

## Article

# Improving Skin Cancer Classification Using Heavy-Tailed Student T-Distribution in Generative Adversarial Networks (TED-GAN)

Bilal Ahmad <sup>1</sup>, Sun Jun <sup>1,\*</sup>, Vasile Palade <sup>2</sup>, Qi You <sup>1</sup>, Li Mao <sup>1</sup> and Mao Zhongjie <sup>1</sup>

<sup>1</sup> School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; bilalahmad@stu.jiangnan.edu.cn (B.A.); 7171905005@stu.jiangnan.edu.cn (Q.Y.); wxmaoli@jiangnan.edu.cn (L.M.); 7181905016@stu.jiangnan.edu.cn (M.Z.)

<sup>2</sup> Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry CV1 5FB, UK; ab5839@coventry.ac.uk

\* Correspondence: junsun@jiangnan.edu.cn

**Abstract:** Deep learning has gained immense attention from researchers in medicine, especially in medical imaging. The main bottleneck is the unavailability of sufficiently large medical datasets required for the good performance of deep learning models. This paper proposes a new framework consisting of one variational autoencoder (VAE), two generative adversarial networks, and one auxiliary classifier to artificially generate realistic-looking skin lesion images and improve classification performance. We first train the encoder-decoder network to obtain the latent noise vector with the image manifold's information and let the generative adversarial network sample the input from this informative noise vector in order to generate the skin lesion images. The use of informative noise allows the GAN to avoid mode collapse and creates faster convergence. To improve the diversity in the generated images, we use another GAN with an auxiliary classifier, which samples the noise vector from a heavy-tailed student t-distribution instead of a random noise Gaussian distribution. The proposed framework was named TED-GAN, with T from the t-distribution and ED from the encoder-decoder network which is part of the solution. The proposed framework could be used in a broad range of areas in medical imaging. We used it here to generate skin lesion images and have obtained an improved classification performance on the skin lesion classification task, rising from 66% average accuracy to 92.5%. The results show that TED-GAN has a better impact on the classification task because of its diverse range of generated images due to the use of a heavy-tailed t-distribution.

**Keywords:** variational autoencoder; generative adversarial networks; melanoma detection; skin cancer classification; student t-distribution; heavy-tailed distribution; t-distribution; informative noise vector; deep learning; convolutional neural networks; GANs; VAE



**Citation:** Ahmad, B.; Jun, S.; Palade, V.; You, Q.; Mao, L.; Zhongjie, M. Improving Skin Cancer Classification Using Heavy-Tailed Student T-Distribution in Generative Adversarial Networks (TED-GAN). *Diagnostics* **2021**, *11*, 2147. <https://doi.org/10.3390/diagnostics11112147>

Academic Editors: Shang-Ming Zhou, Haider Raza, Honghan Wu and Ayman El-Baz

Received: 4 October 2021

Accepted: 9 November 2021

Published: 19 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Melanoma is the least common but most brutal of all skin lesions, with the highest mortality rate per year worldwide [1]. With effective early-stage diagnoses, the survival rate of patients increased substantially; reported to be 98.5% [2]. In contrast, the 5-year survival rate decreases to 19.9% only, with the failure of timely detection of melanoma [2]. Patients' skin color and similarities among various skin lesions make it hard for medical experts to diagnose it correctly at the initial stage. Highly expert dermatologists can diagnose melanoma visually with an accuracy of 60% only [3].

A technique with better accuracy commonly used by dermatologists is dermoscopy. It is an in-vivo and non-invasive technique that eliminates skin surface reflection and magnifies it up to 400% for better examination. Dermoscopy is also called epiluminescence microscopy or dermatoscopy. In this method, the diagnosis accuracy of melanoma is up to 84% [4,5]. It is a technique with better accuracy but is time and resource-consuming,

and the diagnosis efficiency is easily affected by various physical and human factors [6]. Therefore, it is necessary to develop more accurate and faster computer-aided diagnosis (CAD) systems to reduce the burden on healthcare systems and assist dermatologists in making better decisions.

In classic CAD and machine-based systems, researchers use several image processing filters to extract an image's characteristics and features. For example, the Harris Corner detector is a Gaussian window function for detecting the edges and corners of the image. Median or averaging filters were used to reduce the noise. Such methods were difficult, time-consuming, and useful for small data sets only. In addition, it was difficult to transfer the algorithm's learning to another unseen challenge.

Recently, Convolutional Neural Networks (CNNs) and image processing techniques gained tremendous attention from researchers of all the fields of science due to their good performance, especially in medical imaging applications [7–12]. Deep learning models are composed of many layers and can have millions of parameters. The deeper the model, the more data is needed to train it. Otherwise, the model starts to exhibit overfitting. Overfitting is a common problem in deep learning architectures that are trained on small datasets.

A CNN-based end-to-end system requires a large quantity and high quality of labeled data to function well. Having a high-quality labeled dataset, especially in medicine, is expensive and time-consuming. Melanoma, for example, is the most serious type of skin cancer. In actual practice, only 20% of patients are diagnosed with melanoma following a biopsy, and not all patients go for a biopsy in the first place [6]. This indicates that if a biopsy is not recommended to the patient in the first place, the odds of melanoma growing and progressing to the worst stage increase; conversely, recommending a biopsy for each patient increases the burden on the healthcare system. To address the limited availability of medical data and overfitting, researchers focus on testing and proposing novel ways to use deep learning models with small datasets.

Data augmentation and transfer learning are two clear strategies that academics are increasingly employing to handle the challenge of limited data. In these methods, researchers extend the training set in data augmentation by modifying the training images by scaling, cropping, and rotating them. Transfer learning involves using a pre-trained model that has already been trained on a big dataset and retrains a few layers of the model on a small training set. It works well for classification tasks involving common generic species such as cats, dogs, trees, etc. Naturally, this is not the case with a medical dataset. We do not usually have pre-trained models that can be employed for transfer learning in medical image classification tasks because of the restricted availability and difficulty of obtaining medical datasets [13,14].

One of the possible solutions recently adopted by the researchers to solve the limited availability of medical datasets is to generate artificial data. Among several generative models, Generative Adversarial Networks (GANs) [15] have gained the attention of medical image processing researchers. Recently, GANs were used in medical image (MI) generation [16,17], medical image editing in latent space [18], MI segmentation [19], and MI classification [20], because of their better performance among all generative models. A typical GAN consists of a generator (G) and a discriminator (D) network. The generator's job is to produce more realistic images from random noise by learning the distribution of real images. In contrast, the discriminator's job is to differentiate whether the images fed to it are real or fake (generated by the generator). They are trained alternatively to reach the final convergence. Its popularity is mainly because it automatically learns the image distribution in an unsupervised manner without employing, for example, a Markov Chain Monte Carlo (MCMC) approximation. Another beauty of GANs which makes them stand out in several traditional generative models is that they generate the image as a whole, not pixel by pixel, which provides more diversity in the generated images.

Mode collapse, on the other hand, is a common issue in GANs. They need a significant amount of training data to avoid this. In the event of mode collapse, a GAN is unable to generate clear images. Some scholars have addressed these issues and proposed GAN

extensions [21]. All of them used random Gaussian noise as an input to sample the noise vector for the generator. Because random Gaussian noise does not have a strong tail, the image distribution learned from it cannot guarantee diverse image production.

Heavy-tailed distributions such as the Cauchy, log-normal, and t-distributions have been utilized to replace the random noise Gaussian distribution with excellent results. Student t-distribution is an infinite mixture of Gaussian distribution. The t-distribution was utilized in machine learning research as a replacement for the random noise Gaussian distribution [22,23]. Because of its extended tail features, it achieves excellent results in terms of diversity and low computational cost. For example, among all other income distributions, log t-distribution provides the greatest fit for predicting the income distribution of European Union countries [22]. Van et al. presented a modification to Stochastic Neighbor Embedding that uses the t-distribution to show high-dimensional data in a low-dimensional latent space by assigning a place in a two-dimensional map to each data point. This modification was much faster to optimize because of the t-distribution instead of Gaussian noise [23].

