2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-1179-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EMBC46164.2021.9630147

Classification of Non-tumorous Facial Pigmentation Disorders Using Generative Adversarial Networks and Improved SMOTE*

Jiawei Peng, Ruihan Gao, Steven Thng, Weimin Huang, Zhiping Lin

Abstract—**The diagnosis of non-tumorous facial pigmentation disorders is crucial since facial pigmentations can serve as a health indicator for other more serious diseases. The computer-based classification of non-tumorous facial pigmentation disorders using images / photographs allows automated diagnosis of such disorders. However, the classification performance of existing methods is still not satisfactory due to the limited real-world images available for research. In this paper, we proposed a novel approach to applying generative adversarial network (GAN) with improved synthetic minority over-sampling technique (Improved SMOTE) to enhance the image dataset with more varieties. With the application of Improved SMOTE, more data is provided to train GAN models. By utilizing the GAN to perform data augmentation, more diverse and effective training images can be generated for developing classification model using deep neural networks via transfer learning. A significant increase in the classification accuracy (>4%) was achieved by the proposed method compared to the state-of-the-art method.**

I. Introduction

Non-tumorous skin pigmentation disorders are one of the common diseases which affect a large population around the world. It not only affects a person's facial appearance but also affects one's mental health as stress may be developed with prolonged non-tumorous facial pigmentation disorders. In addition, the different types of facial pigmentation are treated differently as one form of treatment might aggravate another cause of facial pigmentation. Moreover, some non-tumorous facial pigmentation disorders might be an indication of more serious disease for the patient and thus requires more investigations. Therefore, it is important to classify non-tumorous facial pigmentation disorders to assist the proper diagnosis and treatment. Furthermore, as manual classification is time-consuming [1], the development of an automatic and accurate computer-aided classification method for non-tumorous facial pigmentation disorders is of great significance.

Related existing work in classification of non-tumorous facial pigmentation disorders started with the development of a Voting based Probabilistic Linear Discriminant Analysis (V-PLDA) method [2] which focuses on solving the problem of large intra-class variance. However, due to the use of handcraft features such as color and texture, and the limited training data, the result was not good enough. To address the problem of the available small dataset, a method incorporating transfer learning and data augmentation using Synthetic Minority Oversampling Technique (SMOTE) was proposed in [3]. An improved SMOTE method [4] was further proposed to enhance the effectiveness of the enlarged dataset. The methods in [3,4] have overcome the issue of small dataset to some extent. However, the data augmentation schemes with both SMOTE and improved SMOTE [3,4] are limited by linear combination of existing data samples and thus provide insufficient variance among the new data generated.

In this paper, we propose to apply Generative Adversarial Network (GAN) [5] with improved SMOTE [4] to achieve data augmentation for the classification of non-tumorous facial pigmentation disorders. The learning process of GAN involves automatically discovering the patterns and distribution in input so that the model can be used to generate new samples containing information drawn from the original dataset but with more varieties. GAN technique has been applied to generate artificial biomedical images such as retinal images and skin lesions [6-8]. Progressive GAN (PGAN) were adopted in this paper. PGAN [9] provides possibility to generate images with high resolution. By applying PGAN to generate proper training images, we effectively prevent the model from overfitting and overcome the shortcomings of the limited dataset. Specifically, since GAN model still requires a large training set, we propose to use SMOTE to generate some new images as an enlarged training dataset for the GAN models. We show by experiments that the newly proposed method in this paper has improved the classification accuracy by more than 4% compared to [4] for the same clinical image set of non-tumorous facial pigmentation disorders [10].

II. Methodology

In this section, we propose a two-step data augmentation approach that enlarges a very small training dataset to enhance the performance of a deep learning model for the classification of non-tumorous facial pigmentation disorders. We use the improved SMOTE first, followed by PGAN [9] to automatically augment images with more diversity.

