approaches can be divided into two categories, namely input space data augmentation and feature space data augmentation. Input space resampling involves manipulating the original data instances themselves before any feature extraction. Meanwhile, feature space resampling is applied after feature extraction . Geometric transformations and GAN-based augmentations are categorized within input space data augmentation, whereas SMOTE is classified under feature space data augmentation. The benefit of input space data augmentation is its independence from the feature extraction method, providing greater flexibility in choosing feature extraction methods. Therefore, in this study we propose a two-step augmentation, including geometric and GAN-based augmentation, for early detection of skin cancer.

 doi:10.20944/preprints202311.0161.v1



## 2. Related Works

The utility of image augmentation and oversampling approaches in the identification of malignant skin melanoma stems from their ability to address challenges such as limited data and class imbalance. Abayomi et al. [10] proposed data augmentation strategy entails creating a new skin melanoma dataset using dermoscopic images from the publically available PH2 dataset. The SqueezeNet deep learning network was then trained using these modified images. The experiments in a binary classification scenario show a significant improvement in melanoma detection ability, with significant gains in accuracy 92.18%, sensitivity 80.77%, specificity 95.1%, and F1-score 80.84%.

SMOTE (Synthetic Minority Over-sampling Technique) is an oversampling method used to balance the number of samples between the majority and minority classes in a dataset. SMOTE randomly selects samples from the minority class and creates new synthetic samples by combining them with their nearest neighbors. This helps improve the classification performance on imbalanced datasets. However, SMOTE tends to introduce noise and affect classification prediction performance. Therefore, K-means SMOTE was developed to address SMOTE's limitations by using k-means clustering to group samples and generate synthetic samples only within clusters with fewer minority class instances [11]. K-means SMOTE has been shown to enhance classification performance on imbalanced datasets.

A study on DSCC_Net was proposed by Tahir et al. [12], a multi-classification deep learning model for diagnosing skin cancer using dermoscopic images. The proposed model was trained and tested on a dataset of 10,015 dermoscopic images of skin lesions, and the results showed that DSCC_Net outperformed other state-of-the-art models in terms of accuracy, sensitivity, specificity, and F1-score. The SMOTE Tomek technique is used to balance the dataset by generating synthetic samples for the minority class and removing noisy and borderline examples from both the minority and majority classes. The study concludes that DSCC_Net has the potential to be implemented in the medical field to improve skin cancer diagnosis.

Alam et al. [13] aimed to find a solution for classifying skin lesions using images with efficient performance. A novel framework was proposed to solve the problem of data imbalance. The classes in the dataset were not balanced, limiting the performance of deep learning models. Data augmentation techniques were used to increase the size of the dataset and resolve the data imbalance issue. The proposed framework was trained on the Skin Cancer MNIST: HAM10000 dataset. AlexNet, InceptionV3, and RegNetY-320-based deep learning models were trained on balanced and imbalanced datasets. The proposed framework was tuned on different hyperparameters, i.e., the learning rate, epochs, and batch size in which the learning rate was changed, but the epochs and batch size were kept constant. The results showed that deep learning-based models performed better on a balanced dataset than on an imbalanced dataset.

Sae-Lim et al. [14] discusses a skin lesion classification approach using a lightweight deep Convolutional Neural Network. The authors propose a modified version of MobileNet that achieves higher accuracy, specificity, sensitivity, and F1-score than the traditional MobileNet. They evaluate their model using the HAM10000 dataset and achieve promising results. The best accuracy only achieved 83.23%.

## 3. Materials and Methods

### *3.1. Dataset*

HAM10000 is a dataset containing clinical images of various pigmented skin lesions, including both malignant (cancerous) and benign cases. The dataset consists of 10,015 dermatoscopic images of various skin lesions. These images vary in their types and characteristics. The data is categorized into seven categories based on the type of skin lesion. These categories include melanocytic nevi (melanocytic nevus), melanoma, benign keratosis-like lesions (benign keratosis lesion), basal cell

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2023 doi:10.20944/preprints202311.0161.v1



carcinoma (basal cell carcinoma), actinic keratoses (actinic keratosis), vascular lesions (vascular lesion), and dermatofibroma. Figures 1 and 2 show the number and image samples of each category, respectively. Each image in the dataset is accompanied by clinical metadata that includes information such as patient age, gender, and the location of the skin lesion. Dermatology experts have provided annotations and diagnoses for each image in this dataset. These annotations include information about the type of lesion (whether it is malignant or benign) and its characteristics. The images in the HAM10000 dataset are of high resolution and good quality, making them suitable for in-depth analysis and diagnosis. HAM10000 is widely used by researchers and machine learning practitioners to develop and evaluate algorithms for skin cancer diagnosis. This data has played a crucial role in advancing the field of computer-aided skin cancer diagnosis. HAM10000 is a publicly available dataset, allowing researchers and developers to access and use it for non-commercial purposes.