Even though numerous enhancements to GANs have been developed, the noise distribution it uses to sample the input noise vector is still random Gaussian distribution. This area, input noise distribution, has not yet been explored in generative models.

We used two generative adversarial networks (GANs) and one variational autoencoder (VAE) in the proposed methodology in this paper. As VAEs are easy to train compared to GANs, we first train the variational autoencoder (VAE) separately on our training dataset. Then, we swapped the encoder-decoder network into a decoder-encoder network and let the GAN<sub>1</sub> (the first generative model of our framework) sample the input from the output of VAE. This way, GAN<sub>1</sub> has fewer chances to collapse and tends to adopt the domain information easily. For the training of our main generative adversarial network GAN<sub>2</sub>, we used student t-distribution instead of random Gaussian noise. GANs tend to perform better generally if we have a large amount of training data. We increased the training data of GAN<sub>2</sub> by using the images generated by GAN<sub>1</sub> at the previous stage. Furthermore, we added an auxiliary classifier network to the discriminator. The main generative adversarial network (GAN<sub>2</sub>) and the auxiliary classifier were trained together. The block diagram of the proposed methodology is presented in Figures 1 and 2. The main contributions of this work can be summarized as follows:

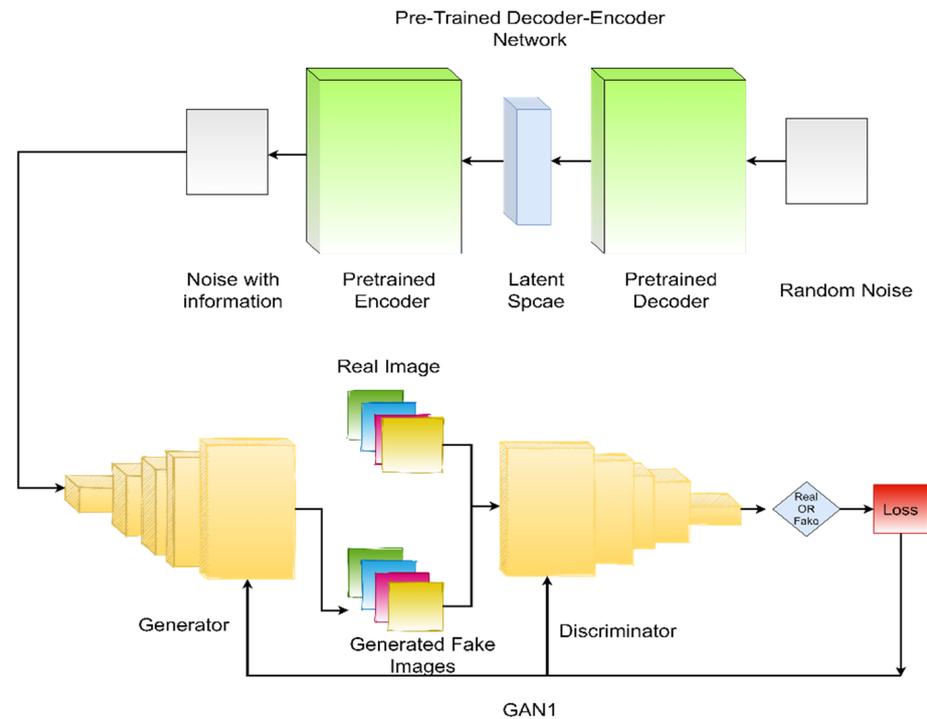
- We used the VAE network to produce a noise vector that has the domain information.
- We used heavy-tailed student t-distribution to add diversity in the generated medical images.
- We used an auxiliary classifier to push the network to produce images from a specific category.
- To the best of our knowledge, this is the first time that, instead of using random noise, a separate network was trained to obtain domain information and used that informative noise for the generation of medical images.

The rest of this paper is organized as follows. Section 1.1 reviews some previous work related to generative models, including GANs and their applications in medical image generation. Section 2 reports the proposed TED-GAN and experiment settings in detail. Results, discussion, and conclusion are presented in Sections 3–5, respectively.

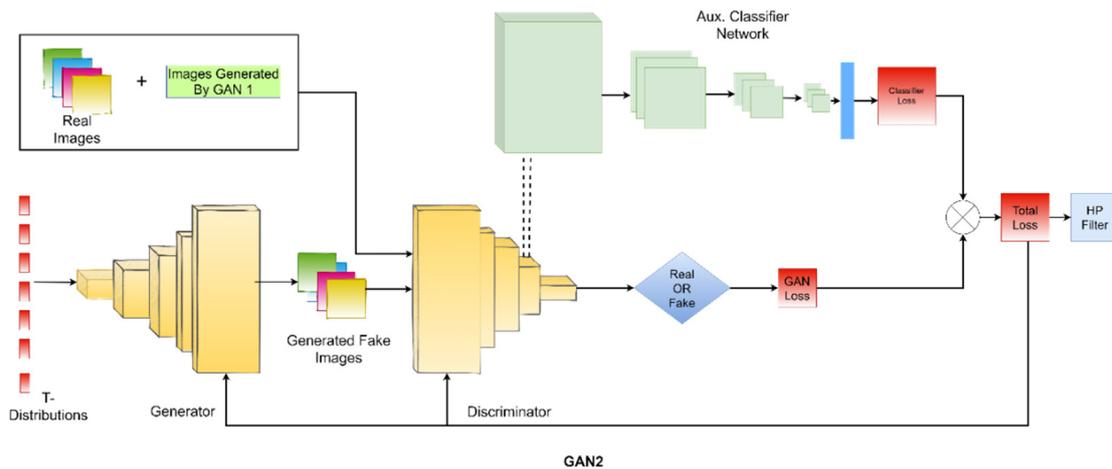
### 1.1. Related Work

Before the dawn of generative adversarial networks (GANs), Deep Belief Networks (DBNs), Restricted Boltzmann Machines (RBMs), Deep Boltzmann Machines (DBMs), Generative Stochastic Network (GSN), Autoencoders (AE), Denoising Autoencoders (DAEs), and variational autoencoders (VAEs) were heavily used for generating new samples of image data. Among them, several generative models are modeled by Markov Chain Monte Carlo (MCMC) based approximations. When the gradient vanishes during the training process, MCMC-based algorithms approximate the gradient of log-likelihood [24]. In image and video datasets, the likelihood-based approaches face the curse of dimensional-

ity. Moreover, the Markov chain approximation in high dimensional spaces is inaccurate, blurry, and computationally slow [25].



**Figure 1.** Encoder-decoder of VAE is swapped after the training so that output noise from the decoder could be used as an input to the generator of GAN<sub>1</sub>. The figure represents the VAE and the GAN<sub>1</sub> of the proposed methodology. It’s a partial block diagram representation of the proposed TED-GAN.



**Figure 2.** The generative adversarial network, GAN<sub>2</sub>, of the proposed framework and the auxiliary classifier. GAN<sub>2</sub> samples the input noise vector from student t-distribution and uses the images generated by GAN<sub>1</sub> for training. Generated images are passed through a high pass filter to improve imperceptibility.

In 2007, Hays et al. [26] proposed a scene completion algorithm that uses huge data of non-annotated images. The algorithm was efficient enough to find similar images from the database to complete the task of scene completion in the host image.

These traditional algorithms have a limited capacity of image estimation, as they generate the new images pixel-to-pixel instead of estimating the image as a whole.

Deep Autoencoders are based on deep learning in which the encoder converts the input images into a latent representation. It reduces the dimensionality of an input representation, and the decoder tries to reconstruct the original images from their latent representation. The objective of the autoencoders is to reduce the reconstruction error.

