

Deep Convolutional Generative Adversarial Networks to Enhance Artificial Intelligence in Healthcare: A Skin Cancer Application

AI VIETNAM - RESEARCH TEAM

October 8, 2023

Date of publication:	08/17/2022
Authors:	Marco La Salvia , Emanuele Torti, Raquel Leon , Himar Fabelo , Samuel Ortega , Beatriz Martinez-Vega , Gustavo M.Callico and Francesco Leporati
Sources:	2022 MDPI
Keywords:	Deep learning, hyperspectral imaging, medical hyperspectral images, synthetic data generation, deep convolutional generative adversarial networks.
Summary by:	Quy Nguyen Dinh

1. Purpose/outputs:

- The paper applies an extended DCGAN architecture for medical hyperspectral images, a novel and underexplored field.
- The paper uses transfer learning technique to initialize the weights for DCGAN from a larger RGB dataset as well as label smoothing technique.
- The paper transforms DCGAN into a conditional network (cGAN) to be able to generate synthetic images according to the desired label.
- The paper uses a ResNet18 network architecture to test the similarity between the synthetic and real data.

2. Contributions:

- The paper applies an extended DCGAN architecture for medical hyperspectral images, a novel and underexplored field.
- The paper uses transfer learning technique to initialize the weights for DCGAN from a larger RGB dataset as well as label smoothing technique.
- The paper transforms DCGAN into a conditional network (cGAN) to be able to generate synthetic images according to the desired label.
- The paper uses a ResNet18 network architecture to test the similarity between the synthetic and real data.

3. Inputs:

- The paper uses a medical in vivo hyperspectral image dataset, consisting of 76 hyperspectral images of 40 benign and 36 malignant skin lesions, taken from different body parts from 61 patients. The dataset was processed and labeled into two classes: benign and malignant. The dataset also includes RGB and monochrome images of the same skin lesion, taken with different cameras. This dataset is not large, having only 76 images, so the authors use DCGAN to generate synthetic hyperspectral images to train a deep classifier.
- Each image has a spatial resolution of 50×50 pixels and a spectral resolution of 8 nm, covering 125 spectral bands from 450 to 950 nm (Figure 1a). In addition, the camera also integrates a monochrome sensor capable of capturing the same scene with a conventional monochrome image with a resolution of 1000×1000 pixels (Figure 1b). Besides the hyperspectral (HS) and monochrome images, conventional RGB images of 3000×4000 pixels of the same skin lesion were also captured (Figure 1c).
- The HS images were preprocessed and calibrated with white and dark references to standardize the spectral signatures and remove the first five and the last four bands due to high noise, resulting in a final size of 50×50 pixels and 116 bands.
- For labeling, the authors use label smoothing technique. The malignant lesions are assigned the positive class, while the benign ones are assigned zero. The authors replace the positive class with a random value ranging from 0.7 to 1 and the other with random values from 0 to 0.3.

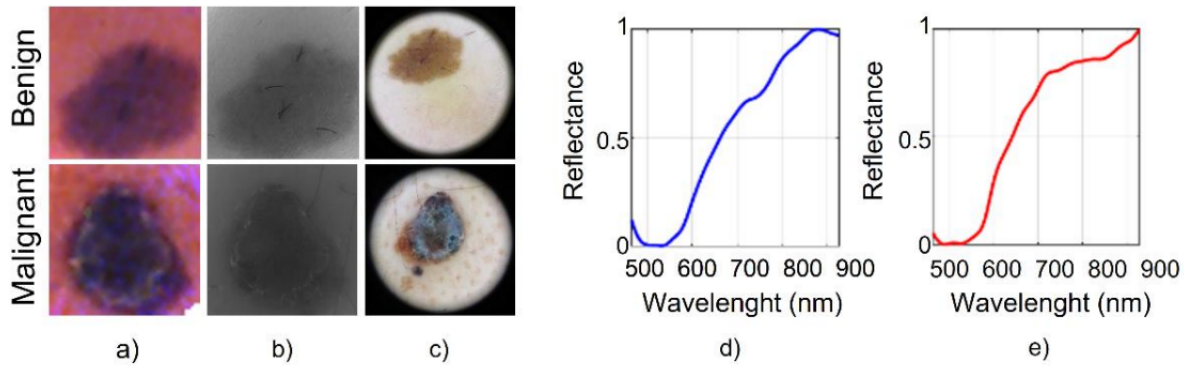


Figure 1: Example images from the dataset. (a) Pseudo-RGB images obtained with the HS camera, (b) grayscale image captured with the monochrome sensor, (c) RGB images obtained with the digital dermoscopic camera, (d) average spectral signature of benign lesion, (e) average spectral signature of malignant lesion.

4. Methodology:

- (a) **Deep convolutional generative adversarial networks:** The GAN model was proposed for the first time in 2014 and is based on two subnetworks: a generator (G) and a discriminator (D). Figure 2 illustrates the basic idea behind GAN. The generator G takes as input a latent space vector z from a standard Gaussian distribution and produces a sample $G(z)$. This sample represents the mapping from the latent space z to the real data space. On one hand, G is optimized to estimate the training data distribution and generate fake samples with the same distribution as the real data. On the other hand, D receives as input the fake data

produced by G or a sample (x) from the real dataset. D outputs the probability estimate of the input data source. Specifically, it estimates whether the sample came from the training data or from G . G tries to minimize the probability that D will predict its outputs as fake, while D tries to maximize its probability to correctly discriminate between real and fake data.

