Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-17067-1

# Tackling the class imbalanced dermoscopic image classification using data augmentation and GAN

**Mostapha Alsaidi**[1] · **Muhammad Tanveer Jan**[1] · **Ahmed Altaher**[1] · **Hanqi Zhuang**[1] · **Xingquan Zhu**[1]

Received: 15 March 2023 / Revised: 21 July 2023 / Accepted: 14 September 2023


**Abstract**
Dermoscopy is a noninvasive way to examine and diagnose skin lesions, *e.g.* nevus and melanoma, and is a critical step for skin cancer detection. Accurate classification of dermoscopic images can detect skin cancer at an early stage and bring social and economic impact to patients and communities. Using deep learning methods to classify dermoscopic images has shown superior performance, but existing research often overlooks the class imbalance in the data. In addition, although a handful of public datasets are available for skin cancer research, these datasets are generally not large enough for deep learning algorithms to produce accurate results. In this paper, we propose to use data augmentation and generative adversarial networks (GAN) to tackle class-imbalanced dermoscopic image classification. Our main objectives are to determine (1) how state-of-the-art fine-tuned deep learning models perform on class-imbalanced dermoscopic images, (2) whether data augmentation and GAN can help alleviate class imbalances to improve classification accuracy, and (3) which method is more effective in addressing the class imbalance. By using public datasets and a carefully designed framework to generate augmented images and synthetic images, our research provides clear answers to these questions. Code and data used in the study are available at: https://github.com/mjan2021/Dermoscopic-image-classification.git

**Keywords** Dermoscopic image classification · Class imbalance · Data augmentation · Generative adversarial networks · Deep learning · Transfer learning

Mostapha Alsaidi and Muhammad Tanveer Jan are equal contributors.

✉ Muhammad Tanveer Jan
mjan2021@fau.edu

[1] Department. of Electrical Engineering & Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

## 1 Introduction

Skin cancer is a common and potentially deadly disease affecting millions worldwide. Melanoma is the most frequent type of skin cancer involving an uncontrolled proliferation of pigmented cells. A potentially fatal condition, it is more common in people with lighter skin tones and can spread swiftly throughout the body [1]. Consequently, melanoma patients have a better chance of survival if their disease is detected and treated early.

A skin examination requires the expertise of medical specialists and access to diagnostic equipment [1]. Dermoscopy is a non-invasive diagnostic method that magnifies skin lesions, specially pigmented and non-pigmented ones. A dermoscope magnifies and illuminates the skin lesion using polarized light. This lets the dermatologist evaluate the lesion's features and detect characteristic patterns and structures to help diagnose skin disorders. Melanoma, basal cell carcinoma, and squamous cell carcinoma are diagnosed with dermoscopy. It can detect moles, warts, and eczema. The dermatologist can better diagnose a lesion by studying its blood vessels, pigmentation patterns, and other morphological traits. This can lead to early and more successful therapy [33]. Dermatologists examine dermoscopy images for specific characteristics, i.e., asymmetrical patterns, erratic borders, lots of colors, and different pigmentations. They may also search for "dots and globules" to suggest blood vessels in the skin and "streamers" to show malignant cell spread. Scaling, crusting, and thickening may also indicate skin cancer. Dermatologists can diagnose skin cancer by examining these photos [34]. Manual diagnosis could be incredibly time-consuming and costly. Therefore, it is essential to develop methods to reliably differentiate between malignant skin cancers like Melanoma and less severe conditions like Melanocytic Nevi, Benign Keratosis, Dermatofibroma, Actinic Keratosis, and vascular lesions.

Deep learning has recently revolutionized the medical field as well as many other domains. In light of this, numerous studies have been conducted with Convolutional Neural Network (CNN) being primarily used as a base model to detect and classify skin diseases [58]. CNNs represent the backbone of visionbased deep learning models and are particularly effective for image recognition tasks. Other approaches also employ numerous deep learning models to improve model performance and address data imbalance by using data augmentation techniques such as rotating, flipping, or resizing images [45].

Dermoscopy image processing frequently suffers from data imbalance, especially when it comes to determining whether a skin lesion is benign or malignant. The dataset is skewed since benign lesions are far more common than malignant ones. Any data-driven algorithm used to categorize skin lesions would suffer a bias toward benign classification caused by the skewed dataset used for training. In turn, this could mean that the algorithm is less reliable in identifying malignant tumors, which could result in failure to detect cancer at an early stage. Oversampling the minority class (malignant lesions) or undersampling the majority class (benign lesions) and applying methods like costsensitive learning or class weighting are all viable options for resolving data imbalance in dermoscopic pictures. It's possible that the best outcomes could be achieved by employing a combination of these methods [35]. Another way to improve model performance is to address the imbalanced dataset problem directly. A CNN model 1 showed tremendous improvement in classification accuracy due to implementing data augmentation techniques such as rotating, flipping, or resizing images. This method alleviates class imbalance by using augmented images to populate the dataset, thus reducing class imbalance.

In this paper, we study the improvements that could be made to deep learning model performance on imbalanced datasets. For this purpose, four state-of-the-art deep-learning

models were trained to detect Melanoma and other skin abnormalities. We utilize data augmentation techniques following the work presented in [1], showing a huge improvement in classification. Furthermore, we investigate the efficacy of training a deep learning model on synthesized data. New data was generated using a generative adversarial network model (AC-GAN) that was trained to generate dermoscopic images in order to populate the dataset and alleviate class imbalance.

## 2 Related work

### 2.1 Deep learning for dermoscopic image classification

Many medical professionals have benefited from the extensive study of images for diagnostics and detection of cancer by analyzing skin lesions, especially for breast and skin cancers. Image classification is an area where deep learning techniques have been shown to excel [1]. There has been a great amount of work done on Melanoma skin cancer detection over the years. However, many researchers struggle with the lack of accessible, high-quality data sets. Healthcare institutions have been reluctant to make their data publicly available for cancer research for reasons related to patients' right to privacy. These obstacles notwithstanding, advancements in cancer research have yielded significant results that have helped tremendously medical practitioners in combating cancers. In this section, we provide an overview of the published research on the topic of Melanoma skin cancer diagnosis using machine learning techniques. Machine learning algorithms can analyze images of skin lesions, and the ones that look suspicious can be flagged as probable cases of skin cancer. Using photos labeled as Melanoma or Non-Melanoma, researchers [1, 20, 22] investigated the efficacy of using (CNNs) to detect skin cancer. On the topic of skin lesion classification, CNNs have been found to perform better than more conventional machine learning techniques like support vector machines (SVMs) and decision tree classifiers. Unfortunately, other types of skin diseases are ignored by this method of binary categorization [23].

Deep learning models are now commonly used in different fields [74, 75, 77]. In previous studies, researchers employed VGG16, VGG19, and InceptionV3, and obtained accuracy of 77%, 76% and 74%, respectively [76]. Many industrial professionals are also reviewing such studies, pointing out the gaps between the research and industrial usage of those studies. [78–80]. Another deep learning model-based study used for the classification of skin lesions was conducted on the HAM10000 dataset by employing the pre-trained ResNeXt101 and the ensemble IneptionResNetV2 With ResNeXt101. They were able to obtain an accuracy of 93.20% and 92.83% respectively [17].

In addition to CNNs, other types of machine-learning algorithms have also been explored for skin lesion classification. For example, some studies have used random forests, a type of ensemble learning algorithm that combines the predictions of multiple decision trees to improve the model's overall accuracy. Other studies [14, 15, 24, 66] have explored the use of hybrid algorithms that combine elements of CNNs with other types of machine learning algorithms. For example, some studies have used a combination of CNNs and support vector machines (SVMs) to improve the performance of skin lesion classification models.

Some of the studies used a two-stage approach by using segmentation to extract features of the skin lesion and then using classification to identify the type of lesion [59]. A survey

conducted on segmentation models for deep learning divides the research into several types, i.e., single Network models, multi-model networks, transformer models, and hybrid feature models [60]. Most of the single network models are either fully connected [61, 62] or Ushaped [63, 64]. Multi-model networks are divided into standard ensembles [65, 67], multi-task model [68, 69] and GANs [70, 71]. Transformers initially introduced for natural language processing have performed pretty well in other areas, especially in computer vision. TransNet was one of the initial works that used transformers alongside CNNs for medical images [72], which was further proven by studies [73]

This approach, however, is not limited to only skin lesion classification but can also be used for detecting COVID-19 [15]. while another study is being conducted to analyze the driving pattern of older drivers using such algorithms to detect early-stage dementia [14, 36]. They are also quite accurate in recognizing emotions [53] as well in detecting levels of diabetics [54]

### 2.2 Sample bias in skin lesion detection

One challenge in using machine learning algorithms for skin lesion classification is the limited availability of high-quality, annotated datasets. The HAM10000 dataset [13] is a commonly used dataset for research on skin lesion classification, but it is relatively small and may not be representative of the full range of skin types and conditions. As a result, many studies [26–28] have focused on developing methods to improve the generalization capability of machine learning models trained on this dataset.

Skin cancer disproportionately affects certain populations, such as people with lighter skin tones, and this can lead to imbalances in the training data. As a result, one issue identified in using machine learning algorithms for skin lesion classification is the potential for bias in the training data due to the limited number of samples for certain labels. These imbalances can cause machine learning models to be biased towards the majority class and may lead to lower accuracy on the minority class. To address this issue, some studies have used techniques such as oversampling or undersampling to balance the training data and improve the performance of the machine learning models. Some studies were conducted in which novel approaches like multi-weight new loss and endto-end learning strategies were introduced to deal with the bias of the data [28] while another study introduced supervised contrastive loss and focal loss to deal with the issue of bias in the dataset [37]. Also, by deep-clustering, the cluster separation in embedding space improves the metrics in detection when there is bias in the data [38]. Attention-based models can be very useful when dealing with bias, and they are, as of today, in high focus. Using Grouping of multi-scale attention blocks helps extracts feature on a fine-grained level of the lesions, which leads to improved performance [39].

While using novel approaches and state-of-the-art methods increases the performance and accuracy of the lesion detection areas, a typical approach like data augmentation should also be kept in mind as it could significantly affect the metrics. Such methods are considered a low-cost plug-and-play approach that can be used to increase performance and accuracy by choosing the best arguments among the list of augmentations for skin lesion classification [40].

In a previous study [41], a deeply discriminated GAN (DDGAN) is used to synthesize high-resolution images while keeping the image's features as much realistic as possible. DDGAN employs several discriminators, which help the generator recreate much more accurate synthetic images [41]. Alongside DDGAN, skin lesion style-based GAN is used that is

based on the same architecture of the style-based GAN, which can synthesize high resolution and feature enriched skin lesion images [42].