A variation of autoencoders was proposed in 2013 by Kingma et al. [27], named variational autoencoders (VAEs). In a variational autoencoder, the encoder maps the image representation into a Gaussian vector, and the decoder maps the noise vector to the new image. These generative models suffer from the problem of generating blurry images.

However, with the advent of deep neural networks, image generative models have been revitalized in recent years. Particularly, Generative Adversarial Networks (GANs) have shown promising results in synthesizing realistic images [15]. To improve the image generating capability of GANs, researchers proposed various variations of GANs.

In a deep convolutional generative adversarial network (DCGAN), Radford et al. integrated a convolution operation into the GAN to improve GAN's performance [15,21]. Moreover, DeLiGAN samples the noise vector from random noise and inputs it into the generator to improve the diversity of generated images [28]. Ma et al. combined the meta-learning with CGAN and proposed a new variant of GAN called MetaCGAN. MetaCGAN can transfer the information and style it learned during training on a large dataset to the new task with a small dataset [29]. Recently, some researchers used recurrent and convolutional neural networks (RNNs, CNNs) to generate high-resolution images. However, the algorithms generate the images pixel by pixel instead of generating images as a whole [30,31].

#### GAN Applications in Medical Imaging

The GAN and GAN-based networks that generate synthetic medical images have become very popular recently, as they solve the problem of limited availability of medical datasets. Liu et al. [32] proposed a variant of Cycle GAN that uses the Pseudo Cycle consistent module and the domain control module to generate the Computed Tomography (CT) images. In this approach, the Pseudo Cycle consistent module controls the consistency of generated images, and the domain control module provides additional information of the domain.

Jiang et al. [33] proposed a Fused Attentive GAN (FA-GAN) for generating and reconstructing super-resolution magnetic resonance (MR) images. He introduced local and global feature extraction modules at different levels to extract useful features. In FA-GAN, they used 40 sets (consisting of 256 slices) of 3D magnetic resonance (MR) images for training the network, and PSNR and SSIM are used as performance measure matrices.

Ting et al. used the GAN synthesized ultrasound images of the breast and used them at the augmentation stage in the classification problem of breast lesions [20].

Ali et al. proposed Cascade Ensemble Super-Resolution CESR-GAN to reconstruct the high-resolution skin lesion images from low-resolution images. They introduced a new lossfunction based on features of the images [16].

Simulation of medical images in diverse areas of medicine is a very challenging and hot area of research these days. Synthesizing mass images in mammograms is one of them. Shen et al. used GANs to produce mass images and then fill them with contextual information by adding the synthetic lesions to healthy mammograms. They claim that their proposed network could learn the shape, context information, and distribution of real images [17]. Other researchers used GANs in various fields of medicine, including Mahapatra et al. [34], who generated retinal fundus images; Shin produced abnormal MRI images [35]; Han et al. used two-step GAN to produce MR images of brain segments with and without tumors [36]; Nie et al. generated pelvic CT images [37].

GANs have been used widely in various fields of medical imaging. Researchers try to make improvements in results by using heavy and deep architectures. Playing around the input noise of GANs is still unexplored in medical imaging.

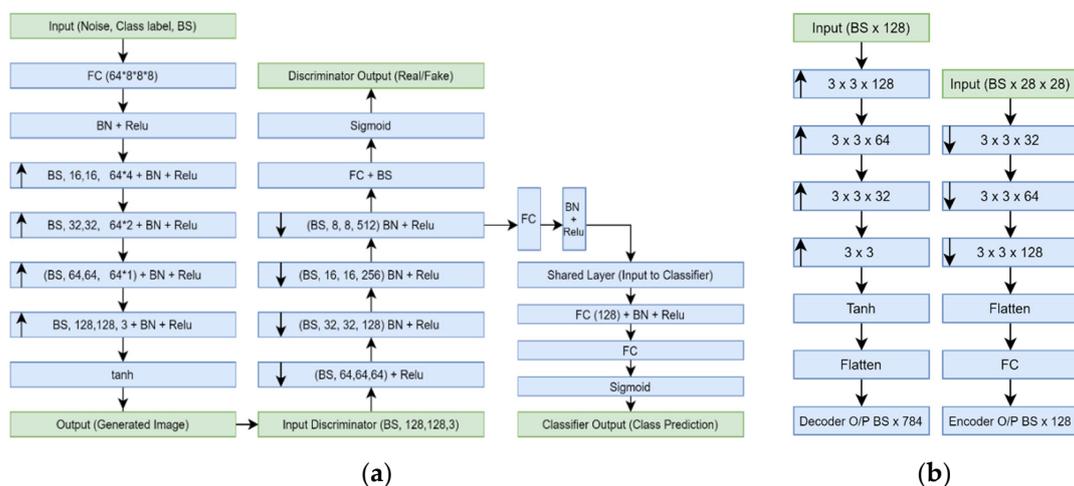
## 2. Materials and Methods

### 2.1. Proposed Method

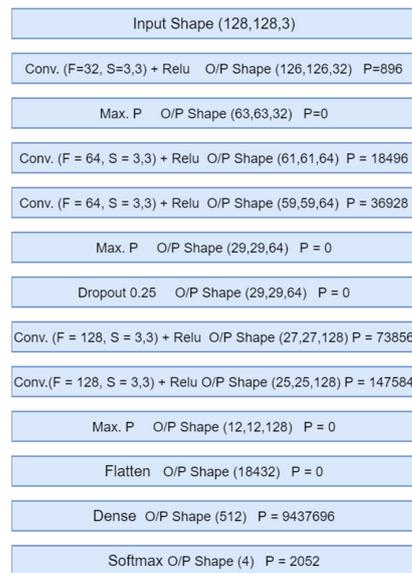
In this section, we introduce our proposed framework, TED-GAN, in detail. As VAEs are easy to train compared to GANs, we first train the variational autoencoder (VAE) on our training dataset to let the network store the information of the image manifold. Then we swap the network from an encoder-decoder to a decoder-encoder network. When the random Gaussian noise vector passes through this trained decoder-encoder network, it now produces a noise vector that is no longer random, but has the information of the domain. Zhong et al. [38] proved mathematically that the noise produced by swapping the encoder-decoder network has the information of image manifold. GAN<sub>1</sub> would sample the noise vectors from this informative noise to produce realistic images. GAN<sub>1</sub> now has a very minute chance of collapse and tends to adopt the domain information easily as its input is sampled from a latent vector of trained VAE instead of random noise.

For the training of our main generative adversarial network GAN<sub>2</sub>, we used student t-distribution instead of random Gaussian noise. GANs tend to perform better generally if we have a large amount of training data. So, we used the images generated by GAN<sub>1</sub> along with real images of the training set (images from the HAM10000 dataset [39]) to feed the discriminator. Furthermore, we added a classifier network in front of the discriminator that shares the feature layer of the discriminator. This is called an auxiliary classifier. The main generative adversarial network (GAN<sub>2</sub>) and the auxiliary classifier are trained together. In the end, generated skin lesion images passed through a high pass filter to improve the imperceptibility. The small step of adding the high pass filter enhanced the quality of images significantly. The whole framework consists of one decoder-encoder network, two GANs, and one classifier. We name it TED-GAN; T from t-distribution, and ED from the encoder-decoder network.

The TED-GAN is used to generate skin lesion images only. We built a separate CNN classifier consisting of a few layers for the fair comparison of classification results with other generative models. This classifier was used to compare the performance of various generative models with the proposed one. The architecture details of the TED-GAN and the block diagram of the CNN classifier are presented in Figure 3a,b and Figure 4, respectively.