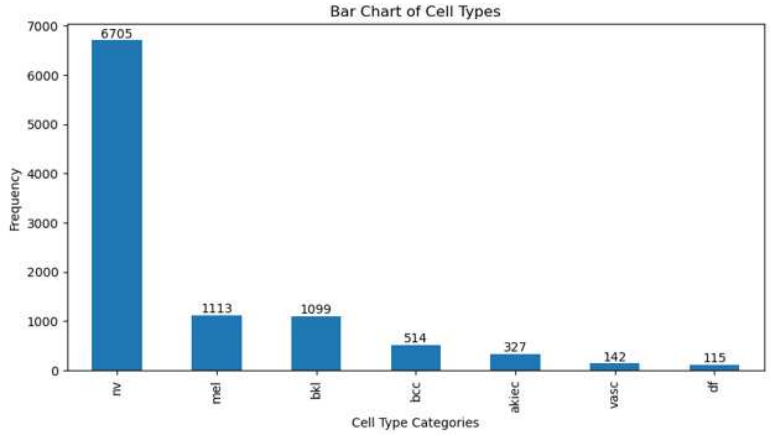



**Figure 1.** Category distribution of HAM10000 dataset

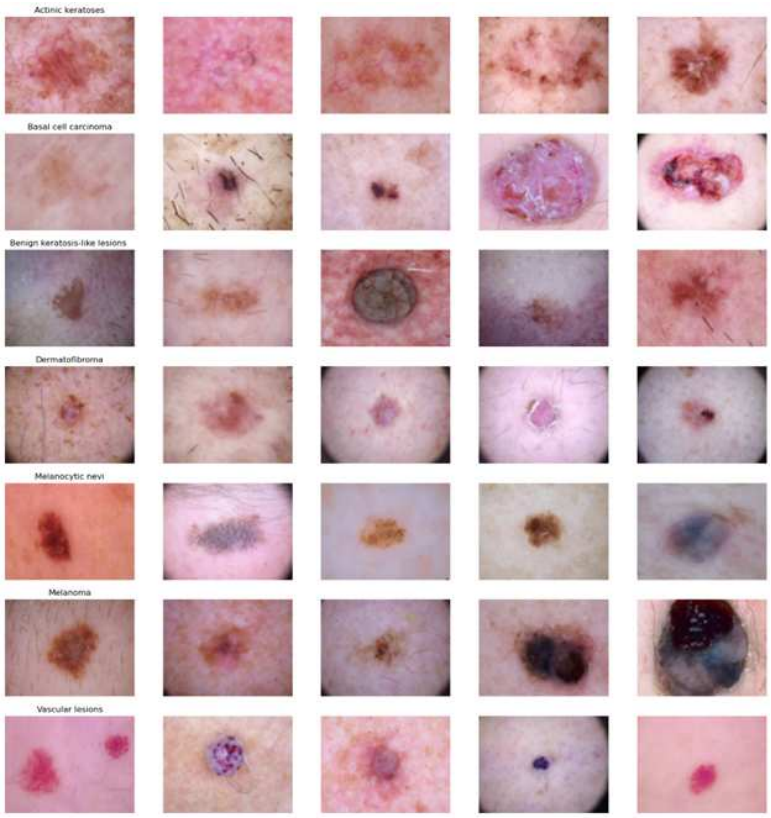



**Figure 2.** Example image of each class in HAM10000 dataset. From top to bottom: akiec, bcc, bkl, df, nv, mel, and vasc.

 doi:10.20944/preprints202311.0161.v1



### 3.2. The Proposed Skin Cancer Detection

Transfer learning is especially beneficial when there is minimal data for the new task or when building a deep model from scratch would be computationally expensive and time-consuming. In this research, six pre-trained CNN models were utilized for skin cancer classification. In this study, skin cancer classification employed six pre-trained CNN models, which included Xception, Inceptionv3, Resnet152v2, EfficientnetB7, InceptionresnetV2, and VGG19. In order to build a robust model, we will apply augmentation techniques to categories that have a limited number of images. Two stages of input space augmentation, namely Geometric and GAN augmentations, are proposed. Figure 3 shows the flow of skin cancer detection with the proposed augmentation.

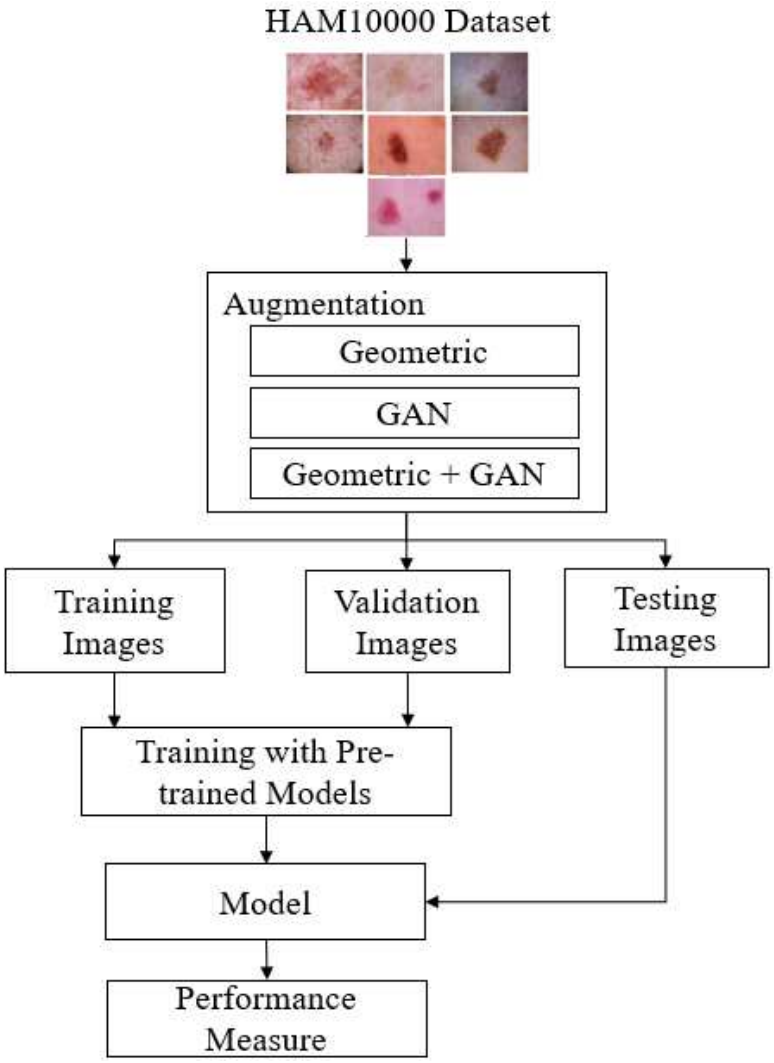



**Figure 3.** The proposed skin cancer detection.

Geometric augmentation is one of the data augmentation techniques used in computer image processing, particularly in the context of deep learning and pattern recognition. The goal of geometric augmentation is to enhance the diversity of training data by altering the geometry of the original image without changing the associated labels or class information related to that image. In this way, machine learning models can learn more general patterns and are not overly dependent on specific poses, orientations, or geometric transformations.

