An exploratory analysis of data augmentation techniques

Mariana Chaves[1]

Université Côte d'Azur `mariana.chaves-espinoza@etu.univ-cotedazur.fr`

1 Introduction

In this study, we explore some data augmentation techniques to compare their impact on the accuracy of an image classification model. Data augmentation, as its name suggests, comprehends techniques used to increase the amount of data by adding modified copies of the original data or by creating synthetic data. It aims, not only to increase the size of the data to feed data-hungry models but also to enhance the quality of the training dataset since is it used to avoid overfitting [4].

There are many techniques, among them we can mention geometric transformations, color space augmentations, kernel filters, mixing images, random erasing, feature space augmentation, adversarial training, generative adversarial networks, and many others. We focus on four:

- Deep convolutional generative adversarial networks (DCGAN)
- Denoising autoencoder (DEA)
- Convolutional variational autoencoder (CVAE)
- Geometric transformations

2 Materials and methods

2.1 Datasets

We choose four datasets from the MedMNIST [5] collection:

- PneumoniaMNIST: it includes 5856 pediatric chest X-Ray black and white images. The task is a binary-class classification of pneumonia against normal.
- BreastMNIST: it contains 780 breast ultrasound black and white images. The task is binary-class of normal/benign against malignant.
- BloodMNIST: this dataset is based on individual normal blood cells, captured from individuals without infection, hematologic or oncologic disease, and free of any pharmacologic treatment at the moment of blood collection. It contains a total of 17092 RGB images and is organized into 8 classes.
- DermaMNIST: it contains 10015 dermatoscopic images of common pigmented skin lesions. It is categorized into 7 different diseases.

2.2 Data augmentation techniques

The first data augmentation technique that we apply is Deep convolutional generative adversarial networks (DCGAN) [3], which follows the same idea as a regular GAN, that is, it trains two networks: a generator and a discriminator. The generator learns to create images that look real, while a discriminator learns to tell real images apart from generated ones. It differs from the original GAN since it uses CNNs instead of only dense layers.

Secondly, we apply a Denoising autoencoder. Autoencoders englobe an encoder and a decoder. The encoder reduces the dimensionality of the data, that is, it brings it to a latent space. Then the decoder takes the latent representation and tries to reconstruct the data, such that the output of the decoder is as similar as possible to the original data. Denoising autoencoders use corrupted data points as input and train to predict the original. In our case, the corrupted data points are images with Gaussian noise. The DEA trains to reconstruct the original, not noisy, images.

The third technique is a Convolutional variational autoencoder (CVAE) [2]. In the same spirit as the autoencoder, an encoder network reduces the data to a latent representation. Nevertheless, in this case, the VAE maps the input data into the parameters of a probability distribution, such as the mean and variance of a Gaussian distribution, as in our code. Instead of forwarding the latent values to the decoder directly, VAEs use them to calculate the mean and the standard deviation. The input to the decoder is then sampled from the corresponding normal distribution.

Finally, we experiment with geometric data augmentation. There are many possible geometric transformations such as flipping, cropping, rotation, noise injection, and others. We implement flippling of the images and then a rotation (which can be of 0, 90, 180, or 270 degrees).

In all cases, we increase the training set by 50%. Therefore, the augmented data contains one-third or generated new images. We take care of incrementing the data proportionally to the class size. In other words, if in the original data a class represented 30% of the images, the augmented data keep that proportion. For the data augmentation using DCGANs, a DCGAN was trained for each class in each dataset. The same strategy was used for the CVAE models.

2.3 The classifier

We use Autokeras [1] as classifier. It was one of the classifiers used by the authors of MedMNIST on the different datasets. AutoKeras is an AutoML tool for deep learning models. It performs a search on Keras models via TensorFlow.

2.4 Code

The code can be found at `https://github.com/m-chaves/data_augmentation_with_generative_methods`

3 Results

In this section we examine the results for each of the datasets.

3.1 PneumoniaMNIST

This data involved a binary classification task for black and white images. In Figure 1a, we see somes examples for each class. Using the DCGAN (Figure 1b), the generated images for each class manage to emulate the originals. The DAE also provides good results (Figure 2). We can see that the reconstructions resemble the original data. The Convolutional Variational Autoencoder partially achieves good results (Figure 1c). The generated images resemble the originals. Nevertheless, the patterns for each class are not as evident as in the true data. For instance, some of the images that were generated for the class 'normal' could be attributed to the 'pneumonia' class. In Figure 3 we see some examples of the geometric data augmentation. In general none of the techniques helped on improving the accuracy (Table 1). The test accuracy even decreased in some cases due to the data augmentation.