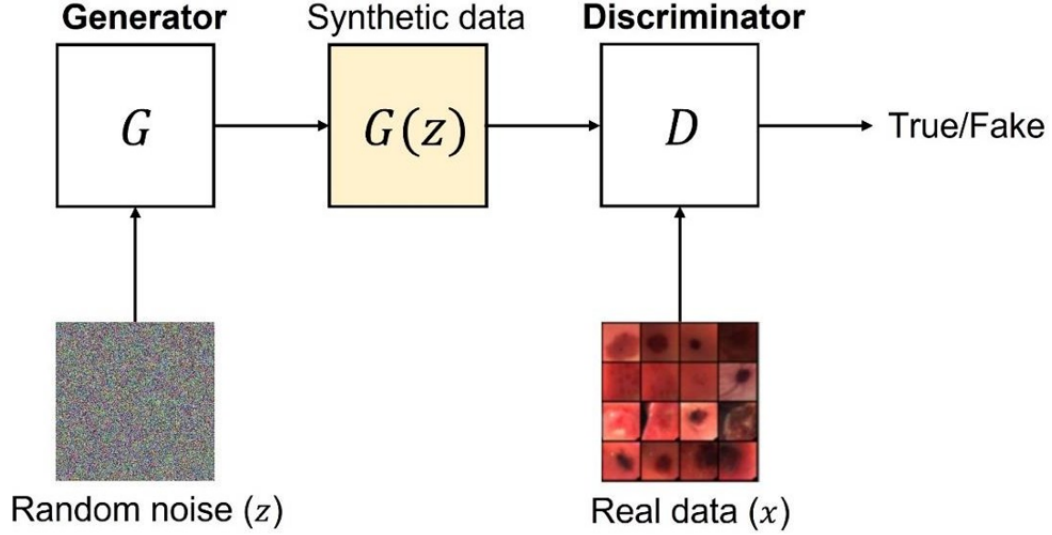


Figure 2: Basic GAN structure

In recent years, deep convolutional generative adversarial networks have emerged as a stable and affordable architecture for generating fake images. Regarding HS images, the conversion from z to the data space performed by G consists of creating fake HS images with the same spatial and spectral dimensions as the training images. Since the authors use the skin cancer dataset described in Section 3 as the training set, G will generate an image with dimensions of $50 \times 50 \times 116$. Figure 3a shows the architecture of G . The deconvolutional layers from 1 to 6 apply batch normalization and ReLU activation function. The last deconvolutional layer uses tanh activation function. On the other hand, D receives as input an HS image with the same size of $50 \times 50 \times 116$, and performs binary classification to determine whether the input image is real or fake. Figure 3b shows the architecture of D . The first convolutional layer uses leaky ReLU activation function. The layers from 2 to 5 apply batch normalization and leaky ReLU activation function. All leaky ReLU functions have a negative slope of 0.2. The final convolutional layer uses sigmoid function.

(b) **Transfer learning:**

The transfer learning method consists of pretraining GAN using RGB skin cancer images and using the obtained parameters as initialization for the final model, which is trained using the HS dataset. Therefore, the authors trained the initial model using HAM10000 dataset, randomly selecting 5000 RGB images from the database. The authors resized the images to 50×50 pixels to have the same spatial dimension as HS dataset. Moreover, they also modified the output layer of G and input layer of D to handle 3 channels instead of 116.

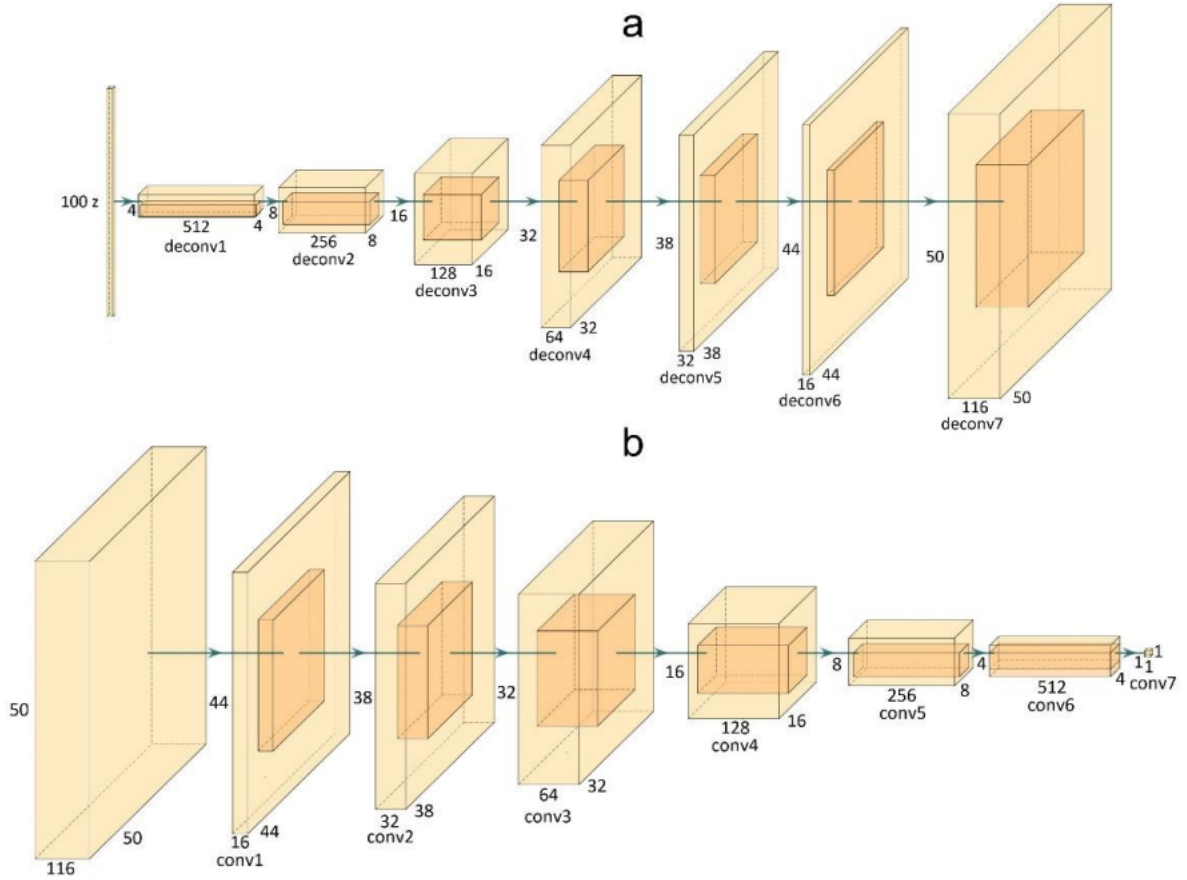


Figure 3: Generator (a) and discriminator (b) architectures

The authors used Adam optimization algorithm, with learning rate of 0.0002 for both networks, and batch size of 128. All hyperparameters were chosen by trial and error, repeating training process with different values. Training process ended after 100 epochs. Finally, they used label swapping technique to avoid overfitting of D network. Figure 4 shows some images taken from original dataset and different images generated by model.

The authors transferred the network weights obtained at the end of this training process to the architectures described in Figure 3 and changed the output layer of G and the input layer of D to the size of 116; therefore, the values obtained by training with the RGB dataset were used to initialize the weights related to the channels corresponding to the green, red, and blue wavelengths. The remaining values were initialized randomly. In this phase, the batch size was reduced to 2. Moreover, the authors changed the size of the output layer of G from 116 to 117. The new channel was used to generate the segmentation mask related to the synthetic image. The mask generation is very important because it includes information that can be used in the training process of a generic segmentation network, highlighting the contours of the lesion. Finally, the proposed architecture was transformed into a conditional GAN (cGAN). It means that G receives as input, along with the random noise vector, the smoothed label value of the class that the synthetic image belongs to. In other words, G can generate fake data related to benign or malignant classes. The architecture of the proposed cGAN is shown in Figure 5.