Some commonly used geometric augmentation techniques in deep learning include:

1. Rotation : images can be rotated by a certain angle, either clockwise or counterclockwise.

 doi:10.20944/preprints202311.0161.v1



2. Translation : images can be shifted in various directions, both horizontally and vertically.
3. Scaling : images can be resized to become larger or smaller.
4. Shearing : images can undergo linear distortions, such as changing the angles.
5. Flipping : images can be flipped horizontally or vertically.
6. Cropping : parts of the image can be cut out to create variations.
7. Perspective Distortion: Images can undergo perspective distortions to change the viewpoint.

By applying these geometric augmentation techniques, training data can be enriched with geometric variations, which helps machine learning models become more robust to variations in real-world images. This allows the model to perform better in pattern recognition tasks, such as object classification, object detection, or image segmentation, even when objects appear in different orientations or poses.

GAN [15] augmentation refers to the use of Generative Adversarial Networks (GANs) as one of the data augmentation techniques in the context of machine learning, especially in image processing. GAN is an artificial neural network architecture consisting of two models, the generator and the discriminator, that compete in a game to improve their capabilities [16]. In the context of data augmentation, GAN augmentation involves using the GAN generator to create additional data that is similar to the existing training data. The GAN generator tries to create images that appear authentic, while the GAN discriminator attempts to distinguish between images generated by the generator and real images.

By combining the images generated by the GAN generator with the training data, the dataset can be enhanced with image variations that appear realistic. GAN augmentation has been proven effective in improving the performance of machine learning models, especially in image recognition tasks such as object classification, object detection, or image segmentation, as it can create more diverse and relevant image variations.

### *3.3. Design of Experiments*

In the experiment, 20% of the 10,015 images, which is 2003 images, were utilized for testing, while the remaining 8,012 images were split into 90% (7,210) for training and 10% (802) for validation. Three methods are used to oversample the data: geometric, GAN, and geometric+GAN augmentations. Experiments were carried out using python and run on NVidia DGX Station A100 with 40 GB GPU, 64 cores CPU, and 512 GB DDR4 RAM.

Several experimental schemes are established to achieve the best performance. In the first scheme, skin cancer detection is conducted using the original data (without augmentation). In the second, third, and fourth schemes, the original data is augmented using Geometric augmentation, GAN augmentation, and Geometric+GAN augmentation. This study use rotation, shift, shear, zoom, flip, and brightness for geometric augmentation with detail parameter values shows in Table 1. During GAN augmentation, a total of 1,000 epochs were run with a batch size of 64. The parameter values of training model, such as optimizer, learning rate, and epoch are shown in Table 2. In this experiment, we also conducted trials with an custom FC layer configuration as shown in Table 3, consisting of a dense layer with 64 neurons, a dense layer with 32 neurons, and a dense layer with 7 neurons [17].

**Table 1.** Summary of geometric augmentation parameters

| Parameter | Value |
|---|---|
| rotation_range | 20 |
| width_shift_range | 0.2 |
| height_shift_range | 0.2 |
| shear_range | 0.2 |
| zoom_range | 0.2 |
| horizontal_flip | True |
| brightness_range | (0.8, 1.2) |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2023 doi:10.20944/preprints202311.0161.v1



**Table 2.** Parameter of training model

| Parameter | Value |
|---|---|
| Optimizer | Adam |
| Learning rate | 0.0001 |
| Optimizer parameters | beta_1=0.9, beta_2=0.999 |
| Epoch | 100 (with early stopping) |

**Table 3.** Summary of custom FC layer

| Layer | Output Shape | Activation |
|---|---|---|
| Dense | (None, 64) | Relu |
| Dense | (None, 32) | Relu |
| Dense | (None, 7) | Softmax |

This study also uses SHAP to explain the skin cancer detection, a technique or approach that utilizes the concept of Shapley values to explain the contribution of each pixel or feature in an image to the model's predictions. CNN are frequently referred to as black boxes due to the difficulty in deciphering their decision-making processes. For the purpose of understanding model behavior and building trust, SHAP assist in improving the transparency and interpretability of CNN decision-making. In SHAP, the concept of Shapley values is applied to measure and understand the influence of each pixel in the image on the model's output or prediction. This technique is valuable for interpreting machine learning models, including convolutional neural network (CNN) models frequently used for image-based tasks. Positive SHAP values signify that the presence of a pixel had a positive impact on the prediction (red pixel), whereas negative values indicate the contrary (blue pixel) [18].

### *3.4. Performance Metrics*

The evaluation of performance was conducted using seven metrics: accuracy (Acc), precision (Prec), recall (Rec), F1-score, SpecificityAtSensitivity, SensitivityAtSpecificity, and G-mean. Accuracy assesses the proportion of true positives and true negatives among all the images. Precision measures the ratio of true positives to all elements identified as positives, which includes false positives. Recall, also known as sensitivity, measures the ratio of true positives to all relevant elements, i.e., the true positives in the dataset. Specificity is a metric that assesses the model's capability to accurately recognize instances that are actually not part of the positive class in a classification scenario. F1-score represents the harmonic mean of recall and precision, providing an indication of classification accuracy in imbalanced datasets. Equations 1-6 define these seven metrics. G-mean is short for Geometric Mean, is used to evaluate the performance of classification models, particularly in situations where imbalanced datasets exist.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$G - mean = \sqrt{sensitivity \times specificity} \tag{6}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2023 doi:10.20944/preprints202311.0161.v1



## 4. Results and Discussion

Before the prediction process, data augmentation is performed on the training, validation, and testing data in the HAM10000 dataset using geometric augmentation, GAN augmentation, and geometric+GAN augmentation. The limited number of images in the skin cancer class will be augmented to bring it closer to the number of images in the class with the highest number of images (*nv* class). The number of images in each class before and after augmentation is shown in Table 4.