In summary, although GAN and data augmentation have been previously studied for dermoscopic image classification, there is no research investigation on the performance comparisons of GAN *vs.* data augmentation, especially in the biomedical domain. A previous study [55] has compared GAN-based generative modeling for dermatological applications and suggested that "the results archived in different scenarios do not differ much", meaning that GANgenerated dermatological images have limited contribution to learning. Our research systematically studies synthetic samples generated from GAN *vs.* from data augmentation and investigates which approach is likely more useful for dermoscopic image classification in tackling class imbalances.

### 2.3 Interpretable skin lesion detection

The interpretability of deep learning models has been a challenge for AI-driven applications. The complexity of the internal components of a deep learning model made it impossible to interpret its decision. Thus, deep learning models have been described as a black box for years.

This lack of transparency between the input and output of the model made the decision-making process vague and uninterpretable, thus increasing the barrier of entry for deep learning-based applications in various fields, such as the medical field. Furthermore, under the General Data Protection Regulations (GDPR) in Europe [49], individuals are entitled to receive an explanation regarding the decisions made by computer algorithms, including those generated by deep learning models.

To tackle this problem of deep learning, interpretability researchers have focused on developing tools and methods that would explain the decision-making process of a deep learning model. GRAD-CAM [50] stands for (Gradientweighted Class Activation Mapping), is a visualization tool that was developed by researchers at Georgia Institute of Technology. GRAD-CAM generates a heatmap highlighting the most important regions leading to the model's decision. A similar approach was introduced [51] where the researchers propose a method for interpretability. This method coined CAV stands for "Concept Activation Vector". It captures the relationship between the model's internal representations and a user-defined set of concepts related to skin lesions.

Another direction in the research community [31, 32] focused on integrating machine learning models that are more interpretable with deep learning models. This approach leverages the deep learning model's ability to extract features, thus using these models for feature extraction, not classification. Furthermore, machine learning models such as decision tree classifiers would be used for classification. These methods rely on the fact that rule-based systems can provide more transparent explanations of their predictions.

Models trained on sensitive data that significantly impacts a person's life need to be verified in sensitive professions like dermatology. One factor is not sufficient to make a decision, and therefore methods need to be verified by techniques that explain and supports the decision made by the model.

## 3 Methodology

Imbalanced data problem has been studied in the literature [6]. For example, HAM10000 [13] dataset used in our experiments is highly imbalanced with class "NV" consisting of 67% of the samples. This leads to biased classification towards the majority class by the model, which is a major problem in the research community. Solutions could be either at the data level where sampling techniques are adopted or at the model level where classifier decision is being weighted in proportion to class distributions [6].

In this research, we present an experimental scheme that allows the comparison, on the classification of skin lesions, between using GAN generated samples *vs.* using data augmentation methods by training different transfer learning models for classification.

### 3.1 Framework for imbalance dermoscopic image classification

HAM10000 dataset contains 10,015 images of size 600*450 stored in one directory and coupled with a CSV file that contains metadata on the image, including the class it belongs to. A custom dataset class was built using the PyTorch library to read the data and be integrated into PyTorch DataLoader throughout all the experiments.

Figure 1 shows the workflow of splitting and balancing the dataset. The original HAM10000 [13] dataset consists of 10 k images belonging to 7 classes. A portion of the HAM10000 dataset is kept aside for test purposes. It is split in a way that the ratio of the classes is kept the same hence the term stratified split. Test split consists of 20% of the original data and the rest 80% is filtered by selecting the minority classes because the majority class NV already consists of takes 67% of the images from the original dataset. The Number of minority class images is 3.3k, which are then used in both approached Augmentation and GAN to generate and balance the dataset. The Augmented images used mentioned techniques (Fig. 2), and to balance it with the NV class, 10×images are generated with each image using multiple techniques for augmentation. The total number of augmented images belonging to 6 classes is 36,981, which are then split into training and validation split with a fivefold cross-validation approach to train the models. The same approach is applied to the GAN model, which is trained on the minority class images belonging to 6

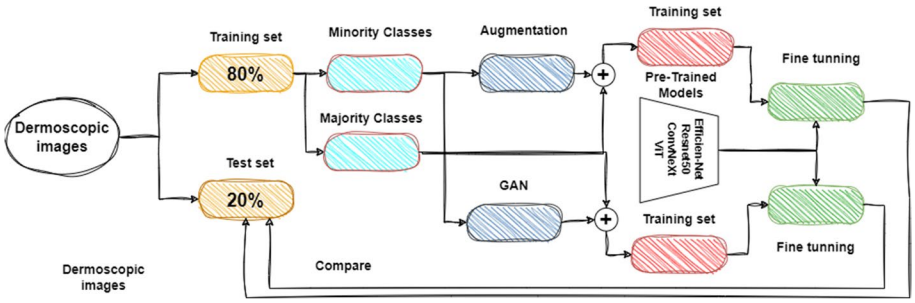


**Fig. 1** Proposed framework and data pipeline for comparative study of GAN *vs.* data augmentation for class imbalanced dermoscopic image classification. The dermoscopic images are split into training and test sets. Samples in the training set are used to train GAN and also generate augmentation samples for minor classes, respectively. The GAN-generated synthetic samples and data augmentation-generated synthetic samples are concatenated with majority class samples to fine-tune pre-trained models. The fine-tuned models are then compared on the same excluded test samples to study the algorithm performance comparatively

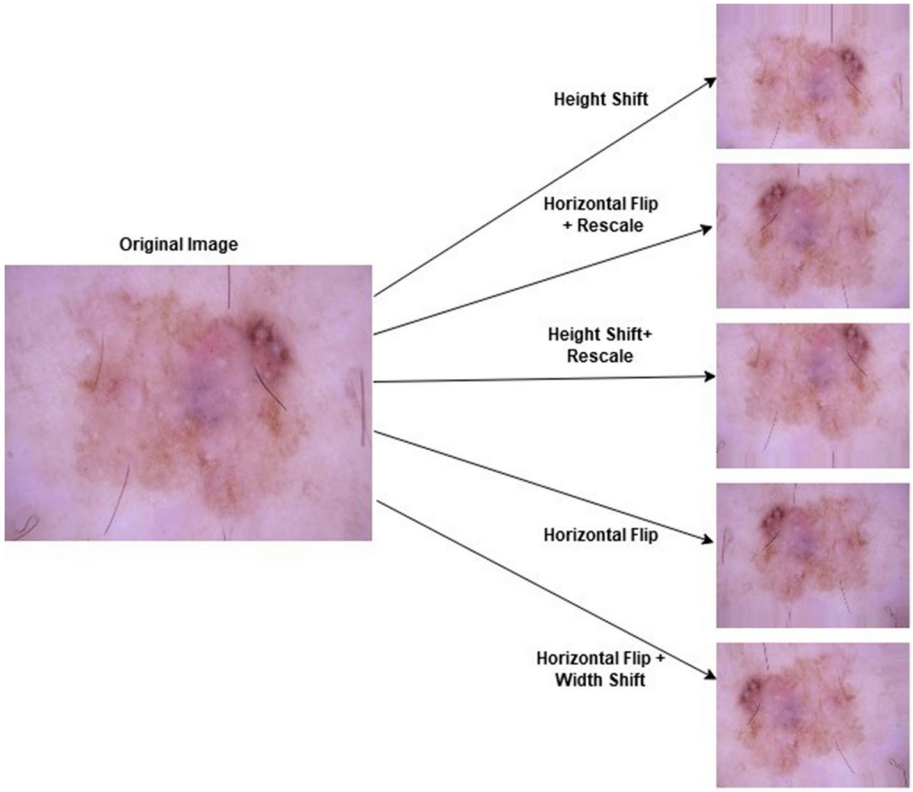


**Fig. 2** Examples of data augmentation using different operators (we use TensorFlow ImageDataGenerator in the experiments). The augmentation is carried out using single operators as well as a random combination of multiple operators

classes, and then it is used to generate an equal number of images keeping the ratio the same as the NV class. The total number of images generated by the GAN model is 36,926.

A few prepossessing steps were introduced to the dataset. Images were resized to 224*224 for all models. This reduces the computation load on the system and maintains proper resolution for the images. Furthermore, images were converted to tensors to be passed to the model. [2] used further prepossessing steps like removing hairlines from the image; however, we found this to be irrelevant to the performance of our models.

### 3.2 Data augmentation

Using deep learning for image classification frequently necessitates a substantial amount of training data for classification models, which is especially true for cancer detection. This method prevents the training model from becoming excessively data-specific [4]. Data augmentation is a technique used to extend the amount of a dataset for computer vision applications. This can be accomplished by transforming the existing data, for as by rotating, scaling, or cropping photographs. By supplying extra examples of the same object in various positions or situations, these changes improve the model's generalization to unseen

data. Data augmentation is especially beneficial when working with tiny datasets since it can boost the model's performance [52].

Vasconcelos et al. [5] provided a data augmentation strategy for the Melanoma dataset. This technique was utilized to rotate, flip, and crop the photos to maximize the likelihood of correctly detecting Melanoma and achieve a greater accuracy rate. Diallo et al. [4] applied various data augmentation techniques, including color modification, to improve the pictures. In addition to the data augmentation options, the stratification method is employed to prevent biased sampling. By establishing explicit criteria used to divide the data into the train and test sets, the stratification approach minimizes any risk of bias in the data-splitting operation. This prevents random data splitting and reduces bias in the data-splitting procedure.

Inconsistency within image datasets structures is another commonly encountered issue. For example, the image's size and form can vary between instances. When such inconsistencies exist during model training, the resulting model is ineffective [3].

### 3.3 Generative adversarial network

Generative Adversarial Network has witnessed wide adoption in the research community since its inception [30]. GANs are used to generate realistic synthetic images that are almost indistinguishable from real images, making GAN a desirable solution to increase training data size by generating new data samples. This process is more cost-effective than acquiring new data and annotating it. GAN models are trained to effectively generate artificial images that are as close to the real images as possible.