A. Improved SMOTE

SMOTE [11] is a technique to perform data augmentation by randomly interpolating similar samples measured by a distance metric. Improved SMOTE is a modified version of it to produce more varieties by removing highly similar new samples generated by SMOTE [4]. Specifically, the range of weight coefficient is restricted to prevent generating repeating images resembling the original images closely. Larger data augmentation multiplier was chosen to synthesize more images. Moreover, highly redundant images are removed from

*Research supported by Nanyang Technological University under the Undergraduate Research Experience on Campus (URECA) program.

Jiawei Peng, Ruihan Gao, Zhiping Lin are with School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639 798 (e-mail: {jpeng004, gaor0007}@e.ntu.edu.sg; ezplin@ntu.edu.sg).

Steven Thng is with National Skin Center, Singapore 308205 (e-mail: steventhng@nsc.com.sg).

Weimin Huang is the corresponding author, with the Institute for Infocomm Research, Singapore 138632 (e-mail: wmhuang@i2r.a-star.edu.sg).

the training set by assessing the similarity between images using SSIM index [12]. It has been demonstrated that improved SMOTE can generate effective data for training but using improved SMOTE alone has its limitation in generating enough new data [4]

B. GAN

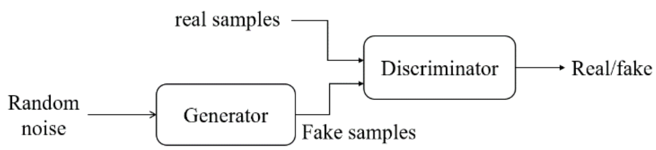


Figure 1. Workflow of GAN

GAN [5] is a generative model to produce consistent and realistic image features by utilizing the concept of adversarial learning. A GAN architecture has two main components as shown in Fig.1: a generator and a discriminator. The generator is trying to generate more realistic samples from a random noise vector input. It learns the latent structure and distribution of the data it aims to reproduce. When a noise vector is fed into it, it generates a sample from the approximated distribution [13]. The discriminator, which is similar to a classifier, is trained to inspect the authenticity of the input samples. It takes both real samples (those in the original dataset) and fake samples (produced by the generator) as input and tries to assess their authenticity. Training a GAN is similar to a min-max game between the generator and the discriminator. The generator is trying hard to produce more realistic samples which the discriminator cannot differentiate as fake. Concurrently, the discriminator is progressing to evaluate the quality of the produced samples compared to the real samples more accurately. The two models are trained together to compete with each other, until they reach the "Nash Equilibrium" when the discriminator has a chance close to 50% to predict accurately the authenticity of the input samples, meaning that the generator is producing plausible samples. Those generated samples output from the generator can be processed for further applications.

C. PGAN

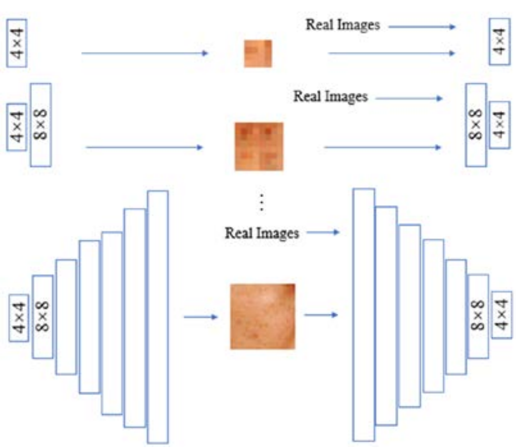


Figure 2. Architecture of PGAN

The work of progressive growing of GAN (PGAN) [9] has shown promising results for high-resolution image generation while maintaining high quality. It has been demonstrated that PGAN can generate celebrity faces up to 1024×1024. As shown in Fig.2, PGAN [9] utilize progressive growing network to generate images from low-resolution to high-resolution step by step. The learning start from small network to generate low-resolution images first. The model gradually grows in depth and complexity when the smaller networks are well-trained. When adding new layers to the model, A fade-in method is applied and hence the models gradually transit to higher-resolution generator and discriminator. This ensures the stable training and avoid sudden shock to the already well-trained smaller network during each growing step. Progressively, the output feature map dimension of the generator and the input feature map dimension of the discriminator is amplified while maintaining stable training and high quality of the generated images.