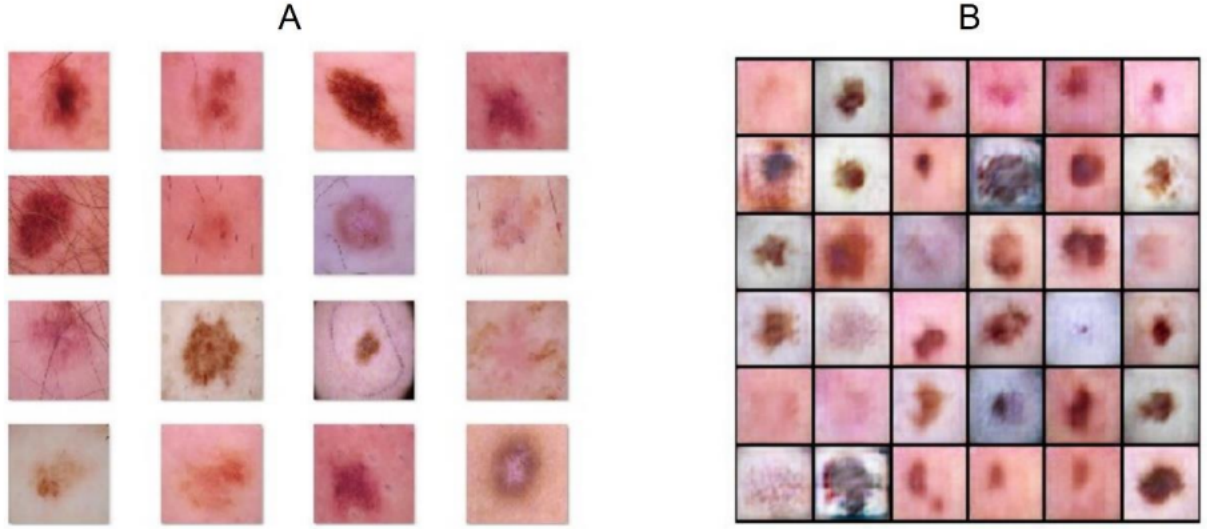


Figure 4: (A) Images taken from the training set, (B) Images generated by the model

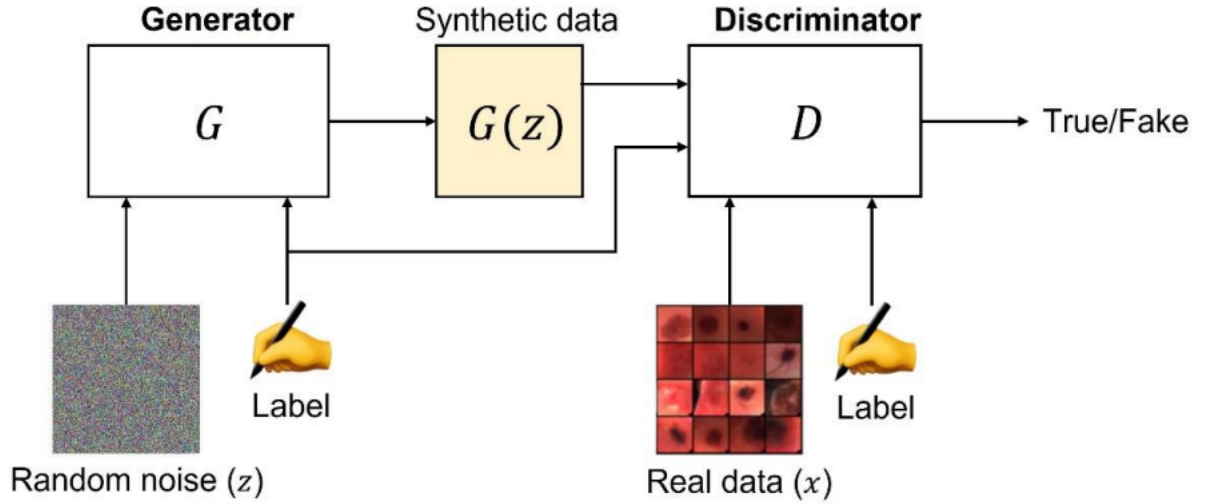


Figure 5: cGAN architecture

The authors trained cGAN for 200 epochs. During training, they used different methods to improve the quality of the synthetic images. The weights of each layer were multiplied by a factor c according to the equalized learning rate rule. Moreover, they applied the two time-scale update rule (TTUR). Specifically, they assigned different learning rate values to the two networks. The learning rate of G was lower than D . Therefore, the weights related to G were updated with more steps than D , to enhance the quality of the synthetic images. To avoid D learning how to discriminate real and fake images in a few training iterations, they swapped labels for 5% random training data. In fact, they treated some fake images as real and vice versa. Finally, they used $L2$ normalization at 10^{-5} level to reduce overfitting.

(c) **ResNet18 classification model:**

The authors used a ResNet18 network to measure similarity between real and fake HS data. Specifically, they only trained this architecture with synthetic HS images to classify real epidermal lesions. Therefore, they exploited overfitting as a method to show how well synthetic data reproduces real statistical distribution. This method indicates whether or not fake dataset represents meaning of real dataset. In this case, overfitting should not be considered a negative impact; overfitting on fake dataset and achieving good results in classifying real dataset means that obtained model has generalized problem under consideration.

In Figure 6, blue arrows indicate that set was used to train model, while green arrow indicates that dataset is used as input for classification.

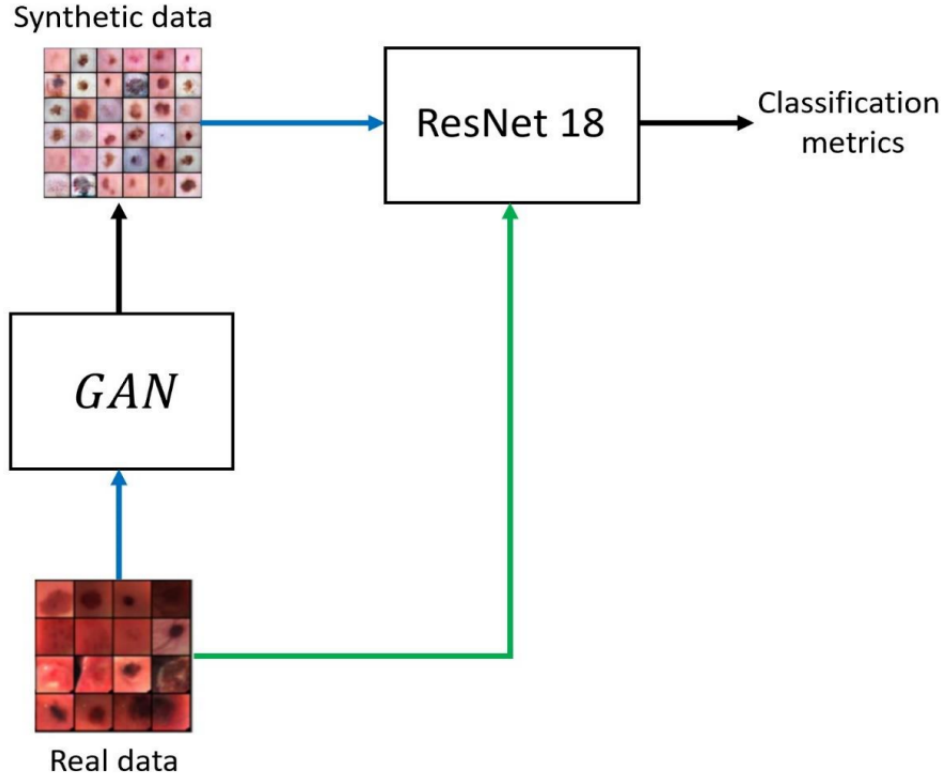


Figure 6: Proposed method for evaluating similarity between datasets

(d) **Evaluation methods:**

FID (Frechét inception distance): Measures the distance between the feature vectors of real and fake images. The lower the FID value, the more similar the two sets of images are.

$$FID = \|\mu_1 - \mu_2\|^2 + \text{Tr}(\Sigma_1 + \Sigma_2 - 2\sqrt{\Sigma_1 \Sigma_2})$$

Where μ represents the mean value, Σ is the covariance matrix and Tr denotes the trace of a matrix. The subscripts 1 and 2 indicate the real and fake images sets, respectively.