A GAN consists of two neural networks: a generator network and a discriminator network. Both networks are trained simultaneously in an adversarial manner. The generator network is trained to generate data samples that are similar to the data used in training, as defined in Eq. (1), while the discriminator network is trained to differentiate between the generated samples and the real data, as defined in Eq. (2) where $X$ is a genuine (*i.e.* $X_{\text{real}}$) or a fake (*i.e.* $X_{\text{fake}}$) image and $S=\{0,1\}$ define a fake or a genuine image, respectively as shown in (Fig. 3).

$$X_{fake} = G(z) \tag{1}$$

$$P(S|X) = D(X);\ S = \{0, 1\} \tag{2}$$

Both networks are trained in an adversarial manner where the generator is learning to generate samples that pass the discriminator without being detected as fake. This is enabled by maximizing the log-likelihood function defined in Eq. (3) until convergence.

$$\updownarrow() = E[logP(S = 1|X_{real})] + E[logP(S = 0|X_{fake})] \tag{3}$$

In this paper, we train Auxiliary Classifier GAN (AC-GAN) [7] architecture to generate synthetic samples to balance the minor classes. AC-GAN is an extension of the DCGAN [43], and a unique characteristic of AC-GAN, compared to GAN, is that it can generate synthetic data belonging to a specific class. This makes AC-GAN a useful model for tasks such as image classification, where the goal is to generate images belonging to a specific class. On the other hand, the DC-GAN is often used to generate realistic images without specific conditions.

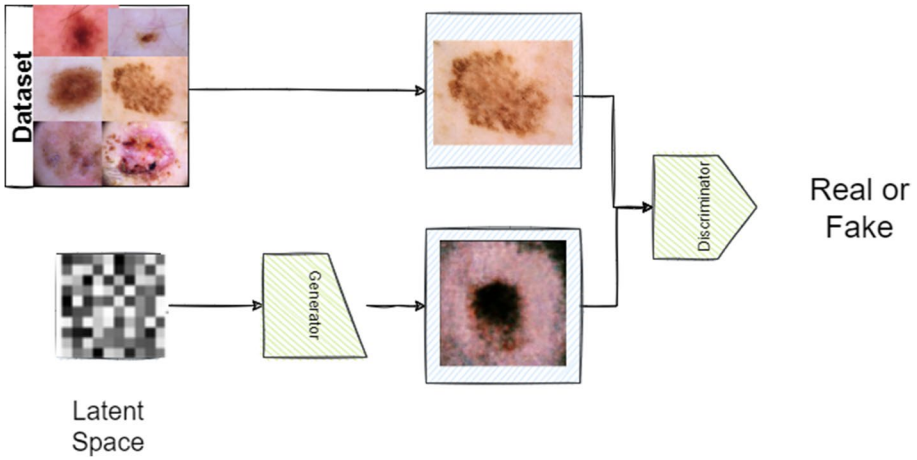


**Fig. 3** Generative Adversarial Network (GAN) for synthetic dermoscopic image generation. A number of genuine dermoscopic images are used to train a generator (*i.e.* a neural network) which generates synthetic images. A discriminator (*i.e.* a second neural network) is trained to differentiate whether an image is genuine or fake. The training of the generator and discriminator iterates until the algorithm converges. Once converged, it is expected that the generator is able to generate fake images that resemble the training images, such that the discriminator cannot differentiate whether the given image is genuine or fake

### 3.3.1 AC-GAN training and data augmentation

An AC-GAN has the same architecture as a traditional GAN model. It is comprised of two models, a "Generator" and a "Discriminator" as shown in Fig. 4. AC-GAN's distinction is the ability to generate images belonging to a specific class rather than generating images randomly. The generator, *i.e.* $G(\cdot)$, takes the class label and random points from the latent space and outputs the generated image belonging to the same class, as defined in Eq. (4).

$$X_{fake} = G(C, z) \tag{4}$$

The discriminator model, on the other hand, takes the generated image as an input and predicts two values, the probability of this image is real or fake, and the class labels the image belongs to. This is enabled by maximizing the sum of the log-likelihood of the fake/real image classification $\ell()$, as defined in Eq. (3), plus a log-likelihood of the correct class $\ell()_C$, as defined in Eq. (5).

$$\updownarrow()C = E[logP(C = c|X_{real})] + E[logP(C = c|X_{fake})] \tag{5}$$

For this study, the AC-GAN model was trained to generate images for the minority classes (all classes but NV). Our goal was to balance the dataset with a 1:1 ratio between all classes. In that manner, we generated samples that, when added to the original data, would equate to a total of 6,705 images, which is the number of NV samples present in the dataset. Figure 12 shows the data distribution of the original dataset as well as the number of generated samples added to the dataset.

Overall, training an AC-GAN is similar to training a regular GAN but with an additional step of conditioning the generator based on class labels. It is important to carefully design the architecture and loss functions of the networks in order to achieve good performance.

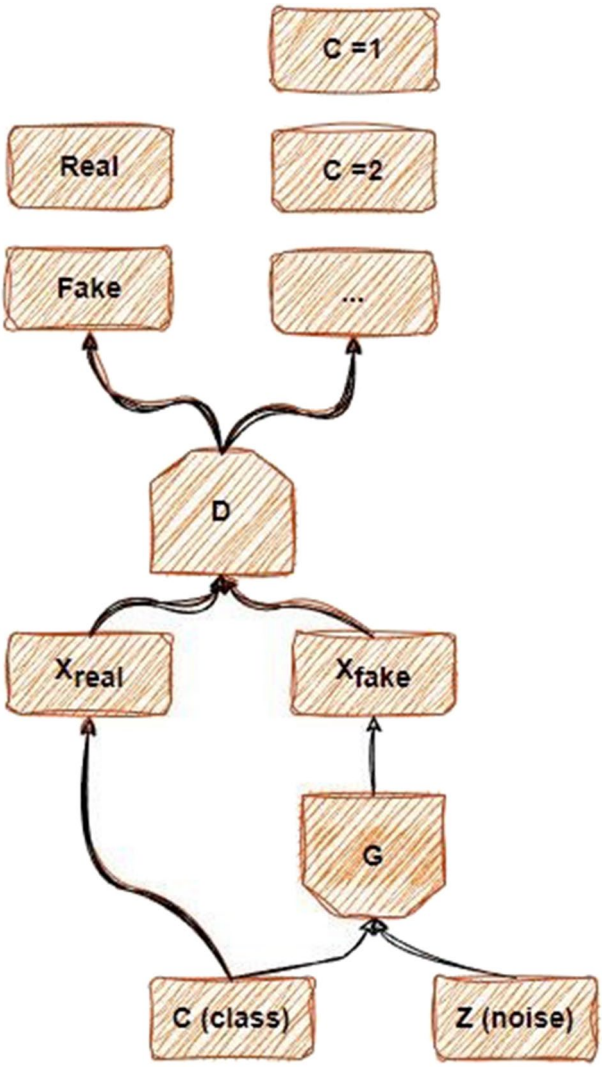


**Fig. 4** A Simple Auxiliary Classifier GAN (AC-GAN) architecture which takes into account the class label of the data into consideration to generate synthetic data

Two loss functions govern the learning process, binary cross entropy to learn the realness of an image and categorical cross entropy to learn which class it belongs to. Both loss functions are tied to the discriminator model.

### 3.4 Transfer learning and fine-tuning

Transfer learning refers to the process of applying a previously-learned model to data from relevant domains. It is now popularly used in deep learning because it can leverage deep neural networks previously trained by others without retraining the models, which require a significant amount of training data and time [56].

In our proposed framework, as shown in Fig. 1, we combine synthetic samples generated from GAN or data augmentation with the majority of samples to form a relatively balanced training set to fine-tune some classical deep learning models, including EfficientNet, ResNet, Vision transformers, and ConvNeXt. Our goal is valid whether synthetic samples, combined with transfer learning, can help boost dermoscopic image classification with imbalanced classification.

### 3.4.1 EfficientNet

About a decade ago, the accuracy performance of deep learning models for image classification improved alongside the complexity of the models for the ImageNet dataset, but these models were mostly inefficient in terms of computational load. The EfficientNet model was among the state-of-the-art CNN models because it achieved 84.4% accuracy with 66 M parameters in the ImageNet classification challenge. The EfficientNet group has eight models from B0 to B7, and while the number of estimated parameters does not rise dramatically with model size increase, accuracy does. While most convolutional neural network (CNN) models rely on the Rectifier Linear Unit (ReLU) activation function, EfficientNet employs a novel activation function dubbed Swish [11].

Compared to other state-of-the-art models, EfficientNet produces more efficient results since it consistently scales down the model in depth, width, and resolution. When working with a limited set of resources, the initial phase of the compound scaling method is to look for a grid that will reveal the connection between the various scaling dimensions of the baseline network. Scaling factors for depth, width, and resolution can be found in the following manner. To uniformly scale depth ($d$), width ($w$), and resolution ($r$), the compound coefficient $\varphi$, which is a user defined is used. where $d=\alpha^{\varphi}$, $w=\beta^{\varphi}$ and $r=\gamma^{\varphi}$. $\alpha$, $\beta$ and $\gamma$ are constants allocated by grid search [11].

Once these coefficients are determined, the starting network can be scaled to meet the desired specifications. Since the FLOPS (floating point operations per second) budget for EfficientNet is higher than some other models, for instance, MobileNetV2 [29], the inverted bottleneck MBConv is the network's primary building piece. Direct connections are utilized between bottlenecks that connect much fewer channels than expansion layers because blocks in MBConv consist of a layer that first expands and then compresses the channels [11].

When compared to conventional layers, the calculations required by these in-depth separable convolutions are reduced by nearly a $k^2$ factor. Here, $k$ is the kernel size, denoting the width and height of the 2D convolution window [11]. In Fig. 5, we show a simplified version of the EfficientNet B0 model.

### 3.4.2 ResNet50

In 2015, the ResNet50 architecture was proposed as a solution to the problems of several non-linear layers failing to learn identity mappings and deterioration. ResNet50 belongs to a family of ResNet, while the 50 identifies the number of layers in the network. ResNet50 is a stack of many residual units, a network-innetwork design. The infrastructure of the network is constructed of Residual units. Both convolutional and pooling layers make up these modules. This design employs the same VGG16-like 3-by-3 filtering architecture for input images of 224 by 224 pixels [12]. ResNet50 uses residual connections to learn residual functions that may be added to layer inputs to produce outputs. Instead of learning the full

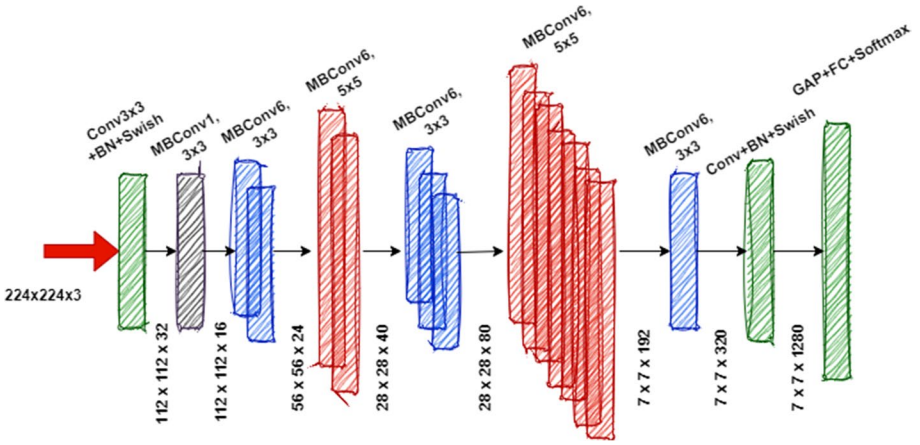


**Fig. 5** Baseline EfficientNet showing different block that makes it up. It comprises of several Inverted Residual Block or MBConv, which takes in a narrowwide-narrow approach to image classification for efficiency reasons. [11]

function, the network can learn to make input adjustments as it passes through each layer. The residual connection improves the gradient, which helps with improving the performance. ResNet achieved state-0f-the-art performance and is used in many fields [46–48]. In Fig. 6, an example of the ResNet50 model is illustrated.