3.2 BreastMNIST

This dataset also involved a binary classification task. The key aspect of this data is that it was the smallest, with only 780 images in total (for training, validation and testing). Also, we can suspect that the classification could be harder. In Figure 4a, we see some examples of its images. For a non-trained eye, it is not that obvious which image would correspond to each class. Figure 4b shows some examples of the images generated by the DCGAN for each class. We observe that the results are not as good as in the previous dataset. One of the reasons for this outcome can be the small size of the dataset. The results for the DAE are not particularlly good either (Figure 5). The reconstructions contain to much blurr. We observe even worse results for the convolutional variational autoencoder (Figure 4c). None of the techniques improved the accuracy of the model (Table 1). Interestingly, the training accuracy increased when using DCGAN as data augmentation technique, which can indicate the tendency of the dataset to overfit then using this data augmentation tool.

3.3 BloodMNIST

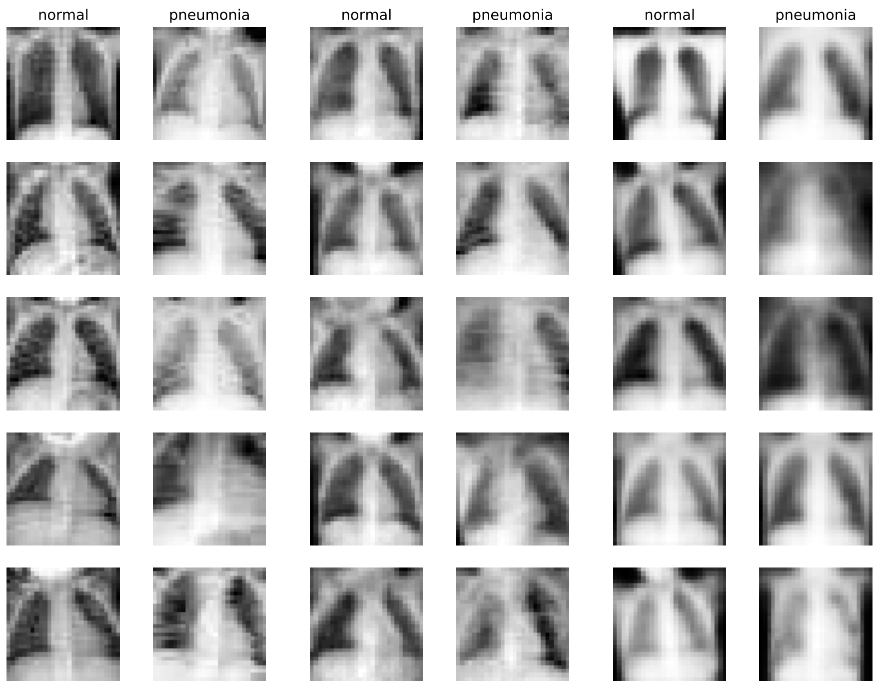
Now we present the results for one of the multi-class classification task for 3 channel images. The DCGAN manages to provide good generated images for each of the classes (Figure 6b). The denoising autoencoder also shows good reconstructions of the images (Figure 7). The variational autoencoder does not shows so good results (Figure 6c). In particular it does not capture correctly the background, and its representations of the classes are too blurry. Once more, none of the techniques aids in improving the accuracy (Table 1). Surprisingly, the CVAE is the only one that increases slightly the validation and test accuracy. Nevertheless, this improvement is not big enough to determine the CVAE as a clear winner.

3.4 DermaMNIST

Finally, we explore the DermaMNIST data. This is also a multi-class collection. For the non-trained eye, the classes are not obvious (Figure 8a). The DCGAN struggles to generate data for some classes and thrives in other ones (Figure 8b). For instance, the images generated for the "melanocytic nevi" class resemble original images, while the "dermatofibroma" examples totally fail to represent the class. The denoising autoencoder produces good reconstructed images (Figure 9). The CVAE, similar to the DCGAN, thrives in the representation of some classes, while it fails for others (Figure 8c). Given that the classes with good results are the same as in the DCGAN, we can assume that this outcome is a result of the size of the class. The classes with more examples are able to train better DCGANs and CVAEs. Once more, none of the techniques helped to increase the accuracy (Table 1).