Accuracy, precision, recall, F1 score: Measure the performance of the ResNet18 network in classifying real images based on fake data. These metrics are high, indicating that the fake data have good quality.

5. Results:

The authors used the FID metric to evaluate the GAN and the accuracy, precision, recall, F1 score for ResNet18, which was trained only with synthetic images, and then tested on the original dataset. In the experiments, the generator produced a total of 1000 synthetic HSI images of epidermal lesions, balanced between benign and malignant classes.

- **Frechét Inception Distance**

The synthetic HS dataset generated by G obtained an FID value of 17.37. To evaluate and compare different FID results, the authors calculated the FID between the original data distribution and its augmented version. Specifically, they simply flipped horizontally each HS image in the dataset. In this case, the FID value measured was 8.96. The two FIDs are close, indicating that the synthetic and real data are similar.

- **Classification performance of ResNet18**

Metric	Value [%]
<i>accuracy</i>	84.21
<i>precision</i>	81.57
<i>recall</i>	86.11
<i>F1 score</i>	83.77

Figure 7: Classification performance of ResNet18 on real HS dataset

- **Spectral signature analysis**

From a visual inspection of the average spectral signatures and their ranges of variation, it is possible to see that the synthetic data can be used to describe the same distribution of the original dataset.

6. Limitations:

- The size of the original HS dataset is very small, only 76 images, making it difficult to ensure representativeness for different skin lesions.
- The use of transfer learning from RGB to HS images may cause spectral discrepancies between the two types of images, affecting the quality of the synthesized HS images.
- The evaluation of the GAN performance is based only on numerical metrics such as FID, accuracy, precision, recall and F1 score, without the involvement of medical experts to assess the quality of the synthesized HS images qualitatively.
- The use of a single type of HS camera may cause differences in spectral distribution among different HS cameras, making it challenging to create synthetic HS images with generalizability.

7. Future researches:

- Collect more data including data from different people, different types of skin lesions and different medical centers. Conduct a rigorous clinical study to test the usefulness of the proposed method.
- Evaluate whether the HS spatial information is related to the morphological features of different skin lesions. Perform qualitative assessments to evaluate the similarity between the distributions of synthetic and original skin lesions through a blind assessment test. Evaluate different types of HS cameras to develop models that generate different data distributions.

Mạng đối sinh tích chập sâu nhằm nâng cao trí tuệ nhân tạo trong chăm sóc sức khỏe: ứng dụng trong ung thư da

AI VIETNAM - RESEARCH TEAM

Ngày 8 tháng 10 năm 2023

Ngày công bố:	17/08/2022
Tác giả:	Marco La Salvia , Emanuele Torti, Raquel Leon , Himar Fabelo , Samuel Ortega , Beatriz Martinez-Vega , Gustavo M.Callico and Francesco Leporati
Nguồn:	2022 MDPI
Từ khóa:	Học sâu, hình ảnh siêu phổ, hình ảnh siêu phổ y tế, tạo dữ liệu tổng hợp, mạng đối sinh tích chập sâu
Người tóm tắt:	Nguyễn Đình Quý

1. Mục đích:

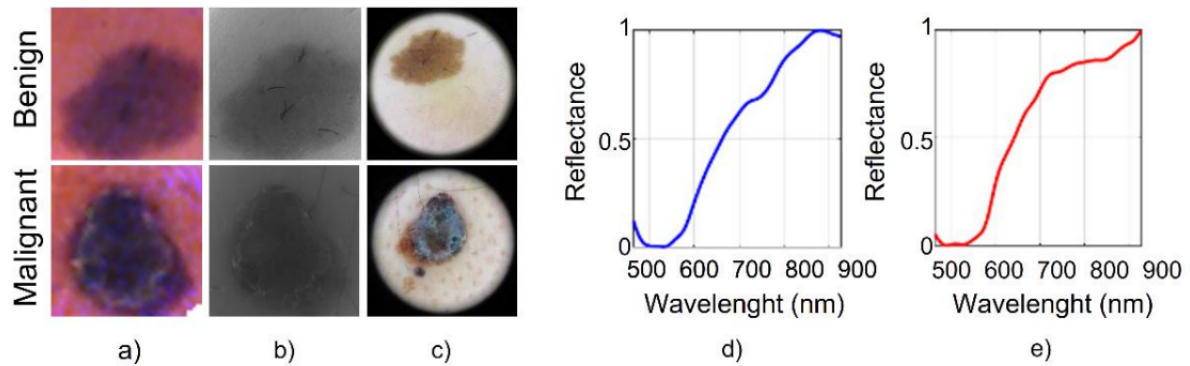
- Mục đích của bài báo là đề xuất một kiến trúc mạng đối sinh tích chập sâu (DCGAN) để tạo ra hình ảnh siêu phổ tổng hợp của các tổn thương biểu bì, dùng cho việc chẩn đoán ung thư da và giải quyết thách thức của các bộ dữ liệu có kích thước nhỏ để huấn luyện các kiến trúc học sâu.
- Bài báo cũng đánh giá hiệu quả của khuôn mẫu được đề xuất bằng việc sử dụng các phương pháp đo lường như khoảng cách Fréchet inception (FID), độ chính xác (accuracy), độ chính xác và độ phủ (precision, recall) cũng như điểm F1 (F1 score). Bài báo cho thấy DCGAN có thể cải thiện đáng kể khả năng chẩn đoán của trí tuệ nhân tạo trong lĩnh vực y tế.

2. Đóng góp:

- Bài báo áp dụng một kiến trúc DCGAN được mở rộng cho hình ảnh siêu phổ y tế, một lĩnh vực mới và ít được nghiên cứu.
- Bài báo sử dụng kỹ thuật học chuyển giao (transfer learning) để khởi tạo trọng số cho DCGAN từ một bộ dữ liệu RGB lớn hơn cũng như sử dụng kỹ thuật làm mịn nhãn (label smoothing).
- Bài báo biến đổi DCGAN thành một mạng có điều kiện (cGAN) để có thể sinh ra hình ảnh tổng hợp theo nhãn mong muốn.
- Bài báo sử dụng một kiến trúc mạng ResNet18 để kiểm tra sự tương đồng giữa dữ liệu tổng hợp được và dữ liệu thực.

