Classification of Non-tumorous Facial Pigmentation Disorders Using Deep Learning and SMOTE

Ruihan Gao*, Jiawei Peng*, Long Nguyen*, Yunfeng Liang§, Steven Thng¶, Zhiping Lin*
* School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
§ Interdisciplinary Graduate School, Nanyang Technological University, Singapore
¶ National Skin Center, Singapore

Abstract— **Non-tumorous facial pigmentation, though not fatal, adversely affects one's quality of life and may indicate concurrence of systemic diseases. Automatic diagnosis method such as voting-based probabilistic discriminant analysis (V-PLDA) has been explored, but the accuracy of classification is not satisfactory due to the limited number of data for training. This paper proposes to use the pre-trained deep learning network of Inception-ResNet-v2 so that information from similar datasets can be utilized. Furthermore, data augmentation using synthetic minority over-sampling technique (SMOTE) is also applied to make full use of available training data. A clinical dataset of five most common types of non-tumorous facial pigmentation disorders in Asia, namely freckles, lentigines, melasma, Hori's nevus, and nevus of Ota, is used for training and testing. The classification accuracy has shown significant improvement (> 7%) compared to the state-of-the-art method.**

Keywords— deep convolutional neural network, SMOTE, facial pigmentation disorders, biomedical images classification.

I. INTRODUCTION

Biomedical image analysis is of vital importance for automatic classification and diagnosis in both medical research and clinical practices [1, 2]. In particular, in the field of facial pigmentation, much attention has been directed to the automatic detection and classification of pigmented facial tumors, such as basal cell cancer, squamous cancer, and melanoma in recent years [3, 4]. However, few studies have been conducted for the automatic classification of non-tumorous facial pigmentation disorders. According to the experiences of dermatologists, patients also tend to overlook the significance of non-tumorous facial pigmentation disorders or resort to non-professional institutes like beauty salons for help. The resulting wrong diagnosis and mistreatment are detrimental since although not as fatal as tumors, non-tumorous facial pigmentation disorders not only impair the facial appearance but also serve as a warning sign of health conditions in vivo.

To address the abovementioned problems, machine learning stands out with computational capability and good performance across diverse applications. It has been widely applied to automatic biomedical image processing for both detection and classification in recent years. In particular, voting-based probabilistic linear discriminant analysis (V-PLDA) is applied to classify non-tumorous facial pigmentation disorders [5]. A voting-based method is used to cope with the high within-class variance of different types of facial pigmentation. Gabor features [6, 7] and features in hue, saturation, value (HSV) color space are also extracted for image training to represent the color and texture information [5]. However, its performance highly relies on human effort for trial and error and is limited by a small set of domain-specific dataset.

To address the problem of limited data and to further improve the classification accuracy for non-tumorous facial pigmentation disorders, we propose in this paper a new method by combining deep learning with data augmentation using Synthetic Minority Over-sampling Technique (SMOTE). Deep learning, particularly convolutional neural network (CNN), uses a deep cascade of multiple layers of neurons to extract and transform features. With its depth and extensive connections between layers of neurons, the deep learning model can obtain generic features which fit a broader set of tasks [8]. Several pre-trained models have been developed and successfully applied to perform biomedical image classification [9-11].

Synthetic Minority Over-sampling Technique (SMOTE) is also implemented in this paper to generate synthetic data [12]. As proposed in [13], SMOTE is an oversampling approach that was initially designed to deal with an imbalanced dataset. In this paper, we use it to generate artificial but related samples at data-level by taking a random point along the line segment joining two near neighbors. It is proven to avoid the overfitting issue effectively and has been successfully employed to prevent the generation of noise [14] and to tackle imbalanced datasets with rough set theory [15].

Specifically, we propose to replace the hand-craft features in V-PLDA by generic features extracted by the deep convolutional neural network. Inception-ResNet-v2 is chosen as the pre-trained model since it combines inception network and residual connections [16] and promisingly extracts more generic features from intricate patterns of facial pigmentation images. Furthermore, SMOTE is employed for data augmentation so that the available information in limited images can be fully utilized. We show by experiments with a clinical dataset of images related to non-tumorous facial pigmentation disorders that the proposed method has achieved significant improvement in terms of overall classification accuracy (> 7%) for non-tumorous facial pigmentation disorders. Since both variation among features and a small available domain-specific dataset are common in biomedical applications, our proposed method is also promising to be applied to other similar tasks.