D. Proposed Architecture

In this paper, PGAN [9] with improved SMOTE [4] was applied to implement data augmentation. Comparing to traditional data augmentation method such as random subsampling and SMOTE, the GAN model is able to learn the training set distribution and hence it can generate more general images with more variations. Statistically, by adding more diverse images to the training set, the distribution of the training set can be closer to the true distribution of the dataset so that the learning error caused by the empirical risk of the class predicting function can be reduced.

Since the training of a GAN model still requires large dataset to avoid problems such as model collapsing, we propose to apply the improved SMOTE to synthesize artificial images to provide more varieties to the dataset before training the GAN model as shown in Fig.3. Both the original images and the improved SMOTE synthesized images were considered as real images to be fed into the discriminator and the discriminator tries to identify the source of the images, i.e., whether the image is generated by the generator.

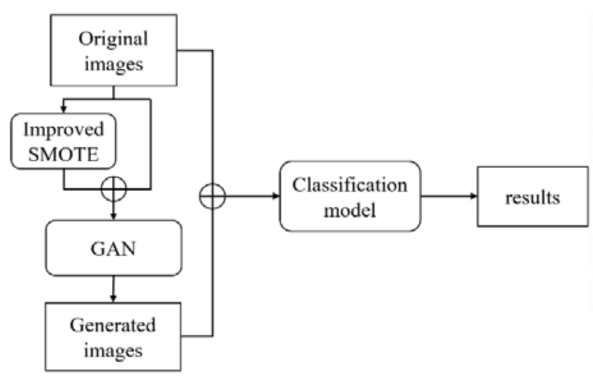


Figure 3. Workflow of the proposed method

III. EXPERIMENT

A. Original Dataset

The dataset used in this research is a real-world clinical dataset collected in National Skin Center, Singapore (see [2]). This experiment focusses on the classification of five common classes of non-tumorous facial pigmentation disorders found in Asia [10], including freckles, lentigines, melasma, Hori's nevus, and nevus of Ota. 30 images in each class are available in the original dataset. Before training, images are resized to a fixed size 100×100×3 and normalized for easier training. A typical image in each class is shown in Fig.4.

Freckles Lentigines Melasma Hori's Nevus Nevus of Ota

Figure 4. Sample images in original dataset [2]

B. Generating Dataset by GAN

Improved SMOTE is applied prior to the PGAN model to provide a larger dataset for the generator to learn a more generalized distribution of the pigmentation data. For each class, 120 images are synthesized using improved SMOTE. Fig.5 shows sample images generated by improved SMOTE. Subsequently, both the 30 original images and the 120 improved SMOTE synthesized images are resized and normalized, and then fed into the discriminator of the PGAN model. GAN models are trained on each class so that every distinct model concentrates on only one class and it is easier to learn the distribution within one class.

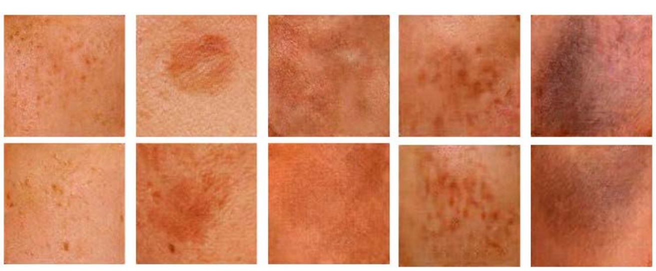
Freckles Lentigines Melasma Hori's Nevus Nevus of Ota

Figure 5. Sample images generated using improved SMOTE

We follow the PGAN network architecture and parameters stated in [9]. The generation starts from 4×4 pixel. Fade-in and stabilization were performed when transforming to the next resolution. As result, 120 images in each class are generated at a resolution of 128×128. All of these newly generated images by PGAN are added into the training set of the classifier while those new images generated by the improved SMOTE are removed so that the total number of training images for each class is 150 which is the same as the existing method [4] for a fair comparison. All the generated images are resized to 100×100 to keep it consistent to the original dataset for classification.