**Figure 3.** ↑, ↓, BS, FC, BN, and O/P indicate upsampling, downsampling batch size, fully connected layer, batch normalization, and output, respectively. (a) The architecture of Generative adversarial networks used in this study. (b) Variational autoencoder architecture.



**Figure 4.** The architecture of a CNN classifier is used to test the classification performance of different generative models. Here P, Max. P, F, and S represent the parameters, maximum pooling, number of filters, and filter size, respectively.

### 2.1.1. Variational Autoencoders (VAEs)

The variational autoencoders (VAEs) consist of two parts; the encoder and decoder. The encoder consists of a separate network that samples the data  $x$  from original data and tries to learn the latent representations ( $z$ ), whereas the decoder network tries to reconstruct the original image  $x'$  from the latent representation  $z$  [40].

$$\begin{aligned} z &\rightarrow Enc(x) = q(z|x) \\ \tilde{x} &\rightarrow Dec(z) = \mathcal{P}(x|z) \end{aligned} \tag{1}$$

Typically,  $z$  is sampled from Gaussian distribution  $N(0,1)$ . The VAE objective function consists of two terms, reconstruction error and a regularization term, given in the following equation.

$$L_{(VAE)} = -\mathbb{E}_q(z|x) \left[ \log \frac{\mathcal{P}(x|z)\mathcal{P}(z)}{q(z|x)} \right] \tag{2}$$

$$= -\mathbb{E}_q(z|x) [\log \mathcal{P}(x|z)] + D_{kl}(q_\phi(z|x) || \mathcal{P}(z)) \tag{3}$$

where the term  $-\mathbb{E}_q(z|x) \log \mathcal{P}(x|z)$  is called reconstruction error ( $L_{(Rec)}$ ) and  $D_{kl}(q_\phi(z|x) || \mathcal{P}(z))$  is the Kullback-Leibler divergence.

### 2.1.2. GAN with Student T-Distribution

As we discussed in the introduction, heavy-tailed t-distribution tends to produce better results than random Gaussian noise. In this section, we will discuss the reparameterization trick of t-distribution. In generative models, backpropagation does not hold as it is. To sample the noise from a student t-distribution instead of a standard normal distribution and to reduce the generative and discriminative loss, we propose a reparameterization of the latent generative space, using a mixture of students' t-distributions. The probability density function (pdf) of multivariate student t-distribution is given by

$$f(x) = \frac{\Gamma\left[\frac{1}{2}(d+n)\right]}{\Gamma\left(\frac{d}{2}\right) (d\pi)^{\frac{n}{2}} |\Sigma^*|^{\frac{1}{2}}} \left[ 1 + \frac{(\underline{x} - \underline{\mu})' (\Sigma^*)^{-1} (\underline{x} - \underline{\mu})}{d} \right]^{-\frac{d+n}{2}} \tag{4}$$

$$f(x) = \frac{\Gamma\left(\frac{d+n}{2}\right)}{\Gamma\left(\frac{d}{2}\right) d^{\frac{n}{2}} \pi^{\frac{n}{2}} |\Sigma^*|^{\frac{1}{2}}} \left[1 + \frac{1}{d} (x - \underline{\mu})' (\Sigma^*)^{-1} (x - \underline{\mu})\right]^{-\frac{d+n}{2}} \quad (5)$$

where  $\underline{\mu}$ ,  $n$ ,  $d$  &  $\Sigma^*$  represents its parameters and is given by:

Mean vector  $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$

Number of variables =  $n$

Degree of freedom =  $d$

Positive-definite symmetric matrix  $\Sigma^*$  of size =  $n \times n$ , respectively.

Covariance matrix ( $\Sigma_s^*$ ) of variables that follow the pdf of t-distribution with a degree of freedom  $d > 2$  is given by

$\Sigma_s^* = d' \Sigma^*$  where  $d' = d(d-2)^{-1}$  for  $d > 2$ . Therefore, multivariate student t-distribution can be written as  $t(\underline{\mu}, d' \Sigma^*)$  with parameters definite matrix ( $\Sigma^*$ ), degree of freedom ( $d'$ ) and mean vector ( $\underline{\mu}$ ).

A t-distribution  $t(\underline{\mu}, d' \Sigma^*)$  is said to be a standard t-distribution if it has a mean vector equal to zero ( $\underline{\mu} = \underline{0}$ ), covariance matrix equal to the identity matrix ( $\Sigma^* = \underline{I}$ ), and can be written as  $t(\underline{0}, d' \underline{I})$ . Where  $\underline{0}$  is a vector with all zeros and  $\underline{I}$  is an identity matrix.

To sample a noise vector, we first randomly select one of the t-distributions and then sample an n-dimensional vector from the selected t-distribution. Thus, as [41] claimed as well, sampling of noise vector and linear transformation of t-variable still follows the t-distribution. Thus, sampling a noise vector  $z^* \rightarrow t(\underline{\mu}_i, d' \Sigma_i^*)$  from the general t-distribution becomes equivalent to sampling  $\varepsilon^* \rightarrow t(\underline{0}, d' \underline{I})$  and calculating  $z^*$  according to Equation (6).

$$z^* = \underline{\mu} + \Sigma_i^* \varepsilon^*, \quad \varepsilon^* \in t(\underline{0}, d' \underline{I}) \quad (6)$$

So, the distribution  $\mathcal{P}_{z^*}(z^*)$  of latent noise can be written as:

$$\mathcal{P}_{z^*}(z^*) = \sum_{i=1}^N \pi_i t_i(\underline{\mu}_i, d' \Sigma_i^*) \quad (7)$$

where  $N$  represents the number of t-distributions and  $\pi_i$  is the weight of each element. We assume  $\Sigma_i^*$  is a diagonal matrix and each diagonal element is initialized with non-zero value of 0.03, close to the suggestion of Sun et al. [42]. We initialize each element of  $\underline{\mu}_i$  by sampling from a uniform distribution of range from  $-1$  to  $1$ . So, this way,  $\underline{\mu}_i$  and  $\Sigma_i^*$ , both can be learned during the learning process of other parameters of the proposed network TED-GAN.

### 2.1.3. The Loss Function

To push TED-GAN to produce an image from a particular category, we need to put some extra information or condition to both generator ( $G$ ) and discriminator ( $D$ ). We represent the output of the generator as  $G(z'|l)$  where ' $l$ ' represents the label of the category to push the network to produce the images from a specific class. So, the generator loss  $G_l$  and the discriminator loss  $D_l$  is given in Equations (8) and (9), respectively.

$$G_l = -\mathbb{E}_{z'|l \sim \mathcal{P}_{z'|l}} [\log(D(G(x|l)))] \quad (8)$$

$$D_l = -\mathbb{E}_{x \sim \mathcal{P}_{data}} [\log(D(x))] - \mathbb{E}_{z'|l \sim \mathcal{P}_{z'|l}} [\log(1 - D(G(x|l)))] \quad (9)$$

### 2.1.4. Auxiliary Classifier Loss Function

As the class label information is encoded into the network, we can add an auxiliary classifier to the discriminator of TED-GAN. In this way, we can push the discriminator to do two tasks; identifying whether it is a real or fake image and predicting the class label of the image. Making the network do additional jobs proved good practice for improving network

performance in the basic tasks. In our case, the auxiliary classifier does the additional supervisory job to push the generator and discriminator to produce realistic and diverse skin lesion images.

Our auxiliary classifier  $L_c$  shares the feature extraction layers with the discriminator, and its loss function is given by [43]

$$L_c = -\mathbb{E}_{x \sim \mathcal{P}_{data}} [\log(D(x))] \mathbb{E}_{z' | l \sim \mathcal{P}_{z'|l}} [\log D(G(z'|l))] \quad (10)$$

## 2.2. Experiment Settings

For experiments, we used an i7-6850 processor supported by the Graphical Processing Unit (GeForce Nvidia GTX 1080 GPU), with an operating system, Ubuntu 18.04.2 LTS, installed on it. We wrote the code in python programming language V3.8 with external libraries, including Keras, TensorFlow v2.0.0, Sci-kit-learn, Pandas, NumPy, and Matplotlib.