**Table 4.** Distribution of each skin cancer category of each augmentation schemes.

| | Original | | | Geometric Aug. | | | GAN | | | Geometric Aug.+GAN | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Category** | **Train** | **Test** | **Val** | **Train** | **Test** | **Val** | **Train** | **Test** | **Val** | **Train** | **Test** | **Val** |
| **vasc** | 110 | 26 | 6 | 4,801 | 1,350 | 554 | 4,843 | 1,359 | 503 | 4,805 | 1,371 | 529 |
| **nv** | 4,822 | 1,347 | 536 | 4,826 | 1,316 | 563 | 4,854 | 1,302 | 549 | 4,856 | 1,300 | 549 |
| **mel** | 792 | 222 | 99 | 4,877 | 1,303 | 525 | 4,858 | 1,319 | 528 | 4,836 | 1,361 | 508 |
| **df** | 83 | 25 | 7 | 4,775 | 1,423 | 507 | 4,831 | 1,325 | 549 | 4,831 | 1,340 | 534 |
| **bkl** | 785 | 224 | 90 | 4,887 | 1,316 | 502 | 4,813 | 1,329 | 563 | 4,831 | 1,337 | 537 |
| **bcc** | 370 | 101 | 43 | 4,783 | 1,360 | 562 | 4,792 | 1,387 | 526 | 4,798 | 1,394 | 513 |
| **akiec** | 248 | 58 | 21 | 4,844 | 1,319 | 542 | 4,802 | 1,366 | 537 | 4,836 | 1,284 | 585 |
| **Num. images** | 7,210 | 2,003 | 802 | 33,793 | 9,387 | 3,755 | 33,793 | 9,387 | 3,755 | 33,793 | 9,387 | 3,755 |
| **Total images** | 10,015 | | | 46,935 | | | 46,935 | | | 46,935 | | |

Table 5 shows the performance comparison of several pre-trained models with the proposed augmentation method. Using the original data, Resnet152v2 performed the best based on accuracy (84.12%), precision (84.77%), recall (83.67%) and F1 score (84.22%). However, when considering sensitivity, specificity, and G-mean, EfficientnetB7 achieved the best metric values with 99.49%, 94.91%, and 97.17%, respectively. Through the augmentation scheme we proposed, the accuracy of skin cancer detection can be enhanced, reaching a range of 96% to 97.95%. Overall, Geometric augmentation produced the best performance based on accuracy, precision, and F1 score metrics, while Geometric+GAN yielded the best metrics in terms of sensitivity, specificity, and G-mean values. SensitivityAtSpecificity, SpecificityAtSensitivity, and G-mean all approach 100% when employing geometric+GAN on a tested pre-trained model. In Table 6, it is evident that by modifying the FC layer, there is an increase in accuracy up to 98.07% when using EfficientnetB7 and geometric augmentation.

**Table 5.** Performance of the proposed augmentation method on several pre-trained models.

| Augmentation Method | Pre-trained Model | Acc | Prec | Rec | F1 | Sensitivity AtSpecificity | Specificity AtSensitivity | G-mean | Epoch |
|---|---|---|---|---|---|---|---|---|---|
| Data Original | Xception | 79.93 | 80.70 | 79.53 | 80.11 | 99.16 | 91.71 | 95.36 | 12 |
| | Inceptionv3 | 78.88 | 79.51 | 78.48 | 78.99 | 99.31 | 92.31 | 95.75 | 11 |
| | Resnet152v2 | 84.12 | 84.77 | 83.67 | 84.22 | 99.33 | 93.56 | 96.40 | 18 |
| | EfficientnetB7 | 78.03 | 79.63 | 77.28 | 78.44 | 99.49 | 94.91 | 97.17 | 11 |
| | InceptionresnetV2 | 79.63 | 80.20 | 78.68 | 79.44 | 99.28 | 91.76 | 95.45 | 19 |
| | VGG19 | 81.73 | 81.83 | 81.63 | 81.73 | 99.12 | 91.26 | 95.11 | 28 |
| Geometric | Xception | 97.05 | 97.06 | 97.01 | 97.03 | 99.87 | 99.12 | 99.49 | 19 |
| | Inceptionv3 | 97.38 | 97.48 | 97.35 | 97.41 | 99.90 | 99.20 | 99.55 | 31 |
| | Resnet152v2 | 96.90 | 96.95 | 96.86 | 96.90 | 99.85 | 98.93 | 99.39 | 28 |
| | EfficientnetB7 | 97.95 | 98.00 | 97.90 | 97.95 | 99.91 | 99.41 | 99.66 | 19 |
| | InceptionresnetV2 | 97.40 | 97.46 | 97.36 | 97.41 | 99.89 | 99.20 | 99.55 | 28 |
| | VGG19 | 97.22 | 97.24 | 97.20 | 97.22 | 99.83 | 98.84 | 99.33 | 32 |
| GAN | Xception | 96.08 | 96.35 | 95.96 | 96.16 | 99.86 | 98.70 | 99.28 | 10 |
| | Inceptionv3 | 96.50 | 96.62 | 96.45 | 96.53 | 99.86 | 98.64 | 99.25 | 16 |
| | Resnet152v2 | 96.30 | 96.47 | 96.23 | 96.35 | 99.79 | 98.37 | 99.08 | 20 |
| | EfficientnetB7 | 96.48 | 96.59 | 96.44 | 96.51 | 99.79 | 98.25 | 99.02 | 20 |
| | InceptionresnetV2 | 96.22 | 96.32 | 96.20 | 96.26 | 99.82 | 98.44 | 99.13 | 18 |
| | VGG19 | 96.22 | 96.26 | 96.20 | 96.23 | 99.70 | 100.00 | 99.85 | 38 |
| Geometric + GAN | Xception | 96.21 | 96.51 | 96.04 | 96.27 | 99.94 | 99.22 | 99.58 | 9 |
| | Inceptionv3 | 96.45 | 96.56 | 96.39 | 96.48 | 99.86 | 98.56 | 99.21 | 20 |
| | Resnet152v2 | 96.59 | 96.75 | 96.45 | 96.60 | 99.86 | 98.87 | 99.36 | 14 |
| | EfficientnetB7 | 96.50 | 96.61 | 96.43 | 96.52 | 99.89 | 98.92 | 99.40 | 14 |
| | InceptionresnetV2 | 96.71 | 96.82 | 96.67 | 96.74 | 99.85 | 98.62 | 99.23 | 21 |
| | VGG19 | 95.39 | 97.36 | 93.89 | 95.59 | 100.00 | 99.93 | 99.96 | 17 |