The remaining part of this paper is structured as follows. Section II presents the methodology of combining deep learning and data augmentation. The structure of Inception-ResNet-v2 (Section IIA) and implementation of SMOTE (Section IIB) are presented. Section III illustrates the clinical dataset used and experiment settings in details. Moreover, experiment results are presented, and a detailed analysis is also conducted in this section. Section IV concludes the paper.

II. Methodology

In this paper, we propose an approach to dealing with a relatively small and domain-specific image dataset: combining transfer learning in deep neural networks with synthetic minority over-sampling technique (SMOTE). Among available pre-trained CNNs, Inception-ResNet-v2 [16] merges the advantages of inception modules and residual connections and has gained popularity for image classification. It is chosen for the implementation in this paper and a brief overview of the proposed method is shown in Fig.1.

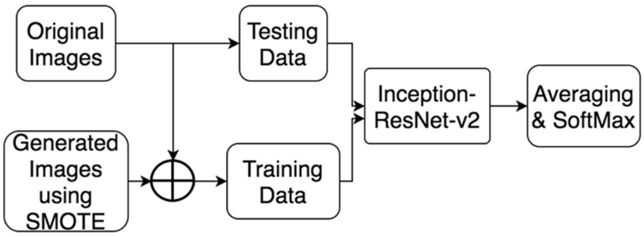


Figure 1. The pipeline of the proposed method.

A. Deep learning with the pre-trained model Inception-ResNet-v2

As stated above, transfer learning is a useful tool to extend the application of Deep Neural Network (DNN) to a small dataset. Training a network with tens or hundreds of layers from scratch requires tens of thousands of images and all of them need to be labelled for supervised learning [17]. With transfer learning which exploits a base model that has been trained on similar but more general data, the data required to perform a specific task can be considerably reduced [18]. This significant reduction abridges the gap of the amount of data between the demand from deep learning networks and the supply or available data from practical tasks or problems to be solved.

The base model chosen in this paper is Inception-ResNet-v2. As its name suggests, Inception-ResNet-v2 is a combination of inception network and residual network. Inception network approximates a sparsely connected CNN and keeps a small number of the convolutional filters for the kernel size, such as 5×5 and 3×3 [19]. In addition, inception network also uses bottleneck layers to help reduce the massive computation requirement and thus allow the network to achieve higher depth and width. The residual network, consisting of building blocks called residual modules that add a direct path between the input and output to imply the identity mapping, is designed to cope with the vanishing gradient problem where the contribution by previous layers to the adjustment of the weights reduces as the networks grow deeper and deeper [19]. The combination of the two networks is further simplified by only performing batch-normalization on the top of traditional layers, and this allows more inception blocks to be built and reduces the overall memory consumed. Moreover, the network is also further stabilized by trimming the residual modules to attain a more robust training process. Inception-ResNet-v2 is chosen in this paper because it is state-of-art among the family of pre-trained models and has recognized the 1000 different classes of objects in the validation dataset of ImageNet 2012 Large Scale Visual Recognition Challenge (ILSVRC) with the lowest top-5 error among inception net and residual net counterparts [16]. In this research, a few parameters like the learning rate are tuned to better fit the dataset. Dropout technique is also used to reduce overfitting. The details of the structure of Inception-ResNet-v2 is shown in Fig.2.

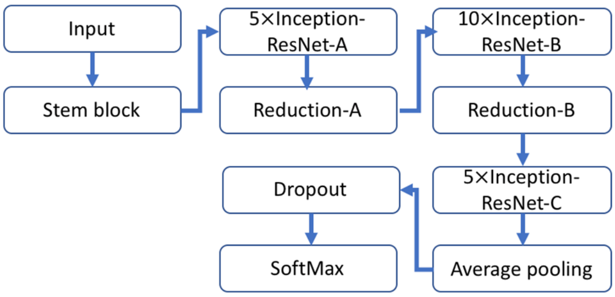


Figure 2. The structure of pre-trained Inception-ResNet-v2 [16].

B. SMOTE

SMOTE, or synthetic minority over-sampling technique, was invented in [13] to handle imbalanced training dataset. As indicated by its name, for a given imbalanced dataset, it oversamples the minority class to generate more synthetic data so that the new dataset becomes balanced. It outperforms most other data augmentation methods such as random oversampling which may result in overfitting when duplicated data are used [14].