3. Dữ liệu đầu vào:

- Bài báo sử dụng một bộ dữ liệu hình ảnh siêu phổ y tế trong cơ thể sống, bao gồm 76 hình ảnh siêu phổ của 40 tổn thương da lành tính và 36 tổn thương da ác tính, được chụp từ các bộ phận khác nhau của cơ thể từ 61 bệnh nhân. Bộ dữ liệu được xử lý và gán nhãn theo hai lớp: lành tính và ác tính. Bộ dữ liệu cũng bao gồm các hình ảnh RGB và đơn sắc của cùng một tổn thương da, được chụp bằng các máy ảnh khác nhau. Bộ dữ liệu này có kích thước không lớn, chỉ có 76 hình ảnh, do đó tác giả sử dụng DCGAN để tạo ra hình ảnh siêu phổ tổng hợp để huấn luyện một mạng phân loại sâu.
- Mỗi hình ảnh có độ phân giải không gian là 50×50 điểm ảnh và độ phân giải phổ là 8 nm, bao gồm 125 dải phổ từ 450 đến 950 nm (Hình 1a). Ngoài ra, máy ảnh cũng có tích hợp 1 cảm biến đơn sắc có khả năng chụp cùng một cảnh với một hình ảnh đơn sắc thông thường có độ phân giải là 1000×1000 điểm ảnh (Hình 1b). Ngoài hình ảnh siêu phổ (HS) và hình ảnh đơn sắc, người ta cũng chụp cả hình ảnh RGB thông thường của cùng một tổn thương da có kích thước 3000×4000 điểm ảnh (Hình 1c).
- Hình ảnh HS được tiền xử lý và hiệu chỉnh với các tham chiếu trắng và tối để chuẩn hoá các đặc trưng phổ (spectral signatures) cũng như loại bỏ 5 dải phổ đầu và 4 dải phổ cuối vì chứa nhiều nhiễu, kích thước cuối cùng của ảnh là 50×50 điểm ảnh và 116 dải phổ.
- Về việc gán nhãn, tác giả sử dụng kỹ thuật làm mịn nhãn. Các tổn thương ác tính được gán lớp dương, trong khi gán các tổn thương lành tính được gán là 0. Tác giả thay thế lớp dương bằng một giá trị ngẫu nhiên trong khoảng từ 0.7 đến 1 và lớp còn lại bằng các giá trị ngẫu nhiên từ 0 đến 0.3.



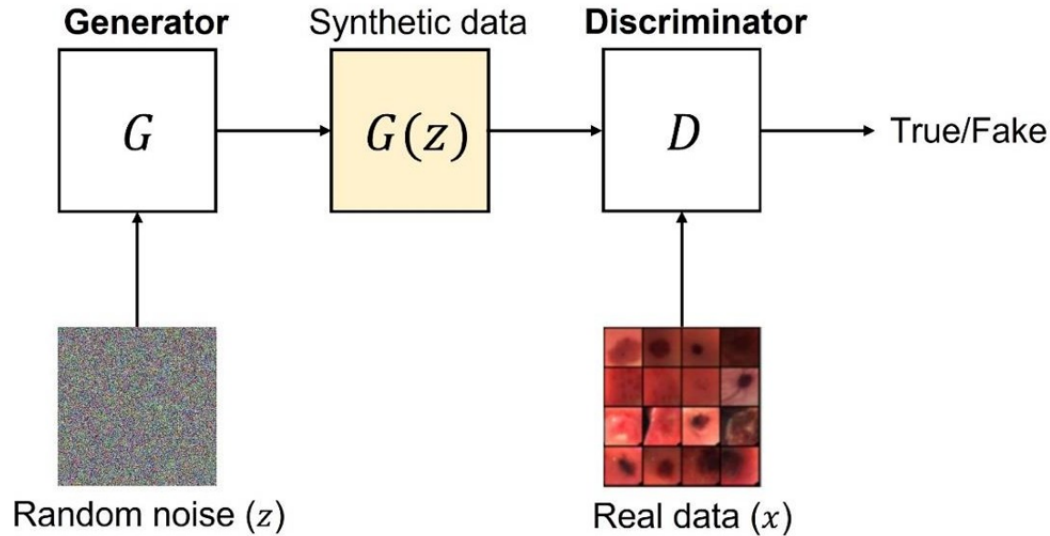
Hình 1: Ảnh ví dụ từ bộ dữ liệu. (a) Ảnh giả RGB (Pseudo-RGB images) thu được từ máy ảnh HS, (b) ảnh xám được chụp bằng cảm biến đơn sắc, (c) ảnh RGB thu được bằng máy ảnh soi da kỹ thuật số, (d) đặc trưng phổ trung bình của tổn thương lành tính, (e) đặc trưng phổ trung bình của tổn thương ác tính.

4. Phương pháp luận:

(a) Mạng đối sinh tích chập sâu:

Mô hình GAN được đề xuất lần đầu vào năm 2014 và dựa trên hai mạng con: một bộ sinh (G) và một bộ phân biệt (D). Hình 2 mô tả ý tưởng cơ bản đằng sau GAN. Bộ sinh G nhận đầu vào là một vector không gian ẩn (latent space vector) z từ phân phối chuẩn Gaussian và tạo ra một mẫu $G(z)$. Mẫu này đại diện cho ánh xạ từ không gian tiềm ẩn z sang không gian dữ liệu thực. Một mặt, G được tối ưu hóa để ước lượng phân phối dữ liệu huấn luyện

và tạo ra các mẫu giả có cùng phân phối với dữ liệu thực. Mặt khác, D nhận đầu vào là dữ liệu giả do G tạo ra hoặc một mẫu (x) đến từ tập dữ liệu thực. D trả ra ước lượng xác suất về nguồn gốc của dữ liệu đầu vào. Cụ thể, nó ước lượng xem mẫu đến từ dữ liệu huấn luyện hay từ G . G cố gắng giảm thiểu xác suất để D dự đoán đầu ra của nó là giả, trong khi D cố gắng tăng xác suất để phân biệt chính xác giữa dữ liệu thực và giả.