Dataset	Train set	Acc. Train	Acc. Val	Acc. Test
PneumoniaMNIST	Original Data	0.9819	0.9637	0.8317
	DCGAN	0.9878	0.9637	0.8349
	DAE	0.8994	0.9656	0.8221
	CVAE	0.8964	0.9561	0.8060
	Geometric Augment.	0.8956	0.9522	0.8060
BreastMNIST	Original Data	0.8717	0.8333	0.8333
	DCGAN	0.9449	0.8589	0.8205
	DAE	0.8353	0.8205	0.8269
	CVAE	0.8317	0.8589	0.8333
	Geometric Augment.	0.8353	0.8589	0.8333
BloodMNIST	Original Data	0.9764	0.9287	0.9257
	DCGAN	0.9692	0.9094	0.9000
	DAE	0.7043	0.9264	0.9199
	CVAE	0.7095	0.9334	0.9280
	Geometric Augment.	0.7056	0.9316	0.9228
DermaMNIST	Original Data	0.7884	0.7467	0.7436
	DCGAN	0.8781	0.7417	0.7336
	DAE	0.7469	0.7427	0.7421
	CVAE	0.7486	0.7407	0.7276
	Geometric Augment.	0.7463	0.7357	0.7216

Table 1: Autokeras accuracy results using the different data augmentation techniques in the four datasets.



(a) True images (b) DCGAN images (c) CVAE images

Fig. 1: PneumoniaMNIST images according to class. For left to right: Examples of true images, images generated using the DCGAN, images generated using the CVAE.

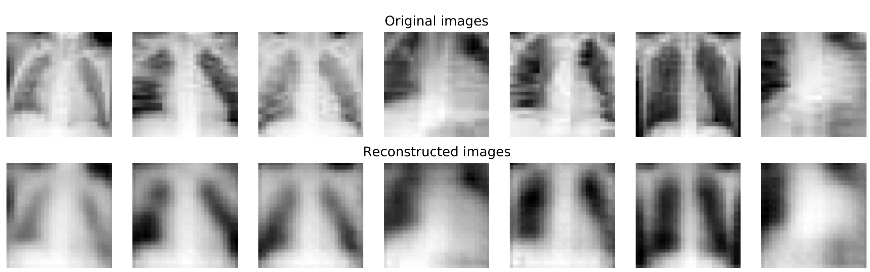


Fig. 2: Original PneumoniaMNIST images and their reconstructed image using DAE.

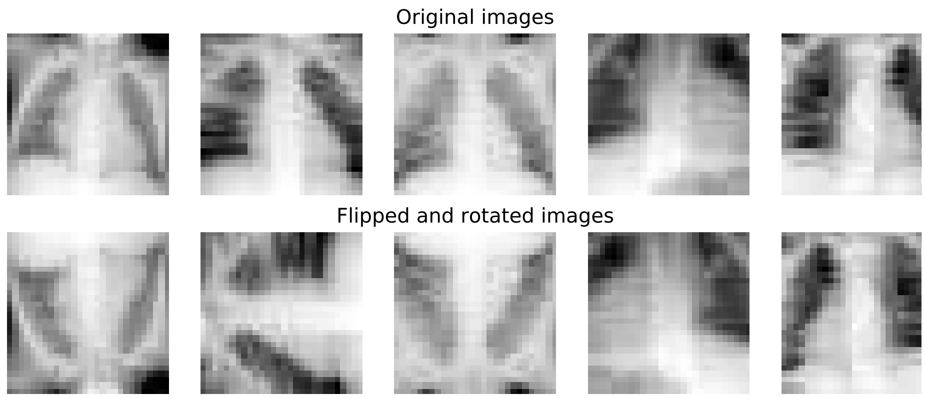
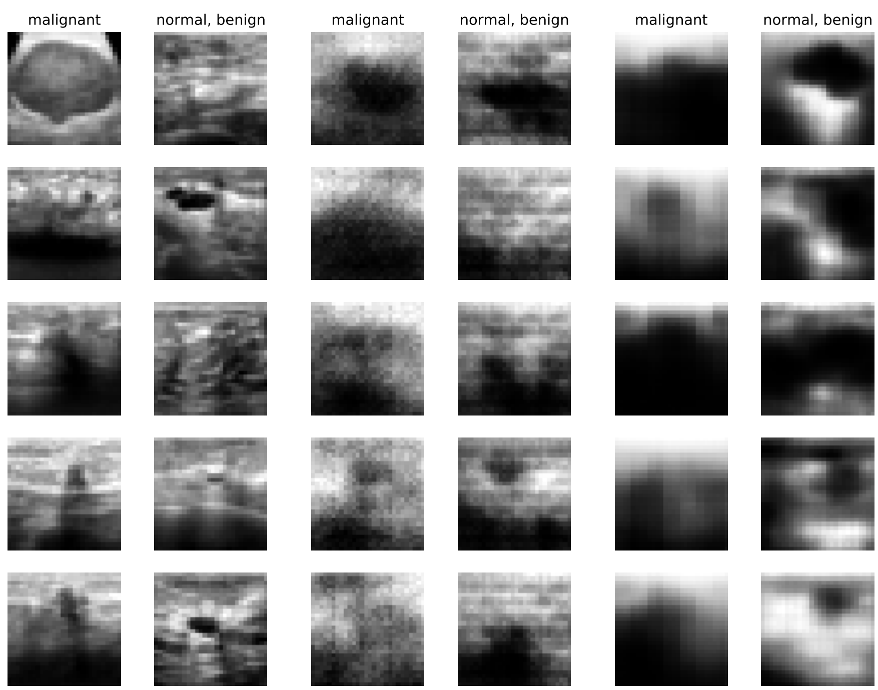