In this research, SMOTE is implemented to enlarge dataset of every class. In addition to applying transfer learning to lower the requirement for the size of the dataset, training data are augmented with SMOTE. With the combination of the two methods, the gap between the required data and the available data is narrowed so that the classification results can be improved.

SMOTE algorithms linearly interpolate a selected sample and its nearest neighbor. Consider one image x_i among the samples. Firstly, k nearest neighbors (KNN) for x_i within the class are determined by matching the selected image x_i and remaining images in the same class pixel by pixel and comparing their Euclidean distances. All the images are normalized to the same resolution and stored according to RGB image format. In this research, k is selected to be 5. Subsequently, 2 out of the 5 nearest neighbors with x_i are chosen randomly [13]. The difference between the attribute matrix of each selected nearest neighbor and the sample x_i is taken. The difference is then added to the attribute matrix of x_i after multiplying it with the coefficient of interpolation of the line segment joining two samples, termed as weight coefficient w_i, which is a random number between 0 and 1, as shown below:

$$x_i{}' = x_i + w_i \cdot (x_{i(nn)} - x_i) \qquad (1)$$

where $x_i{}'$denotes the new sample synthesized by the original sample x_i and one of its nearest neighbors $x_{i(nn)}$. Eqn. (1) can be rewritten as the following equation:

$$x_i{}' = (1 - w_i) \cdot x_i + w_i \cdot x_{i(nn)} \qquad (2)$$

In Eqn. (2), it is clear that $x_i{}'$ is synthesized from the weighted sum of two images x_i and $x_{i(nn)}$. For convenience, both x_i and $x_{i(nn)}$ are referred to parent images of x_i in this paper.

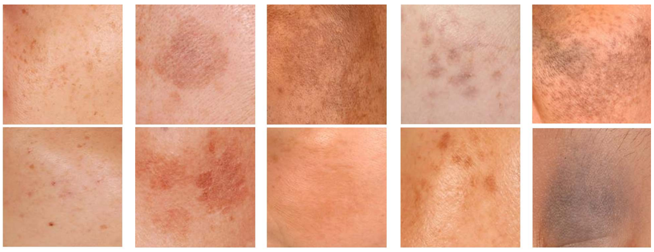
As the result, two new samples, each corresponding to one of the nearest neighbors $x_{i(nn)}$, are synthesized for each image

x_i. The same procedure is repeated for every image in each class of the original image set. A new set of images with twice the number of original images is generated. The respective parent images are also recorded. During the training stage, any new image synthesized with any of its two parent images being a test image during a particular fold of the training stage will be excluded from the corresponding training set. This is to make sure that the training images are truly independent of the testing images.

III. Experiment

A. The original dataset

In this paper, the original dataset of five types of most common non-tumorous facial pigmentation in Asia [20] is adopted. The clinical images are carefully taken and cropped to retain only the region of interest (ROI), mainly the cheeks and lateral planes. The five classes are Freckles, Lentigines, Melasma, Hori's Nevus, Nevus of Ota and each class contains thirty images of different patients. Two images of each class are chosen and displayed in Fig.3 for demonstration.



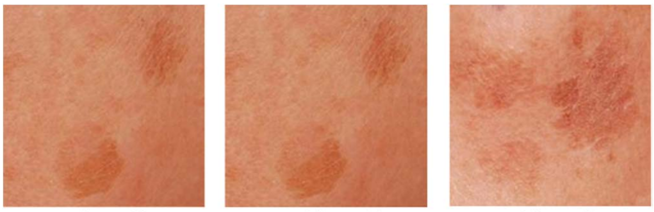
Freckles Lentigines Melasma Hori's Nevus Nevus of Ota

Figure 3. Sample images for each class

B. Data augmentation with SMOTE

In this paper, SMOTE is applied to enlarge the dataset. As stated in Section IIB, SMOTE generates one new image by combining two original samples based on a random weight coefficient. For every image in a 30-sample-class, 2 out of its 5 nearest neighbors are randomly chosen to synthesize two new images with the selected image. 60 new images in total are obtained for each class with a random weight between 0 and 1 for each new image. The respective parent images of the generated images are recorded for training data filtering in Section IIIC.