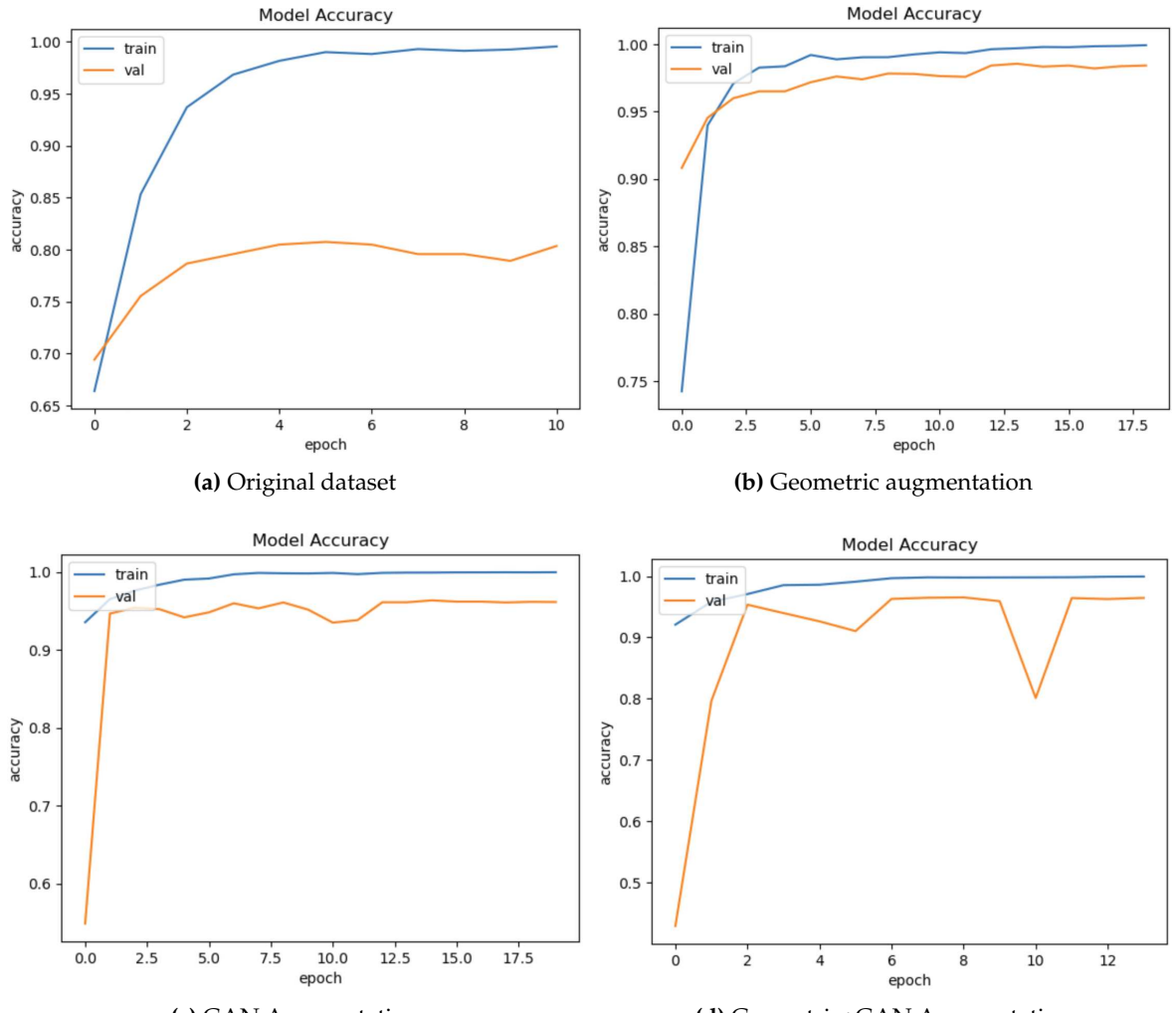
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2023 doi:10.20944/preprints202311.0161.v1



**Table 6.** Performance of the proposed augmentation method on the custom FC layer (three dense layers with 64 neurons, 32 neurons, and 7 neurons).

| Augmentation Method | Pre-trained Model | Acc | Prec | Rec | F1 | Sensitivity AtSpecificity | Specificity AtSensitivity | G-mean | Epoch |
|---|---|---|---|---|---|---|---|---|---|
| Geometric | EfficientnetB7 | 98.07 | 98.10 | 98.06 | 98.08 | 99.92 | 99.46 | 99.69 | 20 |
| GAN | Inceptionv3 | 96.48 | 96.63 | 96.44 | 96.53 | 99.83 | 98.54 | 99.18 | 17 |
| Geometric + GAN | InceptionresnetV2 | 96.90 | 97.07 | 96.87 | 96.97 | 99.86 | 98.90 | 99.38 | 22 |

Figure 5 shows sample accuracy results from the training and validation of EfficientnetB7 on the original dataset and the proposed augmentation. The training and validation accuracies appear to overfitting on the original dataset (Figure 4a). Validation accuracy is improved by geometric augmentation (Figure 4b), thereby reducing overfitting. Training accuracy performance is enhanced through the use of GAN and geometric+GAN (Figure 4c and 4d).