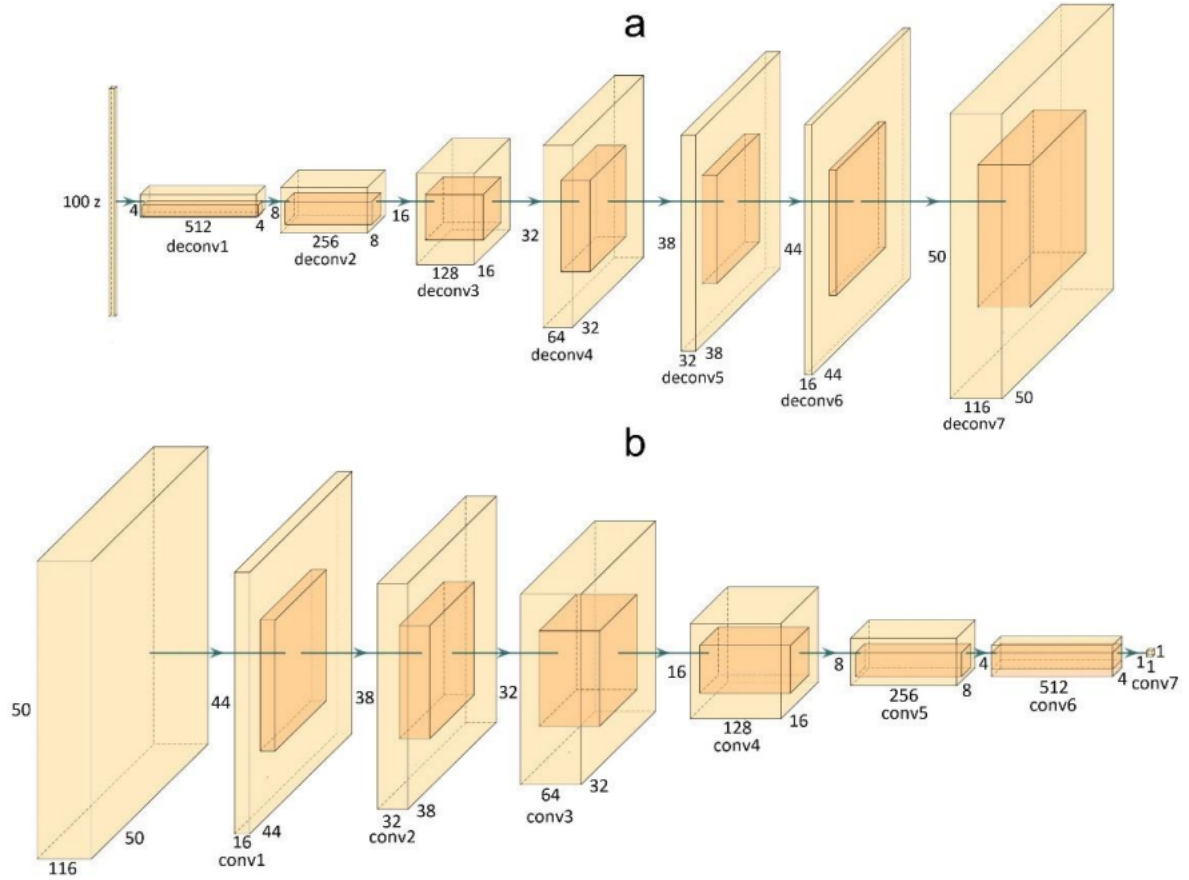


Hình 2: Cấu trúc GAN cơ bản

Trong những năm gần đây, các mạng đối sinh tích chập sâu đã trở thành một kiến trúc ổn định và tiết kiệm cho việc tạo ra ảnh giả. Xét về ảnh HS, việc chuyển đổi từ z sang không gian dữ liệu được thực hiện bởi G bao gồm việc tạo ra ảnh HS giả có cùng kích thước không gian và phổ với ảnh huấn luyện. Vì tác giả sử dụng tập dữ liệu ung thư da được mô tả trong Mục 3 làm tập huấn luyện, G sẽ tạo ra một ảnh có kích thước là $50 \times 50 \times 116$. Hình 3a cho thấy kiến trúc của G . Các lớp giải chập (deconvolutional) từ 1 đến 6 áp dụng chuẩn hóa theo batch và hàm kích hoạt ReLU. Lớp giải chập cuối cùng sử dụng hàm kích hoạt tanh. Mặt khác, D nhận đầu vào là một ảnh HS có cùng kích thước $50 \times 50 \times 116$, và thực hiện phân loại nhị phân để xác định xem ảnh đầu vào là thật hay giả. Hình 3b mô tả kiến trúc của D . Lớp tích chập đầu tiên sử dụng hàm kích hoạt leaky ReLU. Các lớp từ 2 đến 5 chuẩn hóa theo batch và dùng hàm kích hoạt leaky ReLU. Tất cả các hàm leaky ReLU đều có độ dốc âm bằng 0.2. Lớp tích chập cuối sử dụng hàm sigmoid.

(b) **Học chuyển giao:**

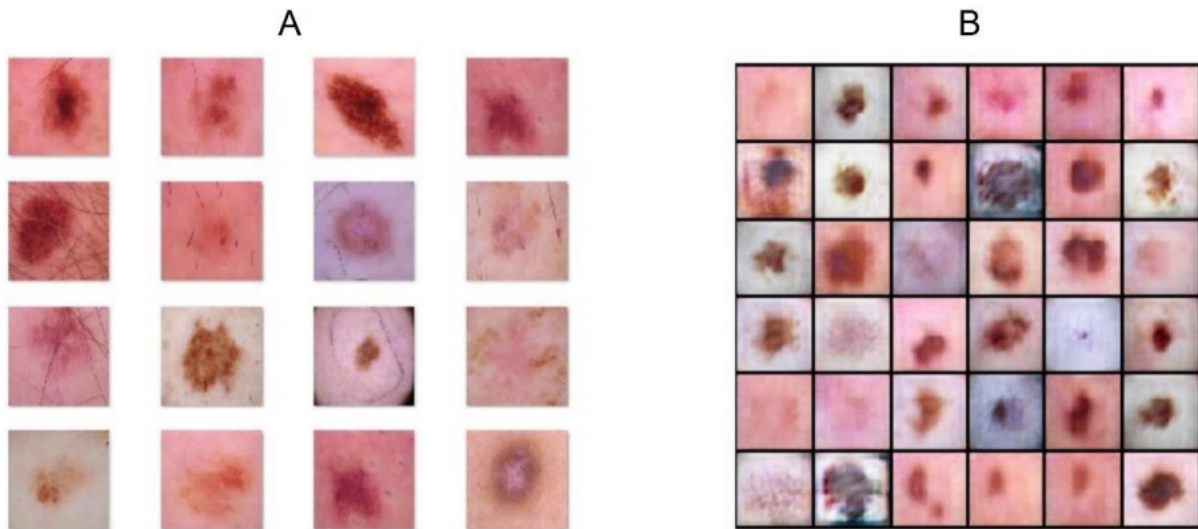
Phương pháp học chuyển giao bao gồm việc huấn luyện trước GAN bằng cách sử dụng các ảnh ung thư da RGB và sử dụng các tham số thu được làm tham số khởi tạo cho mô hình cuối cùng, mô hình này được huấn luyện bằng tập dữ liệu HS. Do đó, tác giả đã huấn luyện mô hình ban đầu bằng cách sử dụng tập dữ liệu HAM10000, chọn ngẫu nhiên 5000 ảnh RGB từ cơ sở dữ liệu. Tác giả đã thay đổi kích thước các ảnh thành 50×50 pixel để có cùng kích thước không gian với tập dữ liệu HS. Hơn nữa, tác giả cũng đã sửa đổi lớp đầu ra của G và lớp đầu vào của D để xử lý 3 kênh thay vì 116.