We used the HAM10000 dataset [39] for skin cancer to evaluate the proposed methodology and conducted extensive experiments. This is a benchmark dataset for skin cancer images that consists of more than 10,000 images of 7 types of skin lesions. We used the four categories, basal cell carcinoma (Bcc, 514 images), benign keratosis (Bkl, 1099 images), melanocytic nevus (Nv, 6705 images), and melanoma (Mel, 1113 images). The other three categories of lesions in this dataset do not have enough images to train a GAN. We split the dataset into training, validation, and test sets with 60%, 20%, and 20% of the data, respectively.

The training set, i.e., 60% of the dataset, is used for VAE and GAN<sub>1</sub> training. After the training of the encoder-decoder network, the network is swapped into the decoder-encoder network. Now the output of the decoder consists of a noise vector, not an image. This noise vector (the output of VAE) that has the information of the image manifold is used in the training of GAN<sub>1</sub>. Additionally, the images generated by GAN<sub>1</sub> are used in the training of GAN<sub>2</sub>. There is no difference in the architectures of the GAN<sub>1</sub> and GAN<sub>2</sub>, except that the GAN<sub>2</sub> has an auxiliary classifier. Moreover, GAN<sub>2</sub> samples the noise vector from the student t-distribution, whereas GAN<sub>1</sub> samples from the pre-trained decoder-encoder latent vector.

The CNN classifier shown in Figure 4 is trained on these images generated by TED-GAN + 60% images of the HAM10000 dataset. This classifier is trained for an equal number of epochs (1000 epochs) for various image datasets generated by other generative models, and classification results are compared. The block diagrams and architecture details of the proposed framework are shown in Figures 1–4, respectively. We used Adam as an optimizer, 0.01 learning rate, batch size 10 with categorical cross-entropy.

To generate the images from other generative models, we used their publicly available code with default parameter settings.

For performance measures, we used precision, recall, average accuracy, and F1-score. The recall is also called sensitivity. Mathematically, sensitivity, specificity, accuracy, and F1-score can be written as:

$$Sensitivity (Recall) = \frac{TP}{TP + FN} \quad (11)$$

$$Specificity = \frac{TN}{TN + FP} \quad (12)$$

$$Accuracy = \frac{TP + TN}{Total} \quad (13)$$

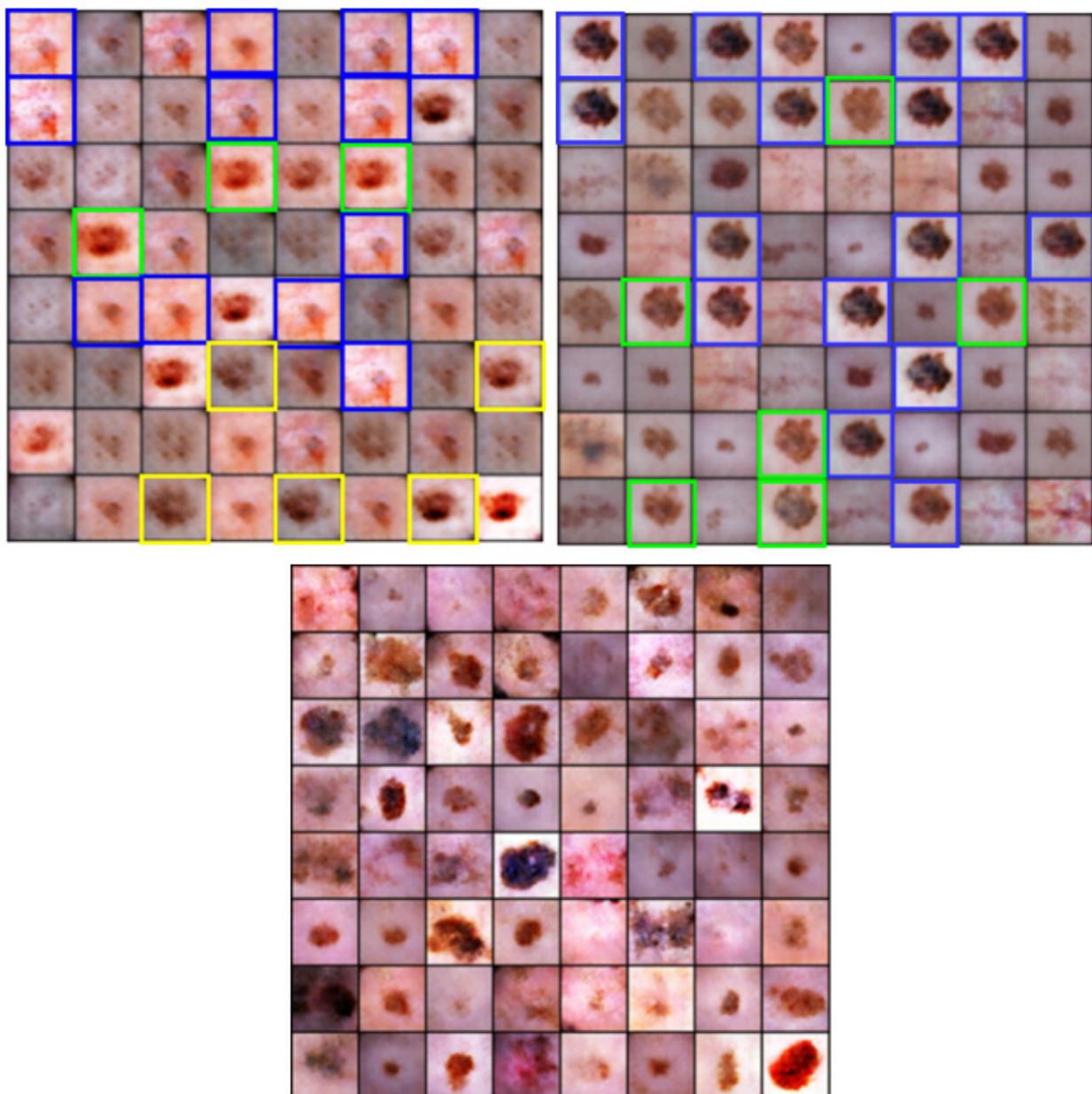
$$F1 \text{ Score} = 2 * \left( \frac{Sensitivity * Specificity}{Sensitivity + Specificity} \right) \quad (14)$$

where:

$FN$  = False Negatives     $TP$  = True Positives  
 $TN$  = True Negatives     $FP$  = False Positives

### 3. Results

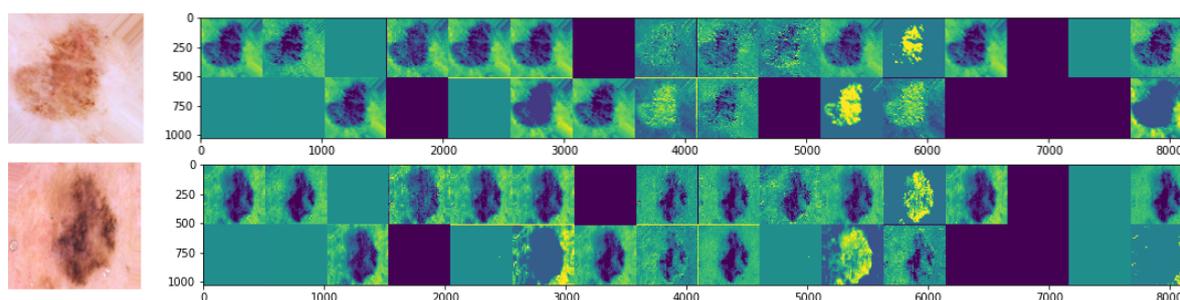
Generating medical images with any GAN is tricky and challenging compared to generating other images, for example, different species images, like dogs, cats, etc. Moreover, using these synthesized images to train the model for disease diagnosis makes this task more crucial. The dataset we used in this study was relatively tiny. We tried to leverage on using the *t*-distribution that generates diverse images because of its fatter tail. These artificially generated and diverse images played an important role in improving the results of deep learning classifiers. The artificial melanoma images generated by various generative models are shown in Figure 5. We can observe that the imperceptibility of images generated by TED-GAN is significantly better than other generative models.