### 3.4.3 Vision transformers

Vision Transformers (ViTs) is a type of neural network architecture that has been developed specifically for image recognition tasks. They are inspired by the transformer architecture, which was originally developed for natural language processing (NLP) tasks such as machine translation and language modeling.

The transformer architecture is based on self-attention mechanisms, which allow the model to attend to different parts of the input and weigh their importance when making a prediction. This allows the model to process the input in a more flexible and efficient way than traditional convolutional neural networks, which rely on fixed-size filters to extract features from the input.

Vision Transformer works by dividing the input image into a grid of patches and treating each patch as a token in a sequence. The model then processes these patches in parallel,

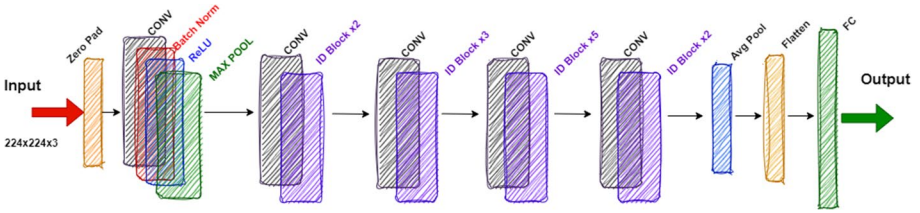


**Fig. 6** ResNet50 model architecture showing different convolutions, batch normalization, and max pooling layers of the model [12]

using self-attention to learn relationships between the different patches and make a prediction based on this information.

One of the key features of ViTs is the self-attention mechanism, which allows the model to attend to different parts of the input and weigh their importance when making a prediction. This allows the model to process the input in a more flexible and efficient way than traditional convolutional neural networks (CNNs), which rely on fixed-size filters to extract features from the input. One of the main advantages of ViTs is that they can process images of any size, as the self-attention mechanism allows the model to attend to any part of the input regardless of its position in the grid. This makes them well-suited for tasks such as object detection, where the size and position of objects in the image can vary significantly.

#### 3.4.4 ConvNeXt

ConvNeXt (Convolutional Neural Network with eXternal memory Translation) [8] is a type of neural network architecture that combines elements of CNNs and external memory networks. It was developed to improve the performance of CNNs on tasks such as image classification, object detection, and segmentation. ConvNeXt consists of two main components: a convolutional neural network and an external memory module. The CNN is responsible for processing the input and extracting features, while the external memory module is responsible for storing and retrieving information from an external memory buffer. The external memory module in ConvNeXt consists of a series of memory cells, each of which is associated with a key and a value. The keys are used to look up information in the memory, while the values are used to store the information. The external memory module can be updated by writing new values to the memory cells or by reading and modifying existing values.

One of the main advantages of ConvNeXt is that it allows the model to store and retrieve long-term dependencies in the external memory, which can be useful for tasks such as image recognition where the relationships between different parts of the input may be complex and varied. This can help the model to make more accurate predictions and improve its generalization ability. ConvNeXt [8] can outperform the Swin Transformer [9]. For example, it outperformed Swin Transformers on COCO detection [10] and ADE20K segmentation [25], achieving 87.8% top-1 accuracy on ImageNet.

## 4 Experiments

In this section, we report experimental results using the baseline approach, the simple data augmentation approach, and the AC-GAN augmentation approach using four deep learning networks fined tuned using the transfer learning approach introduced in the previous section.

### 4.1 Dataset

HAM10000 (Human Against Machine) [13] is a dataset of dermatoscopic images of skin lesions (Fig. 7), which can be used to train machine-learning models for skin cancer diagnosis. The dataset consists of over 10,000 skin lesions images, including benign and malignant tumors. The images were collected from various sources, including dermatologists,

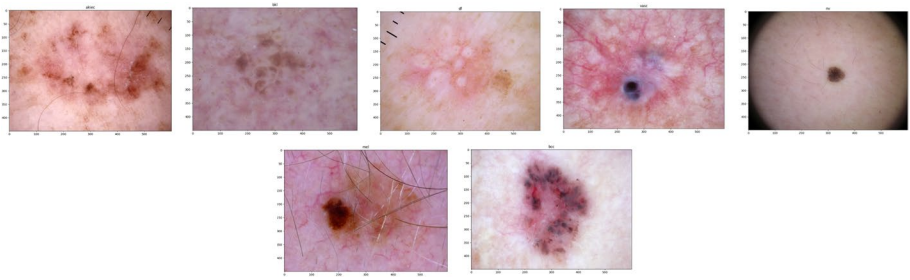

**Fig. 7** Sample Images show different types of dermoscopic images. The above images are arranged as (Left-Side)AKIEC, BKL, DF, VASC, NV, Classified as Non-Cancerous while (Right-Side) Last two MEL & BCC are Classified as Cancerous lesions

hospitals, and the internet, and are representative of a wide range of skin types and conditions.

The HAM10000 [13] dataset is extensively used in cancer research to evaluate the performance of various skin cancer diagnosis algorithms and methodologies. Multiple research has utilized it to compare CNNs, support vector machines, and decision tree classifiers. The unbalanced nature of the HAM10000 dataset presents one of the greatest obstacles for machine learning applications. Compared to benign tumors, the dataset has a relatively limited number of malignant tumors, making it challenging for machine learning algorithms to distinguish minority groups reliably.

As has been mentioned, the HAM1000 dataset has almost 67% of the images belonging to the NV class, which makes it a very imbalanced dataset. Figure 8 shows the class distribution of the original dataset. Training models on an imbalanced dataset make it biased towards the majority class. Imbalance data is a common phenomenon when it comes to medical images.

In this research, various skin pigmentation types have been divided into seven distinct classes for the purposes of multi-class categorization, and those classes are Actinic keratosis (kiec), Basal cell carcinoma (bcc), Benign keratosislike lesions (bkl), Dermatofibroma (df), Melanoma (mel), Melanocytic nevi (nv) and Vascular lesions (vasc) with a number of samples for each as 327, 514, 1099, 115, 1113, 6705 and 142, respectively. All images in the dataset are $450 \times 600$.

### 4.2 Baseline

As the original HAM1000 is imbalanced and most of the images belong to the NV class that's why the models trained on that dataset classify most of the test data as the majority class (NV). Table 1 shows the model metrics and parameters used in the experiments conducted. A general conclusion by taking a look at the accuracy of the model trained on the imbalanced dataset shows that EfficientNet performs better than other models with a test accuracy of 85.82. In Image classification, accuracy is not a good measure, and the reason can be seen by looking at the confusion matrix of the models, which shows that as the NV class has 67% of images so the number of classifications leads to high accuracy.

Figure 9 shows the confusion matrices of EfficientNet, ResNet50, ViT, and ConvNext. All four models have relatively high accuracies, but that is because most of those correctly

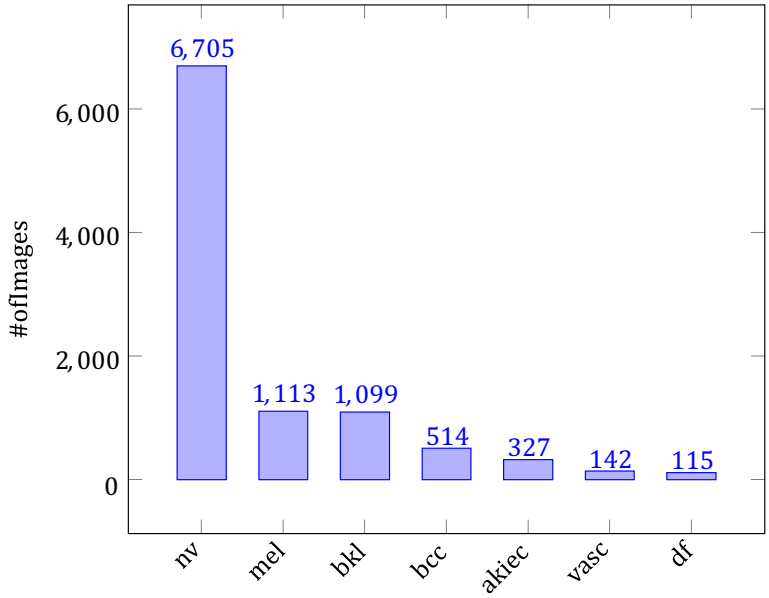


**Fig. 8** Class distributions of the benchmark dataset "HAM10000". The dataset is highly imbalanced with Melanocytic nevi (nv) containing over 67% of samples (nv, commonly called birthmarks or moles, is a non-cancerous disorder of pigment-producing skin cells)

**Table 1** Baseline model parameter settings

| Architecture | Learning Rate | Optimizer | Accuracy |
|---|---|---|---|
| EfficientNet | 0.001 | Adamax | 85.82 |
| ResNet50 | 0.001 | Adamax | 84.72 |
| ConvNext | 0.001 | Adamax | 84.82 |
| ViT | 0.001 | Adamax | 76.04 |

classified images belong to the majority class. The confusion matrix verifies this conclusion that accuracy is not a good measure for image classification.

### 4.3 A simple data augmentation procedure

As the majority class has 6,705 images from the original dataset, which makes 67% images. In order to balance the dataset, we computed a ratio between the majority class to every other class and generated a sufficient number of images to balance the dataset. The augmented images were generated with the TensorFlow ImageDataGenerator module with types shown in Table 2

Several experiments were done by utilizing different combinations of hyperparameters of the deep learning models to get the ones with the most accurate results. Table 3 shows the selected hyper-parameters and results of the respective models.