Freckles Lentigines Melasma Hori's Nevus Nevus of Ota

Figure 6. Sample images generated by PGAN



Freckles Lentigines Melasma Hori's Nevus Nevus of Ota

Figure 7. Sample images generated by PGAN with Improved SMOTE

Fig.6 and Fig.7 shows some images generated by only PGAN (one image in each class) and PGAN with improved SMOTE (two images in each class), respectively. It is observed that PGAN-generated images using only the original dataset do not preserve the basic shape and texture of the skin pigmentation. Since Progressive GAN is a complex model and it requires continuous transition from low resolution to high resolution, this process is easy to collapse. A very small training set cannot guide the model properly. In contrast, images generated by PGAN with improved SMOTE have much better quality. They also have fine-grained details of the pigmentation and skin texture. It is observed that PGAN is sensitive to the size of real dataset. Hence, the deployment of improved SMOTE to enlarge the input data for PGAN makes significant contribution to the stable training of PGAN.

To illustrate the performance of PGAN with the improved SMOTE compared to the method using the improved SMOTE only [4], the similarity between the generated data and original data are evaluated using the same criteria stated in [4]. SSIM index [12] is adopted to gauge the similarity between each generated image and each original image. For every generated image, out of the comparisons with all the original images, the highest SSIM index (HSI) is taken. The HSI range of the generated images in each class is tabulated in Table I where five classes are represented by a capital letter, namely, F: Freckles; L: Lentigines; M: Melasma; H: Hori's Nevus; O: Nevus of Ota. It is shown that by using GAN model with the improved SMOTE, newly generated images are more distinct to the original images. This shows that GAN can add more diversity and variety within each class.

TABLE I. HSI RANGE USING DIFFERENT METHODS

HSI	Improved SMOTE	PGAN with the improved SMOTE
F	0.878-0.998	0.774-0.957
L	0.877-0.997	0.724-0.976
M	0.882-0.997	0.691-0.954
H	0.908-0.998	0.662-0.965
O	0.882-0.999	0.684-0.949

D. Classification model setting

The classification model deployed in this experiment consists of a pre-trained CNN model as the feature extractor and fully connected layers as the classifier. Inception-Resnet-v2 [14] pre-trained on general dataset is chosen as the feature extracting model which is the same as in [3, 4] for a baseline comparison. The model is fine-tuned to fit the enlarged dataset. A grid search is performed to find the optimal hyperparameters. A weight exponential decay mechanism is adopted during the training. Early stopping is triggered when the training loss is not decreasing for a certain number of training steps. To produce statistically reliable results, a ten-fold validation is performed to assess the performance of the classification model and the ten-fold validation process is performed for 10 times. The average results of all the 10 runs are taken as the final value.

IV. RESULTS AND ANALYSIS

To illustrate the excellent performance of our proposed method in this paper, the comparison of the experiment results among the 3 methods, namely V-PLDA [2], Transfer Learning

+ Improved SMOTE [4], Transfer Learning + PGAN with improved SMOTE are presented in this section. Table II shows the overall accuracy and standard deviation attained using the same classification model Inception-ResNet-v2 but with different data augmentation methods.

TABLE II. CLASSIFICATION RESULTS WITH DIFFERENT METHODS

Method	Accuracy (%)	Standard Deviation
V-PLDA [2]	77.33	0.0982
Transfer Learning + Improved SMOTE [4]	87.33	0.0767
Transfer Learning + PGAN with the improved SMOTE	91.67	0.0659

It is observed that our proposed method using PGAN with the improved SMOTE yield a remarkable enhancement of more than 4.3% compared to the state-of-the-art model Inception-ResNet-v2 with improved SMOTE [4]. Compared to the V-PLDA method [2], A more significant accuracy gain of 14.3% is observed using proposed method. Since GAN models are able to discover the distribution of the input data and fill the real image distribution that is not covered in the original dataset, the generated images can provide more information about the true dataset distribution to the classification model. More variations are created and hence the classification model becomes more robust to the unseen test data.