**(a)** Original dataset

**(b)** Geometric augmentation

**(c)** GAN Augmentation

**(d)** Geometric+GAN Augmentation

**Figure 4.** The samples of training and validation accuracy on EfficientnetB7

The sample confusion matrix generated from the original dataset and the best model are shown in Figures 5 and 6, respectively. Both of these confusion matrices were generated using EfficientnetB7. Skin cancer images in the *df* and *vasc* classes can be accurately classified with no classification errors (Figure 6). Only 3 images out of 1319 images in the *akiec* class were misclassified as *bcc*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2023 doi:10.20944/preprints202311.0161.v1



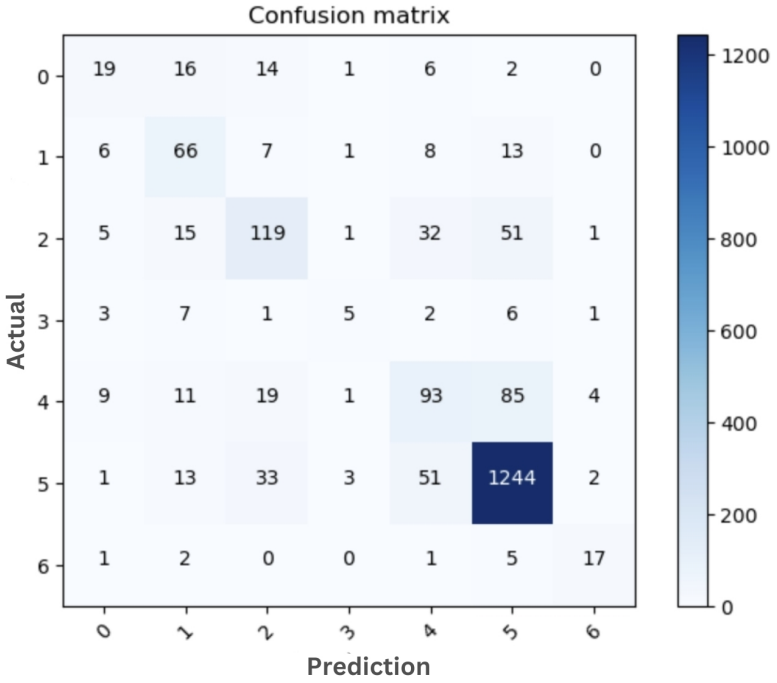



**Figure 5.** Confusion matrix of ori-EfficientnetB7 on original dataset: (0) *akiec*, (1) *bcc*, (2) *bkl*, (3) *df*, (4) *mel*, (5) *nv*, and (6) *vasc*.

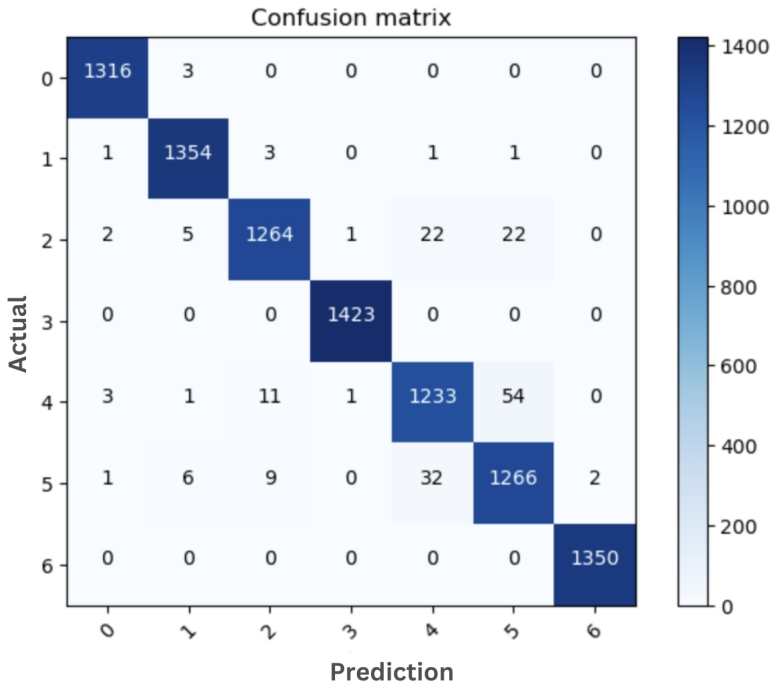



**Figure 6.** Confusion matrix of the best performance model (EfficientnetB7+Custom FC using Geometric augmentation): (0) *akiec*, (1) *bcc*, (2) *bkl*, (3) *df*, (4) *mel*, (5) *nv*, and (6) *vasc*.

We performed a comparative analysis to evaluate the performance of our model by comparing it to the outcomes of earlier research that utilized the use of the HAM10000 dataset, as shown in Table 7. Our proposed approach outperforms earlier findings in a number of metrics. Our limitation is mainly in terms of accuracy when compared to Gomathi et al. [4]. The accuracy rate still needs improvement as we plan to explore other deep learning architectures to enhance skin cancer detection. However, in terms of recall, precision, and F1, our approach outperforms the previous research. The standard deviations of accuracy, precision, and recall in our proposal also indicate low values, suggesting that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2023 doi:10.20944/preprints202311.0161.v1



our proposed approach demonstrates consistent performance across all three metrics. Figures 7 shows the SHAP explanation of *akiec*, *bcc*, *bkl*, *df*, *mel*, *nv*, and *vasc* samples, respectively. The explanations are displayed in a clear gray background, with the testing images on the left.