Table 3 shows hyperparameters and accuracies of the models when trained of a balanced dataset through augmentation. Comparing this with the baseline metric, we can see a clear and significant increase in accuracies. Efficientnet has the highest accuracy among

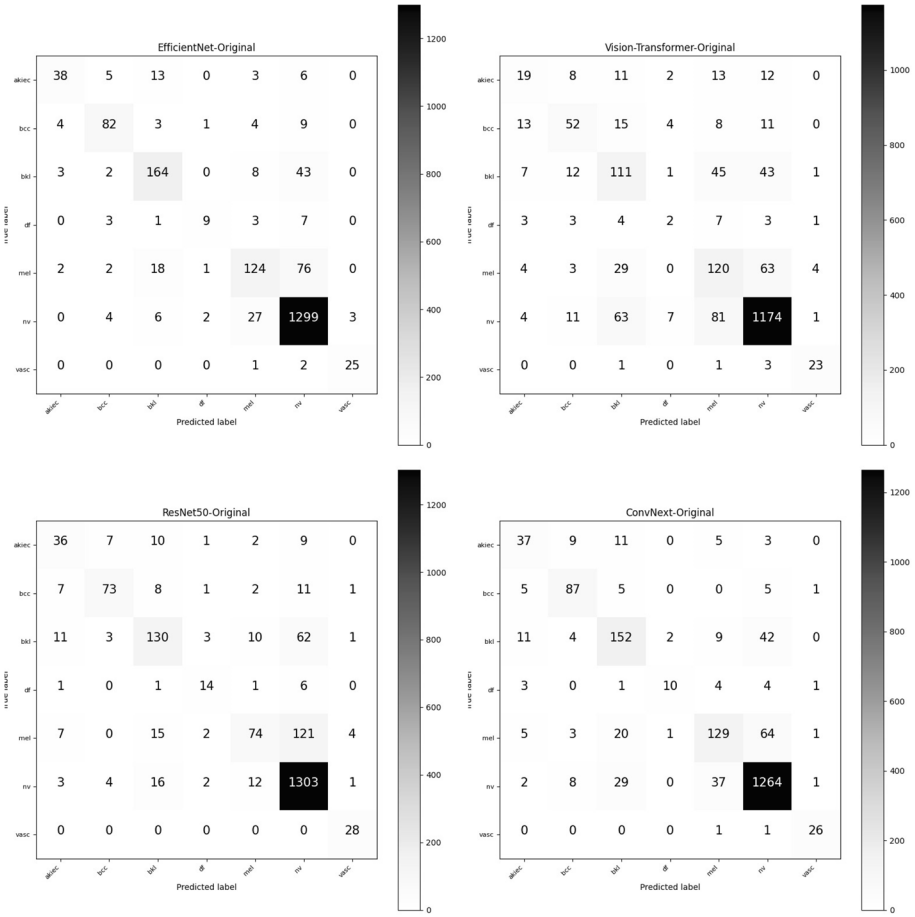


**Fig. 9** Confusion matrix of all four deep learning models when fine-tuned using imbalanced HAM10000 dataset

**Table 2** Different types of augmentations

| Type | Value | Description |
|---|---|---|
| Rotation Range | 10 | Degree of rotation between 0 and value |
| Width Shift | 0.1 | Shift the Image on X-Scale |
| Height Shift | 0.1 | Shift the Image on Y-Scale |
| Zoom Range | 0.1 | Scale upto which the image is zoomed |
| Horizontal Flip | True | Flipping the Image horizontally |
| Rescale | 2 | Rescale the Image on given value |

all, while ConvNeXt stands after. An interesting finding is that ViT outperforms ResNet50, which was not the case when models were trained on the imbalanced dataset.

Figure 10 shows the class distribution of both the original and augmented data. As you can see that the augmented data almost perfectly balances the overall data for model

**Table 3** Models Metrics with Augmented Dataset

| Architecture | L.Rate | Optimizer | Accuracy |
|---|---|---|---|
| EfficientNet | 0.001 | Adamax | 97.73 |
| ResNet50 | 0.001 | Adamax | 83.03 |
| ConvNext | 0.001 | Adamax | 97.63 |
| ViT | 0.001 | Adamax | 90.86 |

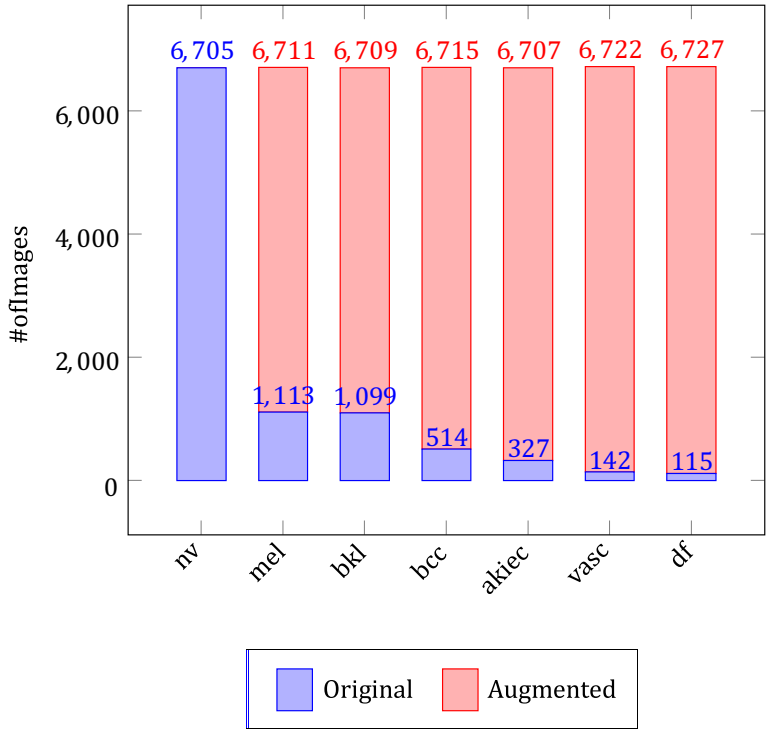


**Fig. 10** Augmented data class distributions (The final number of samples of minor classes slightly vary. This is caused by a randomization augmentation process which applies the same number of augmentations to each input image)

training. With the augmented dataset, all four models that were trained produced much better results than using the original data. Figure 11 shows the confusion matrices of all the models that were trained of 10 epochs with fivefold cross-validation.

### 4.4 Synthetic data using AC-GAN

As has been mentioned, data plays a vital role in training a deep learning model. To balance the class distribution of the imbalance HAM10000 dataset, several studies have been done on skin lesion classification by utilizing GANs. Xiang et al. [16] used AC-GAN to generate images for three classes and then DenseNet [44] to classify the images. In this study, we use GAN to produce synthetic data based on the original data.

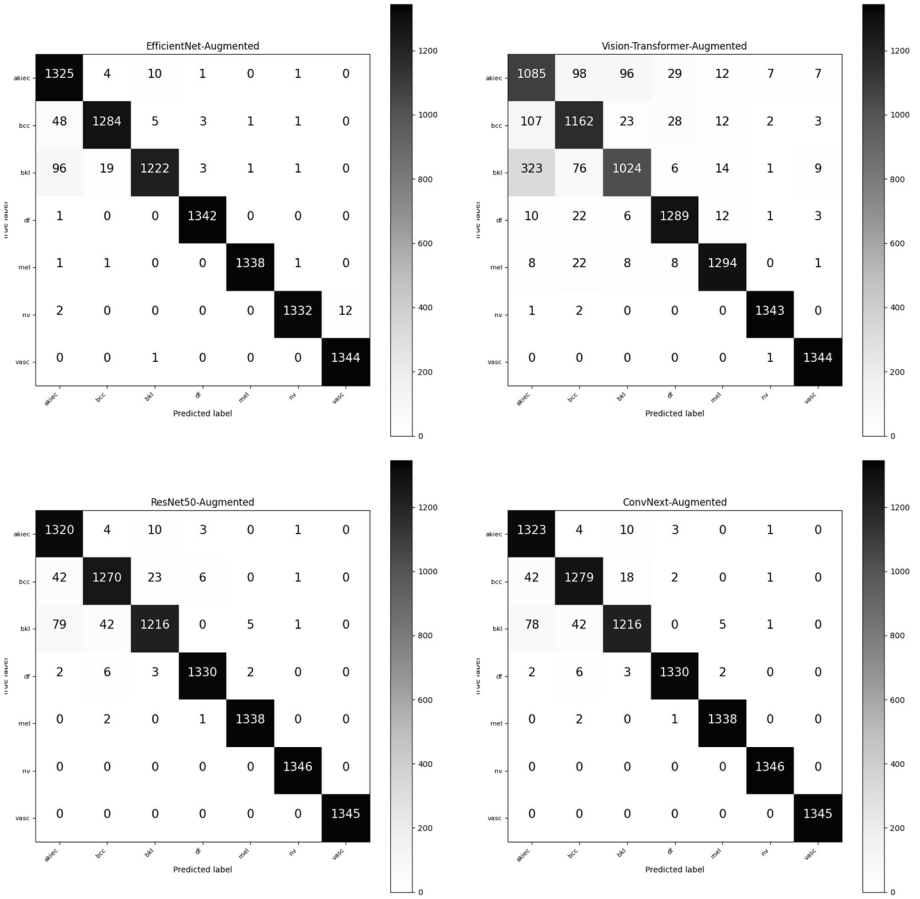


**Fig. 11** Confusion matrices of fine-tuned deep learning models using data combining majority class samples and synthetic samples generated using simple augmentation procedure

The dataset was balanced using AC-GAN as described in Sect. 3.3.1 and was used to train the model with a set of metrics shown in Table 4. Figure 12 shows the class distribution of the data generated with GAN as compared to the original dataset. Similar to the augmented data, the synthetic data was generated with AC-GAN based on its ratio to the original class distribution of the HAM10000 dataset.