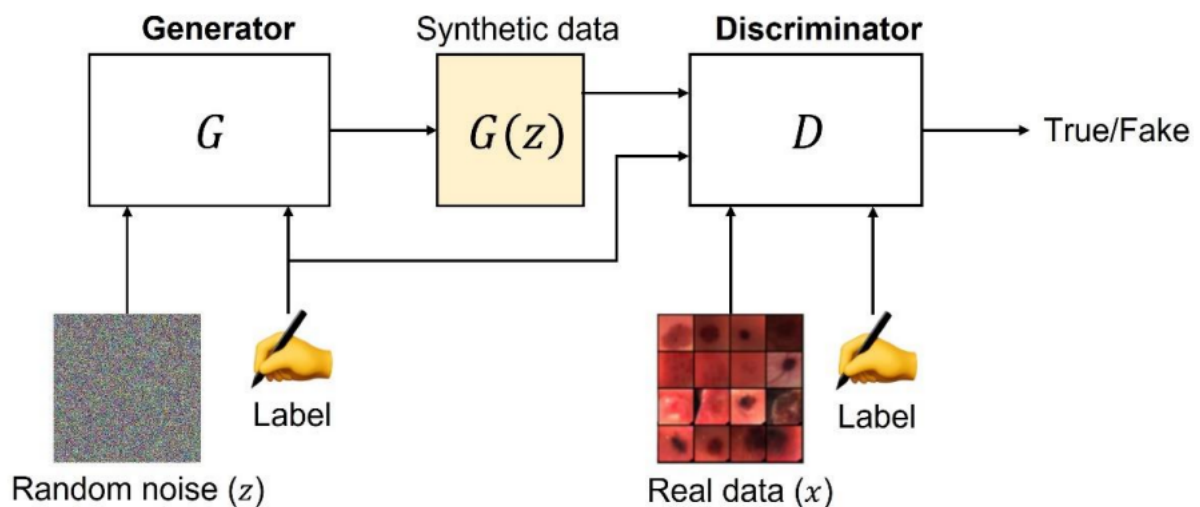
Hình 3: Cấu trúc bộ sinh (a) và bộ phân biệt (b)

Tác giả đã sử dụng thuật toán tối ưu Adam, với tốc độ học là 0.0002 cho cả hai mạng, và kích thước batch là 128. Tất cả các siêu tham số được chọn bằng cách thử sai, lặp lại quá trình huấn luyện với các giá trị khác nhau. Quá trình huấn luyện kết thúc sau 100 epoch. Cuối cùng, tác giả đã sử dụng kỹ thuật hoán đổi nhãn để tránh hiện tượng quá khớp (overfitting) của bộ D. Hình 4 cho thấy một số hình ảnh được lấy từ tập dữ liệu gốc và các hình ảnh khác nhau sinh ra bởi mô hình.

Tác giả đã chuyển các trọng số của mạng được lấy ra ở cuối quá trình huấn luyện này sang các kiến trúc được mô tả trong Hình 3 và thay đổi lớp đầu ra của G và lớp đầu vào của D thành kích thước 116; do đó, các giá trị thu được bằng cách huấn luyện với tập dữ liệu RGB được sử dụng để khởi tạo các trọng số liên quan đến các kênh tương ứng với bước sóng xanh lá, đỏ và xanh dương. Các giá trị còn lại được khởi tạo một cách ngẫu nhiên. Trong giai đoạn này, kích thước batch được giảm xuống còn 2. Hơn nữa, tác giả đã thay đổi kích thước của lớp đầu ra của G từ 116 thành 117. Kênh mới được sử dụng để sinh ra mặt nạ phân vùng (segmentation mask) liên quan đến hình ảnh tổng hợp. Việc sinh ra mặt nạ là rất quan trọng vì nó bao gồm thông tin có thể được sử dụng trong quá trình huấn luyện của một mạng phân vùng tổng quát, làm nổi bật các đường viền của tổn thương. Cuối cùng, kiến trúc được đề xuất đã được biến đổi thành một GAN có điều kiện (cGAN). Có nghĩa là G nhận làm đầu vào, cùng với vector nhiễu ngẫu nhiên, giá trị nhãn đã được làm mịn của lớp mà hình ảnh tổng hợp thuộc về. Nói cách khác, G có thể sinh ra dữ liệu giả liên quan đến các lớp lành tính hoặc ác tính. Kiến trúc của cGAN được đề xuất được hiển thị trong Hình 5.