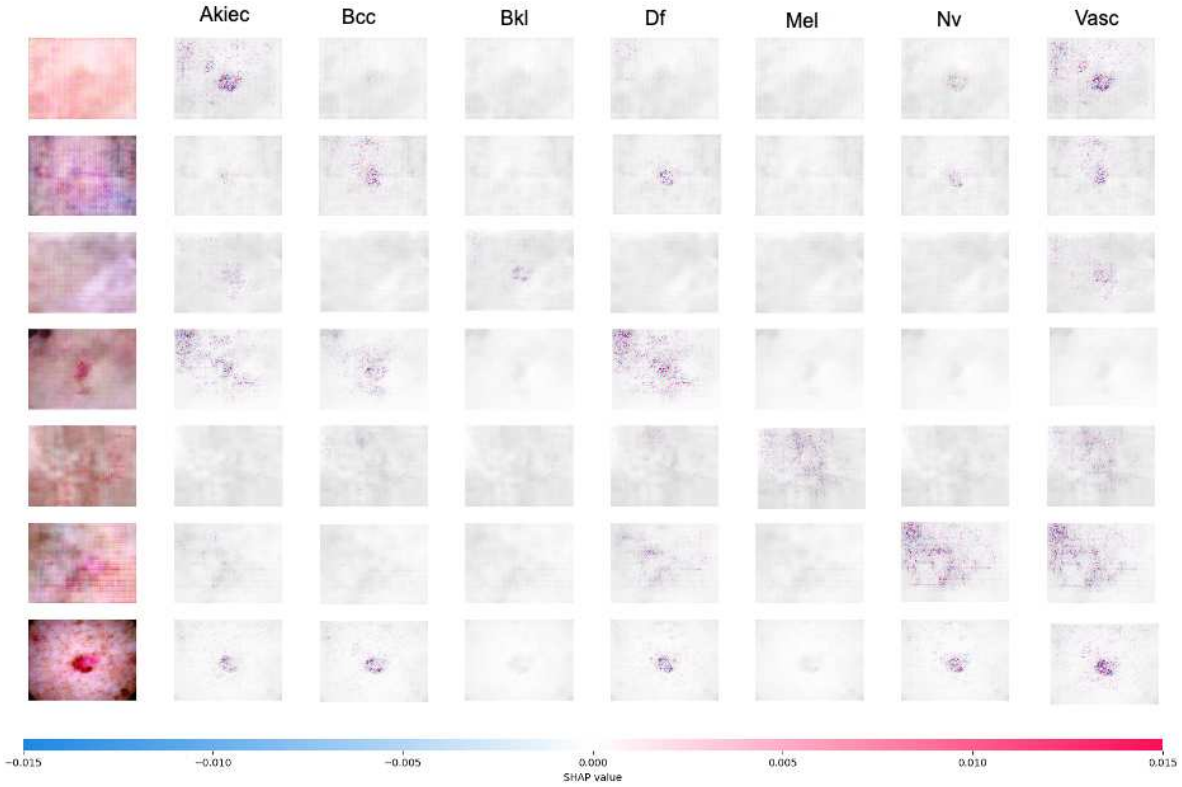



**Figure 7.** The result of SHAP explanation on InceptionResnetV2 using Geometric+GAN augmentation. The sample image is correcly classified as *akiec*, *bcc*, *bkl*, *df*, *mel*, *nv*, and *vasc*, since the high concentration of red pixels is located in the second, third, fourth, fifth, sixth, seventh, and eighth explanation column images, respectively.

**Table 7.** Performance comparison with the state-of-the art models.

| Ref | Method | Acc | Prec | Rec | F1 | Stdev |
|---|---|---|---|---|---|---|
| Alam et al. [13] | AlexNet, InceptionV3, and RegNetY-320 | 91 | - | - | 88.1 | - |
| Kalpana et al. [2] | ESVMKRF-HEAO | 97.4 | 96.3 | 95.9 | 97.4 | 0.7767 |
| Shan et al. [19] | AttDenseNet-121 | 98 | 91.8 | 85.4 | 85.6 | 6.3003 |
| Gomathi et al. [4] | DODL net | 98.76 | 96.02 | 95.37 | 94.32 | 1.7992 |
| Alwakid et al. [20] | InceptionResnet-V2 | 91.26 | 91 | 91 | 91 | 0.1501 |
| Sae-Lim et al. [14] | Modified MobileNet | 83.23 | - | 85 | 82 | |
| Ameri [21] | AlexNet | 84 | - | - | - | |
| Chaturvedi et al. [6] | ResNeXt101 | 93.2 | 88 | 88 | - | 3.0022 |
| Shahin Ali et al. [22] | DCNN | 91.43 | 96.57 | 93.66 | 95.09 | 2.5775 |
| Sevli et al. [23] | Custom CNN architecture | 91.51 | - | - | - | |
| Fraiwan et al. [24] | DenseNet201 | 82.9 | 78.5 | 73.6 | 74.4 | 4.6522 |
| Balambigai et al. [25] | Grid search ensemble | 77.17 | - | - | - | |
| Shaheen et al. [26] | PSOCNN | 97.82 | - | - | 98 | |
| This study | Geometric+EfficientnetB7+Custom FC | 98.07 | 98.10 | 98.06 | 98.08 | 0.0002 |
| | GAN+InceptionV3 | 96.50 | 96.62 | 96.45 | 96.53 | 0.0009 |
| | Geometric+GAN+InceptionresnetV2+Custom FC | 96.90 | 97.07 | 96.87 | 96.97 | 0.0011 |

## 5. Conclusions

In conclusion, this study provides valuable insights into a deep learning approach to early detection of skin cancer using image augmentation techniques. The proposed two-stage image augmentation technique achieved a high performance, in term of accuracy, precision, recall, and

F1-score, which is a promising result for automated skin cancer detection. The other metrics, such as sensitivity, specificity, and G-mean of the proposed augmentation method also achieve the best performance compare to the original dataset. The use of an interpretable technique for skin cancer diagnosis is also a significant contribution to the field, as it can help clinicians understand the reasoning behind the diagnosis and improve trust in the system. Overall, this research paper presents a promising approach to automated skin cancer detection that could have a significant impact on patient outcomes and healthcare costs.

**Author Contributions:** Conceptualization, Catur Supriyanto; Data curation, Abu Salam and Junta Zeniarja; Formal analysis, Adi Wijaya; Investigation, Catur Supriyanto and Adi Wijaya; Methodology, Catur Supriyanto; Resources, Abu Salam; Software, Abu Salam and Junta Zeniarja; Supervision, Adi Wijaya; Validation, Catur Supriyanto and Adi Wijaya; Writing – original draft, Catur Supriyanto; Writing – review & editing, Catur Supriyanto.