**Figure 5.** Artificially generated images of most swear type of skin cancer, melanoma by GANS. The top left, right, and bottom blocks of images represent the melanoma images generated by GAN, DeLiGAN, and TED-GAN (proposed), respectively. By having a close look at upper two blocks of images, we can observe the repetition of generated images. The green, blue, and yellow blocks contain similar images. Diversity in the images can clearly be observed visually in the proposed block (bottom one).

Moreover, the diversity of generated images by TED-GAN (proposed) can be observed in Figure 5 (bottom row). Other generative models used in this study for comparison, GAN [15], DeLiGAN [28], generated several repetitive images highlighted in green, blue, and yellow color boxes in Figure 5. Although these repetitive images are similar in shape, they still have different textures, color temperatures, and saturation. These properties can be observed by having a close look at those highlighted images.

The first convolution layer of CNN learned similar features from both generated and real images. This phenomenon can be seen in Figure 6, representing the feature visualization of the real and generated image. This could be the reason that a simple GAN also has a significant impact on the classification results, though it generates several repetitive images.



**Figure 6.** Features visualization of original and generated melanoma images. Upper and lower rows (original and generated images, respectively) represent the features that the convolutional layer of the classifier learned during the training process.

A detailed quantitative comparison of average accuracy, F1-score, sensitivity, and specificity of various generative models with the proposed method are summarized in Table 1. Table 1 shows that TED-GAN performed best, followed by DeLiGAN [28] and basic GAN [15]. Some GAN variations like deep convolutional generative adversarial networks (DCGAN) [21] can generate better images, but require significantly large amounts of training data (around 60,000 images for training). So, we could not include it in our comparison.

**Table 1.** Sensitivity, specificity, F1-score, and an average score of individual lesion class for various generative models. Nv, Mel, Bcc, and Bkl are the abbreviations of skin lesions. Please consult Section 4 for more details about the dataset and skin lesion names. Augmentation represents the classic augmentation, e.g., rotation, cropping, scaling, etc. Moreover, artificial images generated by various generative models are included in the training set.

		Nv	Mel	Bcc	Bkl	Avg. Acc
Without Augmentation						
GAN [15]	Precision	0.73	0.88	0.86	0.80	81.0%
	Recall	0.85	0.70	0.83	0.86	
	F1-Score	0.78	0.78	0.85	0.83	
DeLiGAN [28]	Precision	0.76	0.91	0.86	0.85	83.75%
	Recall	0.87	0.73	0.86	0.89	
	F1-Score	0.81	0.81	0.86	0.87	
TED-GAN (Proposed)	Precision	0.87	0.94	0.94	0.84	<b>89.25%</b>
	Recall	0.92	0.82	0.90	0.93	
	F1-Score	0.89	0.88	0.92	0.88	
With Augmentation						
GAN [15]	Precision	81	90	90	81	85.25%
	Recall	88	81	85	87	
	F1-Score	0.84	0.85	0.88	0.84	
DeLiGAN [28]	Precision	85	94	94	83	89.3%
	Recall	90	83	89	92	
	F1-Score	0.87	0.88	0.91	0.87	
TED-GAN (Proposed)	Precision	90	94	93	93	<b>92.5%</b>
	Recall	94	89	92	95	
	F1-Score	0.92	0.91	0.92	0.94	

Avg. Acc = average accuracy; Bold indicates the maximum value achieved in the respective experiment settings.

#### 4. Discussion

Melanoma is the most dangerous type of skin cancer. Dermatologists are always interested in its sensitivity and specificity. A simple CNN classifier trained on the HAM10000 images only without any generative images achieved only 53% and 75% sensitivity and specificity, respectively, for the melanoma class. In contrast, for the same class of melanoma, the sensitivity and specificity results improved to 70% and 88% (Table 1, column 4), respectively, when the images generated by GAN were included in the training of the classifier. Whereas the proposed generative model (TED-GAN) achieved 82% sensitivity and 94% specificity values of melanoma. The classification results improved around 3–5% further for all the models (GAN, DeLiGAN, and TED-GAN) when classic augmentation (cropping, scaling, and rotation) was used along with generative images for the training of the classifier.

Apart from comparing with other image generative methods, we compare our classification results with several other studies published in various journals within the last three years. The average accuracy of the proposed method is better than the others, around 2–7%. The classification results are summarized in Table 2.

**Table 2.** Comparison of proposed TED-GAN classification results with various studies published within the last three years.

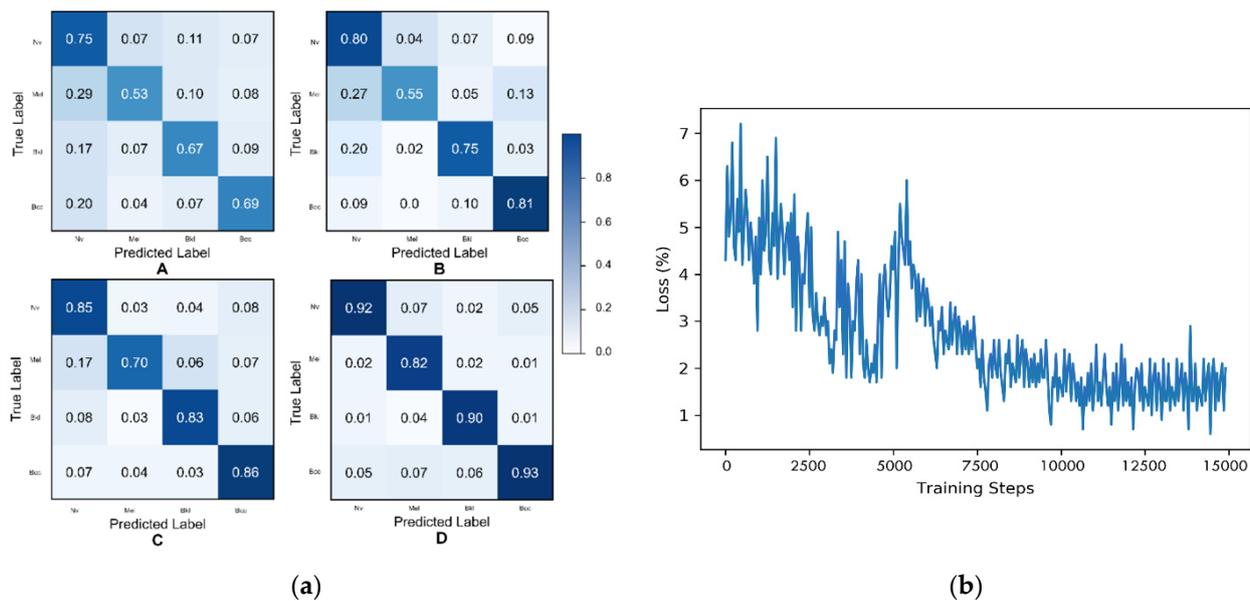
	Source	Dataset	Method	No. of Classes	Sensitivity (Recall)%	Specificity (Precision)%	F1-Score	Accuracy
2019	CNN [44]		CNN	2	92.8	68.2	-	-
	[45]	HAM10000	Physicians	5	66	62	-	-
			CNN	5	86.1	89.2	-	-
2020	[46]	Private	R-CNN	2	-	-	-	86.3
	[47]	HAM10000	GoogLeNet Inception-v3	7	75.57	-	-	-
	[7]	HAM10000	KELM	7	90.2	-	-	90.67
2021	(TED-GAN) Proposed	HAM10000	TED-GAN & CNN	4	89	94	0.91	92.5

Figure 7a represents the confusion matrices of the CNN classifier when it is trained on the images generated by various generative models, including GAN, DeLiGAN, and the proposed TED-GAN.