**Table 4** Models Metrics with Synthetic Dataset

| Architecture | L.Rate | Optimizer | Accuracy |
|---|---|---|---|
| EfficientNet | 0.001 | Adamax | 96.79 |
| ResNet50 | 0.001 | Adamax | 96.32 |
| ConvNext | 0.001 | Adamax | 94.05 |
| ViT | 0.001 | Adamax | 93.98 |

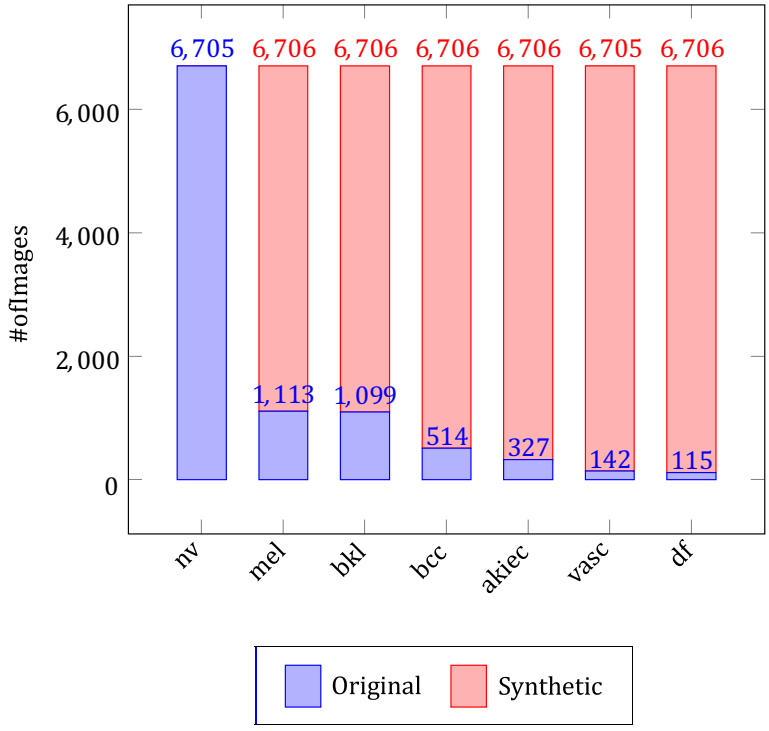


**Fig. 12** Synthetic-generated class distributions (the number of samples of the majority class and minority classes are set to be equal to each other)

Table 4 shows different hyperparameters and the accuracy of the models. Similar to the Imbalanced and Augmented, Efficientnet still performs better than all, while Resnet50 is second instead of ConvNeXt. Cross-analyzing The tables we can see that ViT has slightly higher accuracy on the synthetic data as compared to augmented data while all the other models perform much better on the augmented data.

In our experimental approach, we trained the model on synthetic minor- ity class data generated by an AC-GAN. This minority class data was then carefully combined with the majority class data from the HAM10K dataset to balance the class distribution. Our confusion matrices for models trained with synthetic data are shown in Fig. 15. We found that models trained on this expanded dataset outperformed those trained on the original HAM10K dataset. This considerable improvement shows that adding minority class synthetic data to the majority-class data improved the machine learning models. This strategy reduced class imbalance and improved classification results.

Figure 13 shows the initial batch of the synthetic data generated by the ACGAN after the first epoch, and Fig. 14 shows synthetic images after training AC-GAN for 35 epochs. Comparing Figs. 13 and 14, it is clear that, as the learning continues, AC-GAN indeed learns to improve the image quality, and the synthetic images are becoming more realistic. Another inherent advantage of AC-GAN is that the generated synthetic images are diversified in different forms, and some of them impose structures not exist in the original training samples. This suggests that GAN may generate "new" samples which

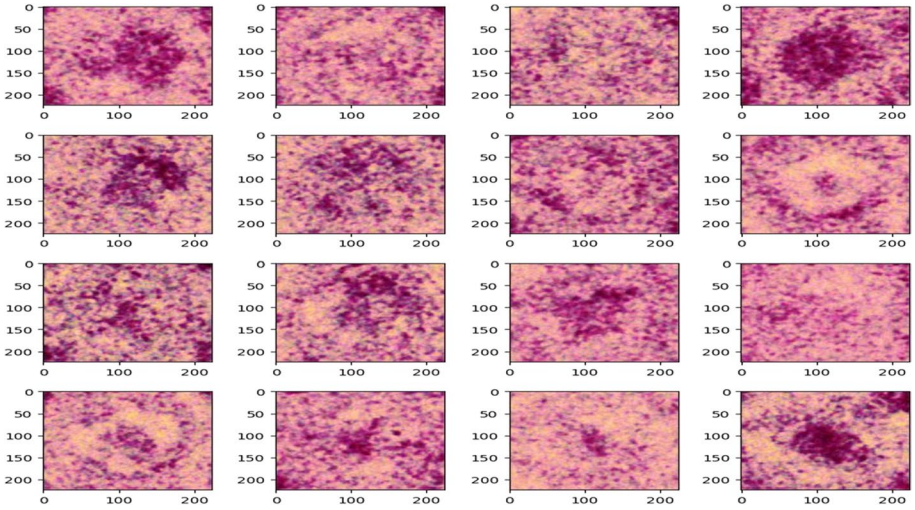

**Fig. 13** Examples of images generated from AC-GAN after the first epoch

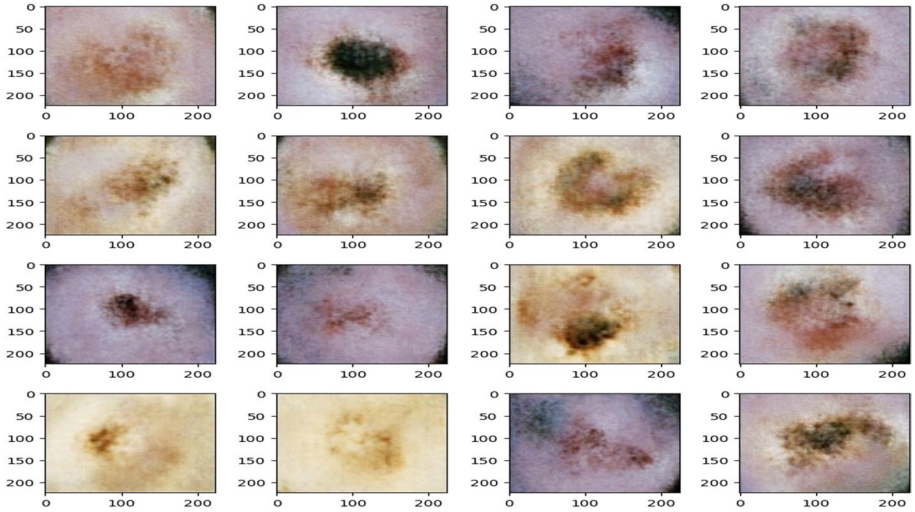

**Fig. 14** Examples of synthetic images generated from AC-GAN after 35 epochs

provide additional information to boost learning, such approaches have indeed shown improved results in several existing studies [55, 57].

Nevertheless, when comparing the quality of images generated from ACGAN (Figs. 14 and 13) *vs.* simple data augmentation (Fig. 2), it is clear that the resolutions of the images from GAN are much lower than augmentation. This is because that data augmentation largely preserves the original image resolutions. This makes images generated from data augmentation keep sufficient details crucial for deep neural network learning. As we will report in the following subsections, the ability to preserve original resolution often makes data augmentation bring better-trained models than the ones

trained from GAN-generated synthetic images as can be evaluated from the confusion matrix in Fig. 15.

Table 5 shows metrics for all models fine-tuned using the original, augmented, and GAN-generated synthetic datasets. The results show that EfficientNet trained on the augmented dataset has a relatively better performance compared to other models where Vision Transformers when trained on the GAN-generated data, perform much better than that on the augmented data (in our experiments, vision transformers were trained for 10 epochs) (Fig. 15).

### 4.5 Comparative analysis

Table 6 reports some previously published results on the benchmark data (HAM10000), including the most recent findings regarding state of the art. It becomes abundantly clear, upon close inspection, that our research has generated the most astonishing results across a wide variety of assessment criteria, which brings this finding to light. In particular, our experimental analysis demonstrates that EfficientNet outperforms its counterparts more consistently than any other model evaluated on the changed dataset. This is the case when accuracy is taken into consideration. Notably, EfficientNet exhibits excellent performance even when contrasted with ResNet50, which achieves a similar.

degree of excellence on the enriched data. This is a significant achievement. A visual similarity in the results is produced due to the accuracy of the reported values, which have been rounded to a single decimal point.

While EfficientNet and ResNet50 demonstrate amazing performance on the augmented data, the Vision transformer demonstrates outstanding performance on the synthetic data. This is something that should be highlighted because it is so impressive. When presented with synthetic data, the Vision transformer demonstrates significant improvements beyond those it has already achieved with the enhanced data. This discovery highlights the adaptability and versatility of the Vision transformer design, which makes it an appealing choice for managing a variety of data types and domains.

**Table 5** Model performance metrics (ORG: results from the original dataset, AUG: results from using original dataset and augmented samples, GAN: results from using original dataset and GAN generated synthetic data)

| Model | Data | F1-Score | Precision | Recall | AUC | Average Accuracy |
|---|---|---|---|---|---|---|
| Efficient-Net-B0 | ORG | 73.7728 | 77.0909 | 72.8951 | 84.1627 | 85.8212 |
| | AUG | 97.7386 | 97.8532 | 97.7329 | 98.6776 | 97.7340 |
| | GAN | 96.8652 | 96.8396 | 96.9202 | 98.1956 | 96.8245 |
| ResNet50 | ORG | 72.3775 | 75.0827 | 70.3324 | 82.9656 | 84.7229 |
| | AUG | 97.5116 | 97.5761 | 97.5202 | 98.5535 | 97.5207 |
| | GAN | 96.3910 | 96.3942 | 96.3956 | 97.8904 | 96.3219 |
| ConvNext | ORG | 71.1941 | 75.3978 | 69.0066 | 82.4979 | 84.8227 |
| | AUG | 97.6205 | 97.6897 | 97.6260 | 98.6153 | 97.6276 |
| | GAN | 94.1802 | 94.8059 | 93.9893 | 96.4949 | 94.0546 |
| ViT | ORG | 50.9759 | 51.9709 | 51.0734 | 72.3858 | 76.0359 |
| | AUG | 89.9892 | 90.2581 | 90.1281 | 94.2240 | 89.9147 |
| | GAN | 94.1068 | 94.2298 | 94.0456 | 96.5189 | 93.9790 |