To further compare the performances among the existing methods and the proposed method, we analyzed the confusion matrices (omitted here due to space limitation). The proposed methods (PGAN with the improved SMOTE) show significant improvement for all the five classes comparing to the method V-PLDA [2]. Compared to the data augmentation using the improved method [4] , the proposed method yields significant gain in accuracy among most of the classes. It is worth noting that for classes poorly classified by [2] which is Hori's of Nevus and Nevus of Ota, the proposed methods increase the accuracy by 27.6% and 18%, respectively. For classes that are poorly classified by [4], the proposed method increases the accuracy by 10% for Melasma.

The proposed method in this work might be transferred to solving other image-based classification and diagnosis problems in the general biomedical field. With faster and more accurate automated diagnosis, we hope to help reduce the risk of inappropriate treatments.

V. CONCLUSION

In this paper, to alleviate the challenge of a limited and domain-specific dataset to classify non-tumorous facial pigmentation disorders, we have proposed to apply PGAN with the improved SMOTE to perform data augmentation on the given small dataset. Compared to the state-of-the-art methods using the improved SMOTE to synthesize training images, image set generated by GAN has more varieties and provides more information to the classification model. The use of enlarged dataset with the improved SMOTE also facilitates the stable training of GAN. Our proposed method yielded a remarkable improvement (>4%) in terms of overall classification accuracy compared to the state-of-the-art method. This improvement is very significant as it leads to the overall classification accuracy of over 90%, which gives higher confidence in treatment and diagnosis of facial pigmentation disorders. Moreover, it might also be applied to solving other similar problems in the general biomedical field where limited and domain-specific datasets are often the case in practice.

REFERENCES

[1] L. Brochez, E. Verhaeghe, L. Bleyen, and J. M. Naeyaert, "Diagnostic ability of general practitioners and dermatologists in discriminating pigmented skin lesions," in Journal of the American Academy of Dermatology, vol. 44, pp. 979-986, 2001.

[2] Y. Liang, L. Sun, W. Ser, F. Lin, S T. G. Thng, Q. Chen and Z. Lin, " Classification of non-tumorous skin pigmentation disorders using voting based probabilistic linear discriminant analysis," Computers in Biology and Medicine, vol. 99, pp. 123-132, 2018.

[3] R. Gao, J. Peng, L. Nguyen, Y. Liang, S. Thng and Z. Lin, "Classification of non-tumorous facial pigmentation disorders using deep learning and SMOTE," in 2019 IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan, 2019.

[4] J. Peng, R. Gao, L. Nguyen, Y. Liang, S. Thng and Z. Lin, "Classification of non-tumorous facial pigmentation disorders using improved SMOTE and transfer learning," in 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in 2014 Conference on Neural Information Processing Systems (NIPS), Montreal, Canada, 2014.

[6] H. Zhao, H. Li, S. Maurer-Stroh and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," in Medical Image Analysis, vol. 49, pp. 14-26, 2018.

[7] H. Rashid, M. A. Tanveer and H. Aqeel Khan, "Skin lesion classification using GAN based data augmentation," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019.

[8] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," Neurocomputing, vol. 321, pp. 321–331, 2018.

[9] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in 2018 6th International Conference on Learning Representations (ICLR), Vancouver, Canada, 2018.

[10] S. G. Y. Ho and H. H. L. Chan, "The Asian dermatologic patient," American Journal of Clinical Dermatology, vol. 10, no. 3, pp. 153-168, June 2009.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, no. 16, pp. 321-357, 2002.

[12] Z. Wang, B. A. Conrad, S. H. Rahim, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[13] F. H. K. S. Tanaka, C. Aranha, "Data augmentation using GANs, " in arXiv preprint arXiv:1904.09135, 2019.

[14] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-4, inception-resnet and the impact of residual connections on learning," in the 31st AAAI Conference on Artificial Intelligence, San Francisco, 2016.