All the performance measures indicate that using any GAN improves the classification performance of the deep learning model. Additionally, if we use the proposed method to generate the skin lesions, results can significantly be improved because of two reasons.

First, the noise from which the TED-GAN sampled previously is already trained and contains domain information. It also reduces the training time of the TED-GAN considerably. We trained the TED-GAN only for 15,000 training steps, which is half of the time used for the training of other competing GAN models. They trained for 30,000 training steps. Figure 7b depicts the generator loss of TED-GAN. After 10,000 training steps, the average loss value stays around 1.5, which is approximately equal to  $2 \log 2$ , the convergence value suggested by Goodfellow [15].

Secondly, the use of the heavy tail distribution, t-distribution, gives diversity to the generated images. The diversely generated images result in better training of the CNN classifier and prevent it from overfitting on the test set.



**Figure 7.** Confusion matrices of various generative models. (a) (A)—Confusion matrix of the CNN classifier when trained only on HAM10000 images without any generative image. (B)—When classifier training includes the HAM10000 training set and the images generated by GAN [15]. (C)—When classifier includes HAM10000 training set and the images generated by DeLiGAN. (D)— When classifier training includes the HAM10000 training set and the images generated by proposed TED-GAN. (b) Training loss of the generator of TED-GAN ( $GAN_2$ ).

#### 4.1. Limitations

Although we tried to leverage the heavy-tailed distribution that tends to generate diverse images, these generated diverse medical images could be tricky, especially in some particular medical imaging applications like this one. There is a very minute difference in the color and texture of various categories of skin lesions. Diverse image generative models may end up generating skin lesions images that may not belong to that specific lesion category. This is a common limitation of all generative models when they deal with medical imaging.

#### 4.2. Future Work

The performance of the proposed method can further be enhanced easily by using a deep architecture or by making some modifications in the architecture, for example, by using an attention mechanism or by introducing skip connection. We purposely used very simple and lightweight architecture as we had to perform extensive experimentation for a fairly large number of iterations to generate skin lesion images. With a sophisticated deep classifier network along with the proposed TED-GAN, we can further improve the classification results.

Moreover, further experiments can be performed to estimate the classifier's generalization capacity for various unseen datasets. The results can be compared when a small portion of each dataset is included in TED-GAN training and when some of the datasets are unseen.

### 5. Conclusions

This paper proposed a framework that consists of three generative models; one VAE and two GANs. Instead of using random noise for the input of GAN, we trained the VAE to produce informative noise and let the GAN sample the input noise vector from this informative noise. This helped the adversarial network avoid mode collapse, and it converged faster. Moreover, in the second generative network, we used the heavy-tail student t-distribution. This added diversity to the generated images. The proposed

framework improved the classification results of skin lesions from 66% average accuracy to 92.5% average accuracy.

TED-GAN performance was compared with other generative models and existing studies published within the last three years. It successfully achieved a better average accuracy of around 2–7% higher than other generative models.

**Author Contributions:** Conceptualization, B.A.; methodology, B.A.; software, B.A.; validation, B.A., Q.Y., L.M. and M.Z.; formal analysis, B.A.; investigation, B.A.; resources, B.A., S.J. and M.Z.; data curation, B.A., M.Z. and Q.Y.; writing—original draft preparation, B.A.; writing—review and editing, B.A., Q.Y., V.P. and M.Z.; visualization, B.A., V.P., Q.Y. and L.M.; supervision, S.J. and V.P.; project administration, S.J., V.P. and L.M.; funding acquisition, S.J. and L.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work was supported in part by the National Natural Science Foundation of China (Projects Numbers: 61673194, 61672263, 61672265), and in part by the national first-class discipline program of Light Industry Technology and Engineering (Project Number: LITE2018-25).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** We used the pubic dataset available at doi:10.1038/sdata.2018.161.

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

## References

- Zhang, J.; Xie, Y.; Xia, Y.; Shen, C. Attention Residual Learning for Skin Lesion Classification. *IEEE Trans. Med. Imaging* **2019**, *38*, 2092–2103. [[CrossRef](#)] [[PubMed](#)]
- Trinidad, C.M.; Torres-Cabala, C.A.; Curry, J.L.; Prieto, V.G.; Aung, P.P. Update on eighth edition American Joint Committee on Cancer classification for cutaneous melanoma and overview of potential pitfalls in histological examination of staging parameters. *J. Clin. Pathol.* **2019**, *72*, 265–270. [[CrossRef](#)] [[PubMed](#)]
- Kittler, H.; Pehamberger, H.; Wolff, K.; Binder, M. Diagnostic accuracy of dermoscopy. *Lancet Oncol.* **2002**, *3*, 159–165. [[CrossRef](#)]
- Qin, Z.; Liu, Z.; Zhu, P.; Xue, Y. A GAN-based image synthesis method for skin lesion classification. *Comput. Methods Programs Biomed.* **2020**, *195*, 105568. [[CrossRef](#)]
- Sonthalia, S.; Yumeen, S.; Kaliyadan, F. *Dermoscopy Overview and Extradiagnostic Applications*; StatPearls Publishing: Treasure Island, FL, USA, 2020.
- Meng, X.; Chen, J.; Zhang, Z.; Li, K.; Li, J.; Yu, Z.; Zhang, Y. Non-invasive optical methods for melanoma diagnosis. *Photodiagnosis Photodyn. Ther.* **2021**, *34*, 102266. [[CrossRef](#)] [[PubMed](#)]
- Khan, M.A.; Sharif, M.; Akram, T.; Damaševičius, R.; Maskeliūnas, R. Skin Lesion Segmentation and Multiclass Classification Using Deep Learning Features and Improved Moth Flame Optimization. *Diagnostics* **2021**, *11*, 811. [[CrossRef](#)]
- Ahmad, B.; Jun, S. Whether the Health Care Practices For the Patients With Comorbidities Have Changed After the Outbreak of COVID-19; Big Data Public Sentiment Analysis. *arXiv* **2021**, arXiv:2104.12559.
- Qureshi, M.A.; Aziz, A.; Ahmed, B.; Khalid, A.; Munir, H. Comparative Analysis and Implementation of Efficient Digital Image Watermarking Schemes. *Int. J. Comput. Electr. Eng.* **2012**, *4*, 558–561. [[CrossRef](#)]
- Kassem, M.A.; Hosny, K.M.; Damaševičius, R.; Eltoukhy, M.M. Machine Learning and Deep Learning Methods for Skin Lesion Classification and Diagnosis: A Systematic Review. *Diagnostics* **2021**, *11*, 1390. [[CrossRef](#)]
- Ahmad, B.; Hussain, I. Design and hardware implementation of ball & beam setup. In Proceedings of the 2017 Fifth International Conference on Aerospace Science & Engineering (ICASE), Islamabad, Pakistan, 14–16 November 2017; pp. 1–6.
- You, Q.; Sun, J.; Pan, F.; Palade, V.; Ahmad, B. Dmo-qpsso: A multi-objective quantum-behaved particle swarm optimization algorithm based on decomposition with diversity control. *Mathematics* **2021**, *9*, 1959. [[CrossRef](#)]
- Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312. [[CrossRef](#)]
- Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
- Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680. [[CrossRef](#)]
- Shahsavari, A.; Ranjbari, S.; Khatibi, T. Proposing a novel Cascade Ensemble Super Resolution Generative Adversarial Network (CESR-GAN) method for the reconstruction of super-resolution skin lesion images. *Inform. Med. Unlocked* **2021**, *24*, 100628. [[CrossRef](#)]
- Shen, T.; Hao, K.; Gou, C.; Wang, F.-Y. Mass Image Synthesis in Mammogram with Contextual Information Based on GANs. *Comput. Methods Programs Biomed.* **2021**, *202*, 106019. [[CrossRef](#)] [[PubMed](#)]