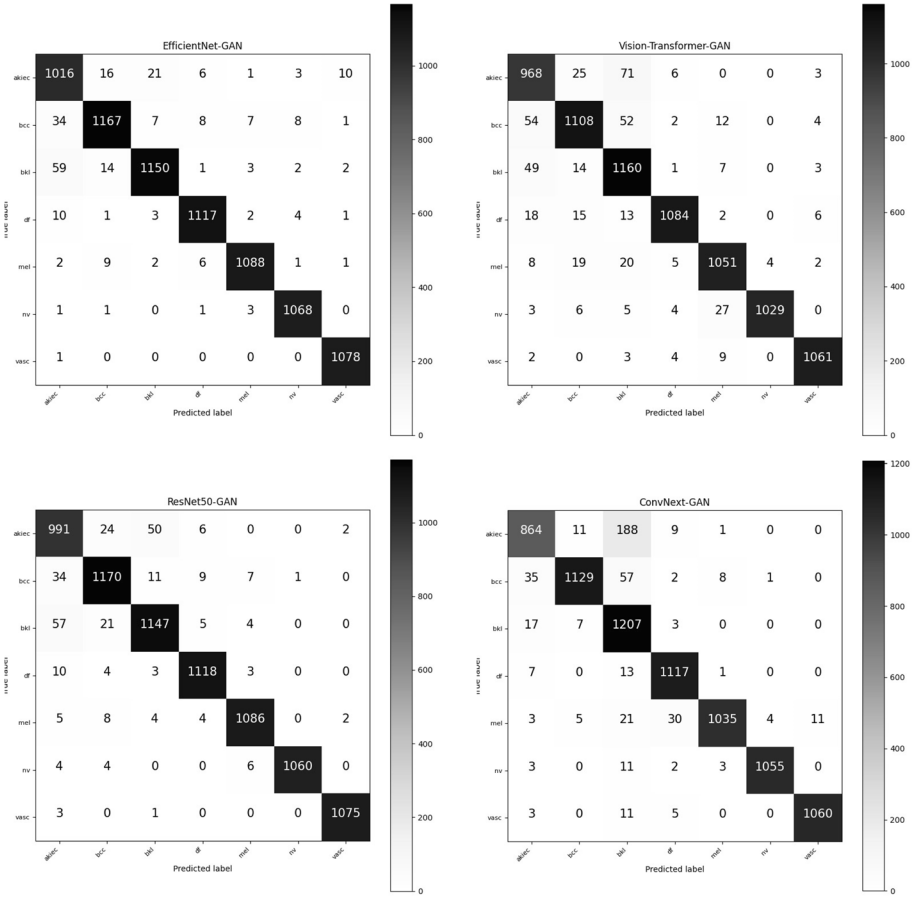


**Fig. 15** Confusion matrices of fine-tuned deep learning models using data combining majority class samples and synthetic samples generated using AC- GAN

In summary, the findings of our research, presented in Table 6 and illustrating the superiority of our suggested method across various evaluation measures, are presented here. On the synthetic data, EfficientNet and ResNet50 prove to be effective, while the Vision transformer emerges as a powerful competitor. All of these models perform relatively well on the enhanced data. These improved results highlight the significance of our research and its possible ramifications for the development of data analysis and machine learning fields.

## 5 Conclusion

Data imbalance is a common challenge in machine learning where the imbalanced class distributions often make the learning emphasize majority class samples to ensure a high classification accuracy. In this paper, we studied class imbalanced dermoscopic image classification by using transfer learning and fine-tuning pre-trained deep neural

**Table 6** Comparative analysis with published results

| Ref | Models | Augmentation | Accuracy | F1 | Prec | Recall |
|---|---|---|---|---|---|---|
| [16] | DenseNet201 | AC-GAN | 81.56 | N/A | N/A | N/A |
| | DenseNet201 | | 80.30 | | | |
| | VGG16 | | 68.38 | | | |
| | SVM Ensemble | | 85.69 | | | |
| [17] | ResNeXt101 | Augmentation | 93.2 | N/A | 88.0 | 88.0 |
| | InceptionV3 | Augmentation | 91.56 | | 89.0 | 89.0 |
| | InceptionResenetV2 | Augmentation | 93.20 | | 87.0 | 87.0 |
| [18] | GoogleNet | | 84.2 | N/A | N/A | 59.2 |
| | AlexNet | | 84.8 | | | 51.8 |
| | ReNet | | 82.8 | | | 52.0 |
| [19] | VGG16 | N/A | 75.6 | N/A | | |
| | Resnet50 | | 86.6 | | | |
| | Dense121 | | 89.2 | | | |
| | InceptiomV3 | | 74.3 | | | |
| [20] | ResNet50 | | 87.1 | | 78.6 | 77.0 |
| | InceptionV3 | | 89.7 | | 84.9 | 80.0 |
| [21] | DenseNet121 | Augmentation | 89.63 | | | |
| | ResNet | | 89.7 | | 84.9 | 80.0 |
| **Ours** | EfficientNet-B0 | Augmentation | 97.7 | 97.7 | 97.7 | 97.7 |
| | EfficientNet-B0 | AC-GAN | 96.8 | 96.8 | 96.8 | 96.9 |
| | ResNet50 | Augmentation | 97.5 | 97.5 | 97.5 | 97.5 |
| | ResNet50 | AC-GAN | 96.3 | 96.3 | 96.3 | 96.3 |
| | ViT | Augmentation | 89.9 | 89.9 | 90.2 | 90.1 |
| | ViT | AC-GAN | 93.9 | 94.1 | 94.2 | 94.0 |
| | ConvNeXT | Augmentation | 97.6 | 97.6 | 97.6 | 97.6 |
| | ConvNeXT | AC-GAN | 94.0 | 94.1 | 94.8 | 93.9 |

networks. Our experiments show that pre-trained models, such as EfficientNet, Resnet, ConvNeXt, and Vision transformers, are vulnerable and sensitive to class imbalance if data used for fine.

tuning are imbalanced. On the other hand, balance class distributions in the fine-tuning dataset bring noticeable improvement to the model performance. When balancing the class distributions, synthetic data generated using GAN and generated using augmentation of original training images are both effective. Nevertheless, using data augmentation often results in better model performance across the four validated deep neural networks compared to GANgenerated samples. We believe that this is mainly attributed to the quality of the synthetic images, where GAN-generated images have relatively low resolutions and lack sufficient details compared to simple data augmentation, such as rotation, shifting, zooming, etc. Because most deep neural networks rely on convolutional filters to learn features from local regions, as the number of convolutional filters is limited, the augmentation provides opportunities to create different local patches for filters to learn features unique to the minority classes. In summary, our research suggests that simple data augmentation is a low-hanging fruit approach to tackling data imbalance in dermoscopic image classification.

## 6 Future direction

Our study shows that GAN is relatively less effective than data augmentation in tackling data imbalance. One hypothesis is that the quality/resolution of images generated from GAN is not at satisfactory levels. Future studies are needed to quantify and investigate image quality and other factors impact on algorithms using generative models to alleviate the data imbalance. Meanwhile, our current study is primarily limited to images. Different types of datasets, including texts, time series data, and tabular data, can be used to cross-compare the results of GAN and augmented data with respect to different models, including generic machine learning classifiers.

**Acknowledgements** This research is partially sponsored by the U.S. National Science Foundation under grant No. IIS-2302786. Special thanks to Zahra Salekshahrezae, Rahmi Alagoz, Ali Salem Altaher, and Nathan Guan for their contributions to the reformatting and proofreading of the manuscript.

**Data availability** The datasets generated and used during the study reported in the paper are available through the following github repository: https://github.com/mjan2021/Dermoscopic-image-classification.git

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Divya G, Liang Q, Wang S, Zhu X (2021) An Empirical Study of Deep Learning Frameworks for Melanoma Cancer Detection using Transfer Learning and Data Augmentation. In 2021 IEEE International Conference on Big Knowledge (ICBK), pp. 38–45. IEEE
2. Ali K, Shaikh ZA, Khan AA, Laghari AA. Multiclass skin cancer classification using efficientNets–a first step towards preventing skin cancer. Neurosci Inf 2022;2(4):100034
3. Devansh B, Choromanska A, Berman RS, Stein JA, Polsky D (2019) Towards automated melanoma detection with deep learning: Data purification and augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0–0
4. Diallo, Papa Abdou Karim Karou, and Yun Ju. "Accurate detection of covid-19 using k-efficientnet deep learning image classifier and k-covid chest x-ray images dataset." In 2020 IEEE 6th International Conference on Computer and Communications (ICCC), pp. 1527–1531. IEEE, 2020.
5. Vasconcelos CN, Nader Vasconcelos B (2017) Convolutional neural network committees for melanoma classification with classical and expert knowledge based image transforms data augmentation." arXiv preprint arXiv:1702.07025
6. Salekshahrezaee Z, Leevy JL, Khoshgoftaar TM (2021) Feature extraction for class imbalance using a convolutional autoencoder and data sampling. In 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI), pp. 217–223. IEEE
7. Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier gans. International conference on machine learning. 2017(pp. 2642-2651). PML
8. Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S (2022) A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11976–11986
9. Ze L, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022
10. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Doll´ar P, Zitnick CL (2014) "Microsoft coco: Common objects in context." In European conference on computer vision, pp. 740–755. Springer, Cham
11. Atila U, Uçar M, Akyol K, Uçar E (2021) Plant leaf disease classification using EfficientNet deep learning model. Ecol Inform 61:101182

12. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778
13. Tschandl P, Rosendahl C, Kittler H (2018) The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data 5(1):1–9
14. Jan MT, Hashemi A, Jang J, Yang K, Zhai J, Newman D, Tappen R, Furht B (2023) Non-intrusive Drowsiness Detection Techniques and Their Application in Detecting Early Dementia in Older Drivers. In Proceedings of the Future Technologies Conference, pp. 776–796. Springer, Cham
15. Alsaidi M, Altaher AS, Tanveer Jan M, Altaher A, Salekshahrezaee Z (2022) COVID-19 Classification Using Deep Learning Two-Stage Approach. arXiv preprint arXiv:2211.15817
16. Xiang A, Wang F (2019) Towards interpretable skin lesion classification with deep learning models. In AMIA annual symposium proceedings, vol. 2019, p. 1246. American Medical Informatics Association
17. Chaturvedi SS, Tembhurne JV, Diwan T (2020) A multi-class skin Cancer classification using deep convolutional neural networks. Multimed Tools Appl 79(39-40):28477–28498
18. Harangi B (2018) Skin lesion classification with ensembles of deep convolutional neural networks. J Biomed Inform 86:25–32
19. Nyíri T, Kiss A (2018) Novel ensembling methods for dermatological image classification. In International conference on theory and practice of natural computing, pp. 438–448. Springer, Cham
20. Shahin AH, Kamal A, Elattar MA (2018) Deep ensemble learning for skin lesion classification from dermoscopic images. In 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), pp. 150–153. IEEE
21. Menegola A, Fornaciali M, Pires R, Bittencourt FV, Avila S, Valle E. Knowledge transfer for melanoma screening with deep learning. In: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017) 2017 Apr 18 (pp. 297–300). IEEE
22. Pomponiu V, Nejati H, Cheung N-M (2016) Deepmole: Deep neural networks for skin mole lesion classification. In 2016 IEEE international conference on image processing (ICIP), pp. 2623–2627. IEEE
23. Milton Md AA (2019) Automated skin lesion classification using ensemble of deep neural networks in ISIC 2018: Skin lesion analysis towards melanoma detection challenge. arXiv preprint arXiv:1901.10802
24. Hasan HA, Ibrahim AA (2020) Hybrid Detection Techniques for Skin Cancer Images. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp. 1–8. IEEE
25. Zhou B, Zhao H, Puig X, Xiao T, Fidler S, Barriuso A, Torralba A (2019) Semantic understanding of scenes through the ade20k dataset. Int J Comput Vision 127(3):302–321
26. Jha D, Riegler MA, Johansen D, Halvorsen P, Johansen HD (2020) Doubleu-net: A deep convolutional neural network for medical image segmentation. In 2020 IEEE 33rd International symposium on computer-based medical systems (CBMS), pp. 558–564. IEEE
27. Gessert N, Nielsen M, Shaikh M, Werner R, Schlaefer A (2020) Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. Methods 7:100864
28. Yao P, Shen S, Mengjuan Xu, Liu P, Zhang F, Xing J, Shao P, Kaffenberger B, Ronald XXu (2021) Single model deep learning on imbalanced small datasets for skin lesion classification. IEEE Trans Med Imaging 41(5):1242–1254
29. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861
30. Goodfellow I, Pouget-Abadie J, Mirza M, Bing Xu, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial networks. Commun ACM 63(11):139–144
31. Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B (2019) Interpretable machine learning: definitions, methods, and applications. arXiv preprint arXiv:1901.04592
32. Du M, Liu N, Xia Hu (2019) Techniques for interpretable machine learning. Commun ACM 63(1):68–77
33. Gajera HK, Nayak DR, Zaveri MA (2023) A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. Biomed Signal Process Control 79:104186
34. Alenezi F, Armghan A, Polat K (2023) A multi-stage melanoma recognition framework with deep residual neural network and hyperparameter optimizationbased decision support in dermoscopy images. Expert Syst Appl 215:119352
35. Emara T, Afify HM, Ismail FH, Hassanien AE (2019) A modified inception-v4 for imbalanced skin cancer classification dataset. In 2019 14th International Conference on Computer Engineering and Systems (ICCES), pp. 28–33. IEEE

36. Jan MT, Moshfeghi S, Conniff JW, Jang J, Yang K, Zhai J, Rosselli M, Newman D, Tappen R, Furht B (2023) Methods and Tools for Monitoring Driver's Behavior. arXiv preprint arXiv:2301.12269
37. Chen K, Zhuang D, Morris Chang J (2022) SuperCon: Supervised contrastive learning for imbalanced skin lesion classification. arXiv preprint arXiv:2202.05685
38. Ozturk S, Cukur T (2022) Deep clustering via center-oriented margin free-triplet loss for skin lesion detection in highly imbalanced datasets. IEEE J Biomed Health Inform 26(9):4679–4690
39. Qian S, Ren K, Zhang W, Ning H (2022) Skin lesion classification using CNNs with grouping of multi-scale attention and class-specific loss weighting. Comput Methods Programs Biomed 226:107166
40. Shen S, Xu M, Zhang F, Shao P, Liu H, Xu L, Zhang C et al (2022) A low-cost high-performance data augmentation for deep learning-based skin lesion classification. BME Frontiers 2022
41. Baur, Christoph, Shadi Albarqouni, and Nassir Navab. "MelanoGANs: high resolution skin lesion synthesis with GANs." arXiv preprint arXiv:1804.04338 (2018).
42. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
43. Yu Y, Gong Z, Zhong Ping, Shan J (2017) Unsupervised representation learning with deep convolutional neural network for remote sensing images. In Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, September 13–15, 2017, Revised Selected Papers, Part II 9, pp. 97–108. Springer International Publishing
44. Gao H, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708
45. Qin Z, Liu Z, Zhu P, Xue Y (2020) A GAN-based image synthesis method for skin lesion classification. Comput Methods Programs Biomed 195:105568
46. Mukti, Zahan I, Biswas D (2019)Transfer learning based plant diseases detection using ResNet50. In 2019 4th International conference on electrical information and communication technology (EICT), pp. 1–6. IEEE
47. Suhita Ray (2018) Disease classification within dermascopic images using features extracted by resnet50 and classification through deep forest. arXiv preprint arXiv:1807.05711
48. Shabbir A, Ali N, Ahmed J, Zafar B, Rasheed A, Sajid M, Ahmed A, Hanif Dar S (2021) Satellite and scene image classification based on transfer learning and fine tuning of ResNet50. Mathematical Problems in Engineering 2021:1–18
49. Council of the European Union (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), available at http://data.europa.eu/eli/reg/2016/679/2016-05-04, Accessed 21 February 2023
50. Selvaraju RR, Cogswell M, Das Ak, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision, pp. 618–626
51. Lucieri, A, Bajwa MN, Braun SA, Malik MI, Dengel A, Ahmed S (2020) On interpretability of deep learning based skin lesion classifiers using concept activation vectors. In 2020 international joint conference on neural networks (IJCNN), pp. 1–10. IEEE
52. Ulus C, Wang Z, Iqbal SMA, Khan KMdS, Zhu X (2022) Transfer Na¨ıve Bayes Learning using Augmentation and Stacking for SMS Spam Detection. In 2022 IEEE International Conference on Knowledge Graph (ICKG), pp. 275–282. IEEE
53. Altaher A, Salekshahrezaee Z, Zadeh AA, Rafieipour H, Altaher A (2020) Using multi-inception CNN for face emotion recognition. Journal of Bioengineering Research 3, no. 1: 1–12.
54. Abidalkareem, AJ, Abd MA, Ibrahim AK, Zhuang H, Altaher AS, Muhamed A (2020.) Diabetic retinopathy (DR) severity level classification using multimodel convolutional neural networks. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1404–1407. IEEE
55. Limeros SC, Majchrowska S, Zoubi MK, Ros´en A, Suvilehto J, Sj¨oblom L, Kjellberg M (2022) GAN-based generative modelling for dermatological applications– comparative study. In Arxiv, arXiv:2208.11702
56. Yosinski J, Clune J, Bengio Y, Lipson H (2014) Howtransferable are features in deep neural networks? In Proc. of NIPS
57. Man Wu, Wang S, Pan S, Terentis AC, Strasswimmer J, Zhu X (2021) Deep learning data augmentation for Raman spectroscopy cancer tissue classification. Sci Rep 11:23842
58. Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, Berking C, Steeb T, Enk AH (2018) Christof von Kalle. Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review, J Med Int Res 20(10):e11936
59. Sikkandar Y, Mohamed, Alrasheadi BA, Prakash NB, Hemalakshmi GR, Mohanarathinam A, Shankar K (2021) Deep learning based an automated skin lesion segmentation and intelligent classification model. J Ambient Intell humanized Comput 12:3245-3255.

60. Mirikharaji Z, Abhishek K, Bissoto A, Barata C, Avila S, Valle E, Celebi ME, Hamarneh G (2023) A survey on deep learning for skin lesion segmentation. Med Image Anal: 102863
61. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440
62. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and ComputerAssisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, pp. 234–241. Springer International Publishing
63. Pollastri F, Bolelli F, Paredes R, Grana C (2020) Augmenting data with GANs to segment melanoma skin lesions. Multimedia Tools and Applications 79:15575–15592
64. Hasan MdK, Dahal L, Samarakoon PN, Tushar FI, Martí R ((2020)) DSNet: Automatic dermoscopic skin lesion segmentation. Computers in biology and medicine 120: 103738
65. Canalini L, Pollastri F, Bolelli F, Cancilla M, Allegretti S, Grana C (2019) Skin lesion segmentation ensemble with diverse training strategies. In Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, Salerno, Italy, September 3–5, 2019, Proceedings, Part I 18, pp. 89–101. Springer International Publishing
66. Seifallahi M, HasaniMehraban A, Galvin JE, Ghoraani B (2022) Alzheimer's disease detection using comprehensive analysis of Timed Up and Go test via Kinect V. 2 camera and machine learning. IEEE Trans Neural Syst Rehabil Eng 30:1589–1600
67. Soudani A, Barhoumi W (2019) An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction. Expert Syst Appl 118:400–410
68. Xie Y, Zhang J, Xia Y, Shen C (2020) A mutual bootstrapping model for automated skin lesion segmentation and classification. IEEE Trans Med Imaging 39(7):2482–2493
69. Jin Q, Cui H, Sun C, Meng Z, Ran Su (2021) Cascade knowledge diffusion network for skin lesion diagnosis and segmentation. Appl Soft Comput 99:106881
70. Lei B, Xia Z, Jiang F, Jiang X, Ge Z, Yanwu Xu, Qin J, Chen S, Wang T, Wang S (2020) Skin lesion segmentation via generative adversarial networks with dual discriminators. Med Image Anal 64:101716
71. Tu W, Liu X, Wei Hu, Pan Z (2019) Dense-residual network with adversarial learning for skin lesion segmentation. IEEE Access 7:77037–77051
72. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille L, Zhou Y (2021) Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306
73. Gulzar Y, Khan SA (2022) Skin lesion segmentation based on vision transformers and convolutional neural networks—A comparative study. Appl Sci 12(12):5990
74. Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W (2022) Development of a wearable belt with integrated sensors for measuring multiple physiological parameters related to heart failure. Sci Rep 12(1):20264
75. Iqbal SMA, Asghar W (2023) Smartphone Integration with Point-of-Care Devices for Disease Diagnostics. In Emerging Technologies In Biophysical Sciences: A World Scientific Reference: Volume 3: Emerging Technologies for Diagnostics, pp. 317–335
76. Jasil SPG, Ulagamuthalvi V (2021) Deep learning architecture using transfer learning for classification of skin lesions. J Ambient Intell Humanized Comput: 1–8
77. Khan Attique, Muhammad Muhammad Sharif, Akram Tallha, Kadry Seifedine, Hsu Ching-Hsien (2022) A two-stream deep neural network-based intelligent system for complex skin cancer types classification. Int J Intell Syst 37(12):10621–10649
78. Naqvi Maryam, Gilani Syed Qasim, Syed Tehreem, Marques Oge, Kim Hee-Cheol (2023) Skin Cancer Detection Using Deep Learning-A Review. Diagnostics (Basel, Switzerland) 13(11):1911
79. Qasim Gilani S, Syed T, Umair M, Marques O. Skin Cancer Classification Using Deep Spiking Neural Network. J Digit Imaging (2023): 1–11
80. Gilani SQ, Marques O (2023) Skin lesion analysis using generative adversarial networks: A review. Multimedia Tools Appl: 1–42