Hình 4: (A) Ảnh lấy ra từ tập training, (B) Ảnh được tạo ra từ mô hình



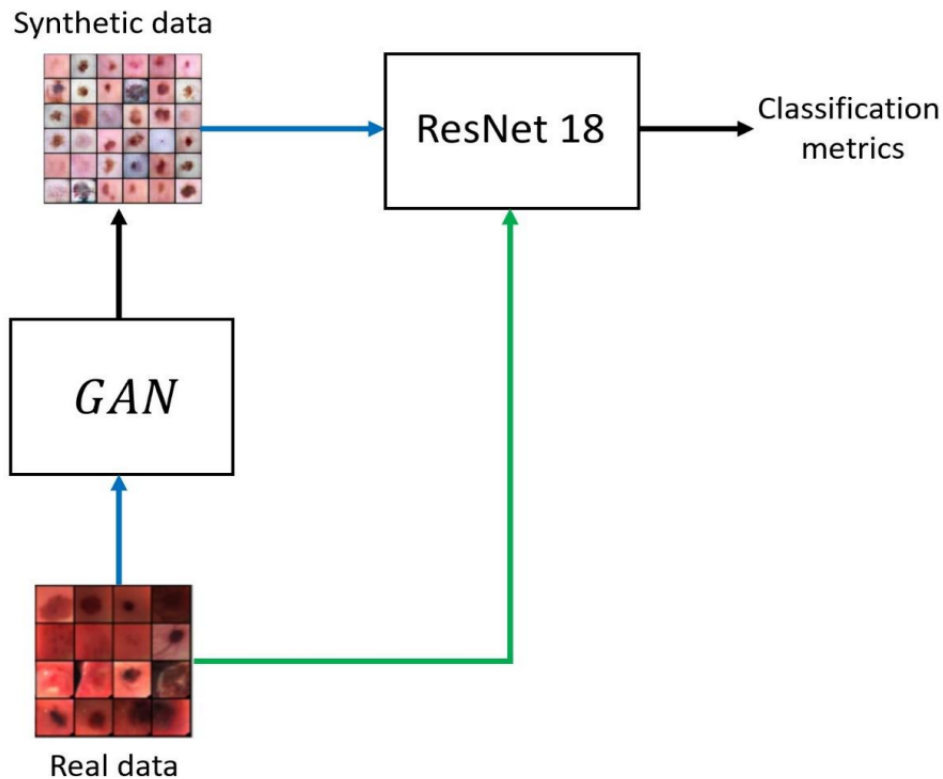
Hình 5: Kiến trúc cGAN

Tác giả đã huấn luyện cGAN trong 200 epoch. Trong quá trình huấn luyện, tác giả sử dụng các phương pháp khác nhau để cải thiện chất lượng của các hình ảnh tổng hợp. Trọng số của mỗi lớp được nhân với một hệ số c theo quy tắc tốc độ học bằng nhau. Hơn nữa, tác giả đã áp dụng quy tắc cập nhật hai thời gian (TTUR - Two Time-Scale Update). Cụ thể, tác giả đã gán cho hai mạng các giá trị tốc độ học khác nhau. Tốc độ học của G thấp hơn so với D . Do đó, trọng số liên quan đến G được cập nhật với nhiều bước hơn so với D , để nâng cao chất lượng của các hình ảnh tổng hợp. Để tránh D học cách phân biệt hình ảnh thật và giả trong một vài vòng lặp huấn luyện, tác giả đã hoán đổi nhãn cho 5% ngẫu nhiên dữ liệu huấn luyện. Thực tế, tác giả đã xử lý một số hình ảnh giả như là thật và ngược lại. Cuối cùng, tác giả đã sử dụng chuẩn hóa $L2$ ở mức 10^{-5} để giảm hiện tượng quá khớp.

(c) **Mô hình phân loại ResNet18:**

Tác giả đã sử dụng một mạng ResNet18 để đo sự giống nhau giữa dữ liệu HS thật và giả. Cụ thể, tác giả chỉ huấn luyện kiến trúc này với các hình ảnh HS giả để phân loại các tổn thương biểu bì thật. Do đó, tác giả đã khai thác hiện tượng quá khớp như một phương pháp để thể hiện rằng dữ liệu giả tái hiện phân bố thống kê của dữ liệu thật như thế nào. Phương pháp này cho biết liệu tập dữ liệu giả có đại diện cho ý nghĩa của tập dữ liệu thật hay không. Trong trường hợp này, hiện tượng quá khớp không nên được coi là một ảnh hưởng tiêu cực, quá khớp trên tập dữ liệu giả và đạt được kết quả tốt trong việc phân loại tập dữ liệu thật có nghĩa là mô hình thu được đã tổng quát hóa vấn đề được xem xét.

Trong Hình 6, các mũi tên màu xanh lam chỉ ra rằng tập hợp đã được sử dụng để huấn luyện mô hình, trong khi mũi tên màu xanh lá cây cho biết rằng tập dữ liệu được sử dụng làm đầu vào cho việc phân loại.



Hình 6: Phương pháp đề xuất nhằm đánh giá sự tương đồng giữa các tập dữ liệu

(d) **Phương pháp đánh giá:**

FID (Frechét inception distance): Đo khoảng cách giữa các vector đặc trưng của hình ảnh thật và giả. Giá trị FID càng thấp, hai tập hình ảnh càng tương đồng.

$$FID = \|\mu_1 - \mu_2\|^2 + \text{Tr}(\Sigma_1 + \Sigma_2 - 2\sqrt{\Sigma_1 \Sigma_2})$$

Trong đó μ biểu thị giá trị trung bình, Σ là ma trận hiệp phương sai và Tr biểu thị vết tích của một ma trận. Các chỉ số 1 và 2 chỉ tập hợp các hình ảnh thật và giả.

Accuracy, precision, recall, F1 score: Đo lường hiệu quả của mạng ResNet18 trong việc phân loại hình ảnh thật dựa trên dữ liệu giả. Các chỉ số này đều cao, cho thấy dữ liệu giả có chất lượng tốt.

5. Kết quả:

Tác giả sử dụng phép đo FID để đánh giá GAN và sử dụng accuracy, precision, recall, F1 score cho ResNet18, được huấn luyện chỉ với các hình ảnh nhân tạo, và sau đó được kiểm tra trên tập dữ liệu gốc. Trong các thử nghiệm, bộ sinh đã tạo ra tổng cộng 1000 hình ảnh HSI nhân tạo của các tổn thương biểu bì, cân bằng giữa các lớp lành tính và ác tính.

- **Frechét Inception Distance**

Tập dữ liệu HS nhân tạo được tạo bởi G đạt giá trị FID là 17,37. Để đánh giá và so sánh các kết quả FID khác nhau, tác giả tính FID giữa phân bố dữ liệu gốc và phiên bản tăng cường. Cụ thể, chỉ đơn giản là lật ngang mỗi hình ảnh HS trong tập dữ liệu. Trong trường hợp này, giá trị FID đo được là 8,96. Hai FID gần nhau cho thấy dữ liệu nhân tạo và dữ liệu thực có sự tương đồng.

- **Hiệu suất phân loại của mạng ResNet18**

Metric	Value [%]
<i>accuracy</i>	84.21
<i>precision</i>	81.57
<i>recall</i>	86.11
<i>F1 score</i>	83.77

Hình 7: Hiệu suất phân loại tập dữ liệu HS thực của ResNet18

- **Phân tích đặc trưng phổ**

Từ việc kiểm tra trực quan các đặc trưng phổ trung bình và phạm vi biến thiên của chúng, có thể thấy dữ liệu nhân tạo có thể được sử dụng để mô tả cùng một phân bố của tập dữ liệu gốc.

6. Hạn chế:

- Kích thước của bộ dữ liệu HS ban đầu là rất nhỏ, chỉ có 76 ảnh, do đó khó đảm bảo tính đại diện cho các tổn thương da khác nhau. Việc sử dụng transfer learning từ ảnh RGB sang ảnh HS có thể gây ra sự sai lệch về phổ giữa hai loại ảnh, do đó ảnh hưởng đến chất lượng của ảnh HS được tổng hợp.
- Việc đánh giá hiệu quả của mô hình GAN chỉ dựa trên các chỉ số số học như FID, accuracy, precision, recall và F1 score, chưa có sự tham gia của các chuyên gia y tế để đánh giá chất lượng của ảnh HS tổng hợp một cách định tính.
- Việc sử dụng một loại máy ảnh HS duy nhất có thể gây ra sự khác biệt về phân phối phổ giữa các máy ảnh HS khác nhau, do đó gây ra khó khăn trong việc tạo ra các ảnh HS tổng hợp mang tính tính tổng quát.

7. Các nghiên cứu trong tương lai:

- Thu thập thêm dữ liệu bao gồm dữ liệu từ nhiều người, từ các loại tổn thương da khác nhau và từ nhiều trung tâm y tế. Thực hiện một nghiên cứu lâm sàng nghiêm ngặt để kiểm tra tính hữu ích của phương pháp được đề xuất. Đánh giá xem thông tin không gian HS có liên quan đến các đặc điểm hình thái của các tổn thương da khác nhau hay không.
- Thực hiện các đánh giá định tính để đánh giá sự tương đồng giữa phân bố của các tổn thương da nhân tạo và tổn thương gốc thông qua một bài kiểm tra đánh giá mù (blind assessment test). Đánh giá nhiều loại máy ảnh HS để phát triển mô hình tạo ra các phân bố dữ liệu khác nhau.