18. Fernández Blanco, R.; Rosado, P.; Vegas, E.; Reverter, F. Medical image editing in the latent space of Generative Adversarial Networks. *Intell. Med.* **2021**, *5*, 100040. [[CrossRef](#)]
19. Zhang, J.; Yu, L.; Chen, D.; Pan, W.; Shi, C.; Niu, Y.; Yao, X.; Xu, X.; Cheng, Y. Dense GAN and multi-layer attention based lesion segmentation method for COVID-19 CT images. *Biomed. Signal Process. Control.* **2021**, *69*, 102901. [[CrossRef](#)]
20. Pang, T.; Wong, J.H.D.; Ng, W.L.; Chan, C.S. Semi-supervised GAN-based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification. *Comput. Methods Programs Biomed.* **2021**, *203*, 106018. [[CrossRef](#)]
21. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Proceedings of the 4th International Conference on Learning Representations, ICLR 2016—Conference Track Proceedings, San Juan, PR, USA, 2–4 May 2016.
22. Callealta Barroso, F.J.; García-Pérez, C.; Prieto-Alaiz, M. Modelling income distribution using the log Student's t distribution: New evidence for European Union countries. *Econ. Model.* **2020**, *89*, 512–522. [[CrossRef](#)]
23. Van Der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2625.
24. Perarnau, G.; van de Weijer, J.; Raducanu, B.; Álvarez, J.M. Invertible Conditional GANs for image editing. *arXiv* **2016**, arXiv:1611.06355.
25. Saxena, D.; Cao, J. Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions. *ACM Comput. Surv.* **2020**, *54*, 1–42. [[CrossRef](#)]
26. Hays, J.; Efros, A.A. Scene completion using millions of photographs. *ACM Trans. Graph.* **2007**, *26*, 4. [[CrossRef](#)]
27. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. In Proceedings of the 2nd International Conference on Learning Representations, ICLR 2014—Conference Track Proceedings, Banff, AB, Canada, 14–16 April 2014; pp. 1–14.
28. Gurumurthy, S.; Sarvadevabhatla, R.K.; Babu, R.V. DeLiGAN: Generative adversarial networks for diverse and limited data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 166–174. [[CrossRef](#)]
29. Ma, Y.; Zhong, G.; Wang, Y.; Liu, W. MetaCGAN: A Novel GAN Model for Generating High Quality and Diversity Images with Few Training Data. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7.
30. Van Den Oord, A.; Kalchbrenner, N.; Kavukcuoglu, K. Pixel recurrent neural networks. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; Volume 4, pp. 2611–2620.
31. Van Den Oord, A.; Kalchbrenner, N.; Vinyals, O.; Espeholt, L.; Graves, A.; Kavukcuoglu, K. Conditional image generation with PixelCNN decoders. *Adv. Neural Inf. Process. Syst.* **2016**, arXiv:1606.05328, 4797–4805.
32. Liu, Y.; Chen, A.; Shi, H.; Huang, S.; Zheng, W.; Liu, Z.; Zhang, Q.; Yang, X. CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. *Comput. Med. Imaging Graph.* **2021**, *91*, 101953. [[CrossRef](#)] [[PubMed](#)]
33. Jiang, M.; Zhi, M.; Wei, L.; Yang, X.; Zhang, J.; Li, Y.; Wang, P.; Huang, J.; Yang, G. FA-GAN: Fused Attentive Generative Adversarial Networks for MRI Image Super-Resolution. *Comput. Med. Imaging Graph.* **2021**, *92*, 101969. [[CrossRef](#)]
34. Mahapatra, D.; Bozorgtabar, B.; Hewavitharanage, S.; Garnavi, R. Image Super Resolution Using Generative Adversarial Networks and Local Saliency Maps for Retinal Image Analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Metzler, J.B., Ed.; Springer: Cham, Switzerland, 2017; pp. 382–390.
35. Shin, H.-C.; Tenenholtz, N.A.; Rogers, J.K.; Schwarz, C.G.; Senjem, M.L.; Gunter, J.L.; Andriole, K.P.; Michalski, M. Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks. In *International Workshop on Simulation and Synthesis in Medical Imaging*; Springer: Cham, Switzerland, 2018; pp. 1–11.
36. Han, C.; Rundo, L.; Araki, R.; Nagano, Y.; Furukawa, Y.; Mauri, G.; Nakayama, H.; Hayashi, H. Combining Noise-to-Image and Image-to-Image GANs: Brain MR Image Augmentation for Tumor Detection. *IEEE Access* **2019**, *7*, 156966–156977. [[CrossRef](#)]
37. Nie, D.; Trullo, R.; Lian, J.; Wang, L.; Petitjean, C.; Ruan, S.; Wang, Q.; Shen, D. Medical Image Synthesis with Deep Convolutional Adversarial Networks. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 2720–2730. [[CrossRef](#)]
38. Zhong, G.; Gao, W.; Liu, Y.; Yang, Y.; Wang, D.H.; Huang, K. Generative adversarial networks with decoder–encoder output noises. *Neural Netw.* **2020**, *127*, 19–28. [[CrossRef](#)]
39. Tschandl, P.; Rosendahl, C.; Kittler, H. Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 1–9. [[CrossRef](#)]
40. Wei, R.; Garcia, C.; El-Sayed, A.; Peterson, V.; Mahmood, A. Variations in Variational Autoencoders—A Comparative Evaluation. *IEEE Access* **2020**, *8*, 153651–153670. [[CrossRef](#)]
41. Joarder, A. A short review of multivariate t-distribution a short review of multivariate t-distribution. *J. Stat. Res.* **2014**, *40*, 59–72.
42. Sun, J.; Zhong, G.; Chen, Y.; Liu, Y.; Li, T.; Huang, K. Generative adversarial networks with mixture of t-distributions noise for diverse image generation. *Neural Netw.* **2020**, *122*, 374–381. [[CrossRef](#)] [[PubMed](#)]
43. Ma, D.; Liu, B.; Kang, Z.; Zhou, J.; Zhu, J.; Xu, Z. Two birds with one stone: Transforming and generating facial images with iterative GAN. *Neurocomputing* **2020**, *396*, 278–290. [[CrossRef](#)]
44. Brinker, T.J.; Hekler, A.; Enk, A.H.; Klode, J.; Hauschild, A.; Berking, C.; Schilling, B.; Haferkamp, S.; Schadendorf, D.; Fröhling, S.; et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur. J. Cancer* **2019**, *111*, 148–154. [[CrossRef](#)] [[PubMed](#)]

45. Hekler, A.; Utikal, J.S.; Enk, A.H.; Hauschild, A.; Weichenthal, M.; Maron, R.C.; Berking, C.; Haferkamp, S.; Klode, J.; Schadendorf, D.; et al. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur. J. Cancer* **2019**, *120*, 114–121. [[CrossRef](#)] [[PubMed](#)]
46. Hartanto, C.A.; Wibowo, A. Development of Mobile Skin Cancer Detection using Faster R-CNN and MobileNet v2 Model. In Proceedings of the 2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, Indonesia, 24–25 September 2020; pp. 58–63.
47. Harangi, B.; Baran, A.; Hajdu, A. Assisted deep learning framework for multi-class skin lesion classification considering a binary classification support. *Biomed. Signal Process. Control* **2020**, *62*, 102041. [[CrossRef](#)]