Fig. 3: Original PneumoniaMNIST images and their modified versions using geometric data augmentation techniques.



(a) True images (b) DCGAN images (c) CVAE images

Fig. 4: BreastMNIST images according to class. For left to right: Examples of true images, images generated using the DCGAN, images generated using the CVAE.

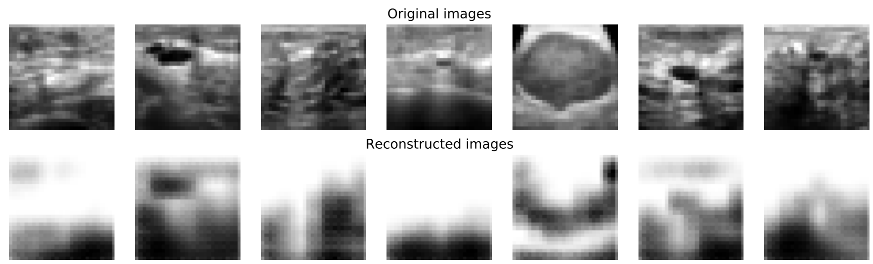


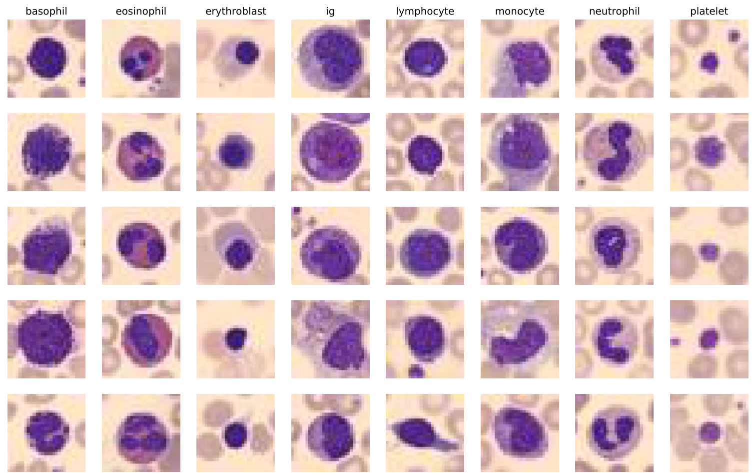
Fig. 5: Original BreastMNIST images and their reconstructed image using DAE.