**Funding:** This research was funded by DRTPM-DIKTI for research funding in 2023.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bozkurt, F. Skin lesion classification on dermatoscopic images using effective data augmentation and pre-trained deep learning approach. *Multimedia Tools and Applications* **2023**, *82*, 18985–19003. doi:10.1007/s11042-022-14095-1.
2. Kalpana, B.; Reshmy, A.; Senthil Pandi, S.; Dhanasekaran, S. OESV-KRF: Optimal ensemble support vector kernel random forest based early detection and classification of skin diseases. *Biomedical Signal Processing and Control* **2023**, *85*, 104779. doi:https://doi.org/10.1016/j.bspc.2023.104779.
3. Girdhar, N.; Sinha, A.; Gupta, S. DenseNet-II: an improved deep convolutional neural network for melanoma cancer detection. *Soft Computing* **2023**, *27*, 13285–13304. doi:10.1007/s00500-022-07406-z.
4. Gomathi, E.; Jayasheela, M.; Thamarai, M.; Geetha, M. Skin cancer detection using dual optimization based deep learning network. *Biomedical Signal Processing and Control* **2023**, *84*, 104968. doi:10.1016/j.bspc.2023.104968.
5. Cassidy, B.; Kendrick, C.; Brodzicki, A.; Jaworek-Korjakowska, J.; Yap, M.H. Analysis of the ISIC image datasets: Usage, benchmarks and recommendations. *Medical Image Analysis* **2022**, *75*, 102305. doi:https://doi.org/10.1016/j.media.2021.102305.
6. Chaturvedi, S.S.; Tembhurne, J.V.; Diwan, T. A multi-class skin Cancer classification using deep convolutional neural networks. *Multimedia Tools and Applications* **2020**, *79*, 28477–28498. doi:10.1007/s11042-020-09388-2.
7. Alsahafi, Y.S.; Kassem, M.A.; Hosny, K.M. Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier. *Journal of Big Data* **2023**, *10*, 105. doi:10.1186/s40537-023-00769-6.
8. Mumuni, A.; Mumuni, F. Data augmentation: A comprehensive survey of modern approaches. *Array* **2022**, *16*, 100258. doi:10.1016/j.array.2022.100258.
9. Zhang, Y.; Wang, Z.; Zhang, Z.; Liu, J.; Feng, Y.; Wee, L.; Dekker, A.; Chen, Q.; Traverso, A. GAN-based one dimensional medical data augmentation. *Soft Computing* **2023**, *27*, 10481–10491. doi:10.1007/s00500-023-08345-z.
10. Abayomi-Alli, O.O.; Damaševičius, R.; Misra, S.; Maskeliūnas, R.; Abayomi-Alli, A. Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES* **2021**, *29*, 2600–2614. doi:10.3906/elk-2101-133.
11. Chang, C.C.; Li, Y.Z.; Wu, H.C.; Tseng, M.H. Melanoma Detection Using XGB Classifier Combined with Feature Extraction and K-Means SMOTE Techniques. *Diagnostics* **2022**, *12*, 1747. doi:10.3390/diagnostics12071747.

12. Tahir, M.; Naeem, A.; Malik, H.; Tanveer, J.; Naqvi, R.A.; Lee, S.W. DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images. *Cancers* **2023**, *15*, 2179. doi:10.3390/cancers15072179.
13. Alam, T.M.; Shaukat, K.; Khan, W.A.; Hameed, I.A.; Almuqren, L.A.; Raza, M.A.; Aslam, M.; Luo, S. An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset. *Diagnostics* **2022**, *12*. doi:10.3390/diagnostics12092115.
14. Sae-Lim, W.; Wettayaprasit, W.; Aiyarak, P. Convolutional Neural Networks Using MobileNet for Skin Lesion Classification. 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2019, pp. 242–247. doi:10.1109/JCSSE.2019.8864155.
15. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. Advances in Neural Information Processing Systems. Curran Associates, Inc., 2014, Vol. 27.
16. Shamsolmoali, P.; Zareapoor, M.; Shen, L.; Sadka, A.H.; Yang, J. Imbalanced data learning by minority class augmentation using capsule adversarial networks. *Neurocomputing* **2021**, *459*, 481–493. doi:10.1016/j.neucom.2020.01.119.
17. Shahin, M.; Chen, F.F.; Hosseinzadeh, A.; Khodadadi Koodiani, H.; Shahin, A.; Ali Nafi, O. A smartphone-based application for an early skin disease prognosis: Towards a lean healthcare system via computer-based vision. *Advanced Engineering Informatics* **2023**, *57*, 102036. doi:10.1016/j.aei.2023.102036.
18. Bhandari, M.; Shahi, T.B.; Neupane, A. Evaluating Retinal Disease Diagnosis with an Interpretable Lightweight CNN Model Resistant to Adversarial Attacks. *Journal of Imaging* **2023**, *9*. doi:10.3390/jimaging9100219.
19. Shan, P.; Chen, J.; Fu, C.; Cao, L.; Tie, M.; Sham, C.W. Automatic skin lesion classification using a novel densely connected convolutional network integrated with an attention module. *Journal of Ambient Intelligence and Humanized Computing* **2023**, *14*, 8943–8956. doi:10.1007/s12652-022-04400-z.
20. Alwakid, G.; Gouda, W.; Humayun, M.; Jhanjhi, N.Z. Diagnosing Melanomas in Dermoscopy Images Using Deep Learning. *Diagnostics* **2023**, *13*, 1815. doi:10.3390/diagnostics13101815.
21. Ameri, A. A Deep Learning Approach to Skin Cancer Detection in Dermoscopy Images. *J Biomed Phys Eng.* **2020**, *10*, 801–806. doi:10.31661/jbpe.v0i0.2004-1107.
22. Ali, M.S.; Miah, M.S.; Haque, J.; Rahman, M.M.; Islam, M.K. An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Machine Learning with Applications* **2021**, *5*, 100036. doi:https://doi.org/10.1016/j.mlwa.2021.100036.
23. Sevli, O. A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation. *Neural Computing and Applications* **2021**, *33*, 12039–12050. doi:10.1007/s00521-021-05929-4.
24. Fraiwan, M.; Faouri, E. On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning. *Sensors* **2022**, *22*. doi:10.3390/s22134963.
25. Balambigai, S.; Elavarasi, K.; Abarna, M.; Abinaya, R.; Vignesh, N.A. Detection and optimization of skin cancer using deep learning. *Journal of Physics: Conference Series* **2022**, *2318*, 012040. doi:10.1088/1742-6596/2318/1/012040.
26. Shaheen, H.; Singhn, M.P. Multiclass skin cancer classification using particle swarm optimization and convolutional neural network with information security. *Journal of Electronic Imaging* **2022**, *32*. doi:10.1117/1.JEI.32.4.042102.