4 Conclusion

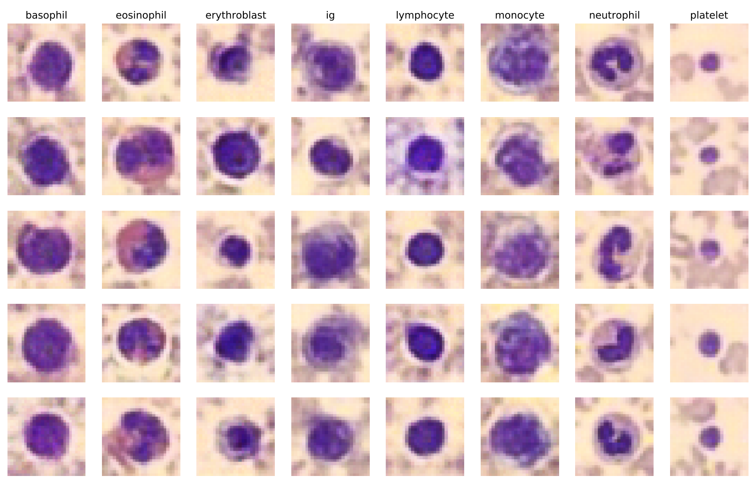
We used four data augmentation techniques with the aim of increasing the quality of our training set and observing better accuracy results on an image classifier. For all the datasets, none of the techniques helped to improve the accuracy of the model. There can be several causes for this outcome. First, in some cases, the data generation techniques do not manage to emulate the true data. This is the case when the data augmentation technique has little data to train. In other cases, the dataset could be particularly complex, and more robust architectures could be necessary. Second, in all the cases, we augmented the data on 50%, perhaps more generated data it's necessary.

References

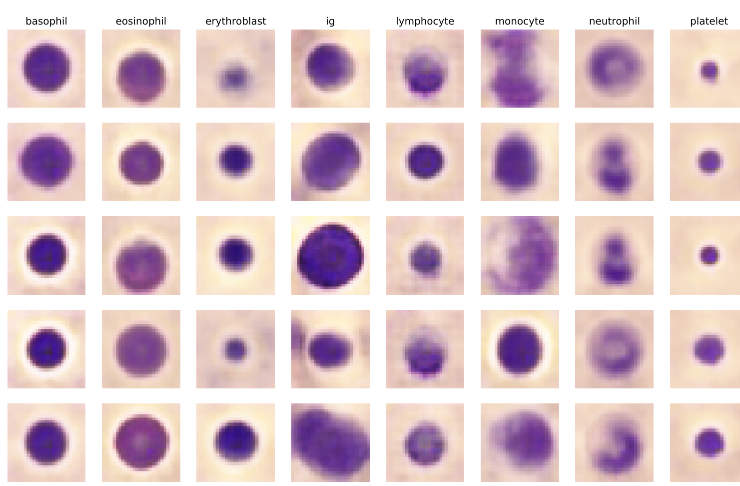
1. Jin, H., Song, Q., Hu, X.: Auto-keras: An efficient neural architecture search system. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 1946–1956 (2019)
2. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
3. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
4. Shorten, C., Khoshgoftaar, T.M.: A survey on Image Data Augmentation for Deep Learning. Journal of Big Data **6**(1), 60 (Jul 2019). https://doi.org/10.1186/s40537-019-0197-0, `https://doi.org/10.1186/s40537-019-0197-0`
5. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification. arXiv preprint arXiv:2110.14795 (2021)



(a) True images



(b) DCGAN images



(c) CVAE images

Fig. 6: BloodMNIST images according to class. For left to right: Examples of true images, images generated using the DCGAN, images generated using the CVAE.

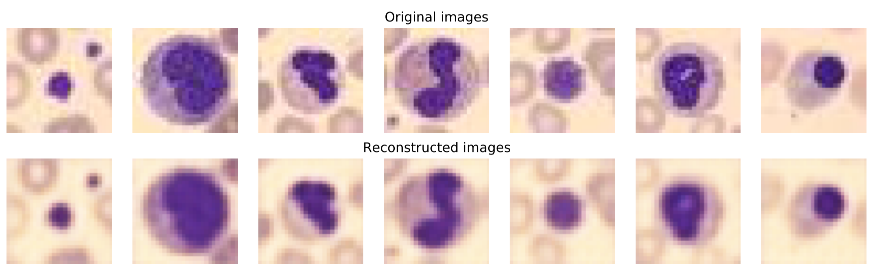
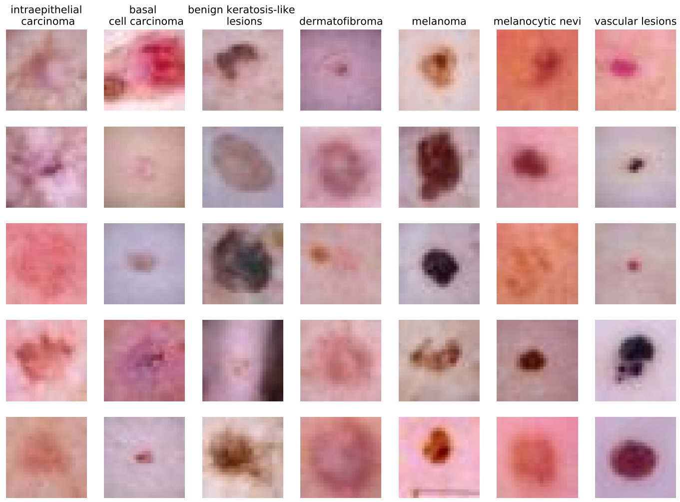
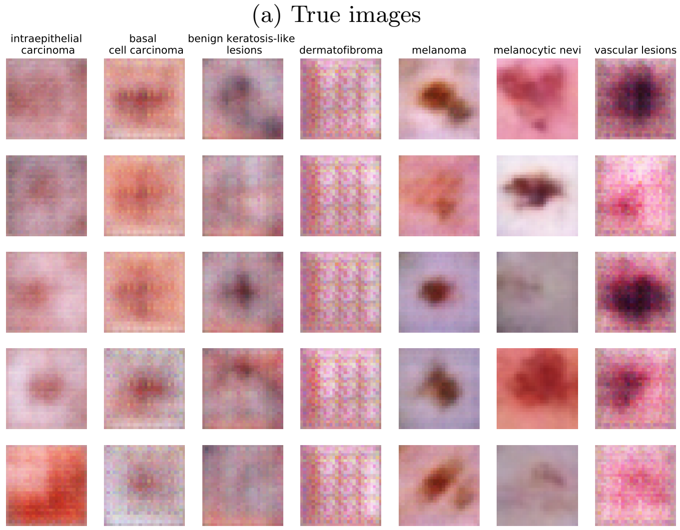


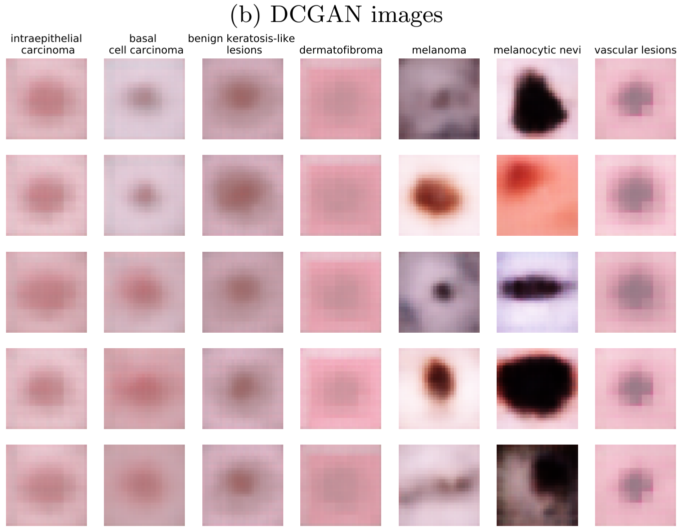
Fig. 7: Original BloodMNIST images and their reconstructed image using DAE.



(a) True images



(b) DCGAN images



(c) CVAE images

Fig. 8: DermaMNIST images according to class. For left to right: Examples of true images, images generated using the DCGAN, images generated using the CVAE.

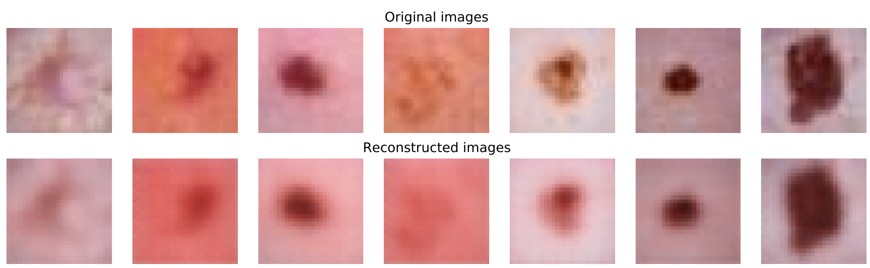


Fig. 9: Original DermaMNIST images and their reconstructed image using DAE.