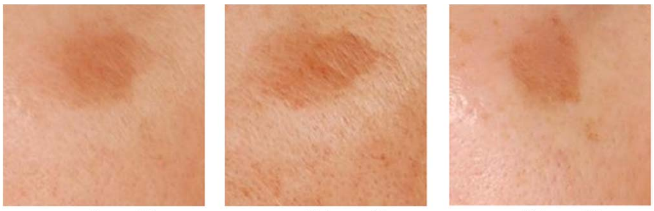
Some examples of generated images using different weight coefficients and their parent images in the Lentigines class are shown in Fig. 4 and Fig. 5.



new image 32 parent1-image 16 parent2-image 30

Figure 4. new image 32 in the class of Lentigines with its parent images

For example, new image 32 is generated from image 16 and 30 using weight coefficient 0.022, which is close to 0. High degree of similarity between image 32 and its parent image 16 is observed in Fig. 4.



new image 22 parent1-image 11 parent2-image18

Figure 5. new image 22 in the class of Lentigines with its parent images

On the other hand, new image 22 is generated from image 11 and 18 using weight coefficient 0.51, which is close to 0.5. In this case, new image 22 is visually more distinct from its parent images but still resembles both in some sense.

For the classification model to be more general and be applicable to real life practice, it is required to be able to classify unpredictable images. Therefore, to correctly evaluate the accuracy of the model, the training set should not contain information of the testing images. As the generated images are similar to their parent images, new images generated with parent images that are later selected as testing images are removed from the training set during the training stage.

C. Model setting for training and testing using Inception-Resnet-v2

As stated above, the pre-trained model Inception-Resnet-v2 is utilized in this research, and parameters like the learning step are tuned. In order to be consistent with the state-of-the-art method in [5] for a fair comparison, the original dataset is randomly divided into ten folds for cross-validation test, with each fold containing three images per class. Each time one fold is chosen as the testing fold, and the rest are added to the training folds. After selecting the testing fold, the indexes of testing images are recorded down and all those generated images derived from parents in the testing images are excluded for training at this stage. The remaining newly generated images are then added to the training set. After finalizing the testing fold and training folds, a grid search using four-fold cross-validation among the training folds is performed for hyper-parameter optimization and early stopping is also performed to avoid overfitting. The criterion for early stopping depends on the four-fold cross-validation, i.e. the training will stop before normal termination if no further improvement of validation performance is observed after 500 iterations. The weights of the processing units in the network are randomly initialized with a Gaussian distribution of zero mean and a standard deviation of 0.001. The learning rate is updated with an exponential decay factor shown in (3):

$$\eta_{adp} = \eta \times decay\_rate^{\left(\frac{step}{decay\_step}\right)} \qquad (3)$$

where η is the learning rate. The decay step is chosen to be 1000. The ten-fold testing is set to run ten independent trials and the results are averaged. In each trial, the testing sets are randomly shuffled in order to produce statistically reliable results.

D. Experiment results and analysis

In this section, the comparison of the classification accuracy between the proposed method of deep learning with enlarged dataset using SMOTE and the state-of-the-art method V-PLDA [5] is presented. The overall classification accuracies and standard deviations for three methods, namely V-PLDA, deep learning with Inception-ResNet-v2 trained on original dataset, and deep learning with Inception-ResNet-v2 trained on the enlarged dataset with generated images by SMOTE are listed in Table I.

TABLE I. RESULTS OF DIFFERENT METHODS

Method	Accuracy %	Standard Deviation
V-PLDA [5]	77.33	0.0982
Inception-ResNet-v2 with original dataset	81.87	0.0889
Inception-ResNet-v2 with SMOTE	84.67	0.0825

Several points can be seen from Table I. i) Performing deep learning with the pre-trained model Inception-ResNet-v2 yields great improvement of about more than 4% compared to the state-of-the-art method V-PLDA with hand-craft features. This improvement is due to the Inception-ResNet-v2 model pre-trained on a much larger dataset and hence having more useful features. ii) By applying SMOTE to generate some synthetic samples, the accuracy further improves by about 3%. This is because the variations of the newly generated images by SMOTE provide more information for training which comes from the original images and thus more specific to the task performed. The overall improvement of more than 7% is very significant and shows promise for practical clinical applications in the near future.

To further compare the performances of the V-PLDA method and the proposed methods (with and without using SMOTE), the confusion matrices for the methods are presented in Tables II, III and IV, respectively. Each of the five classes of pigmentation disorders is represented by a capital letter for simplification: F: Freckles; L: Lentigines; H: Hori's Nevus; O: Nevus of Ota.

TABLE II. CONFUSION MATRIX OBTAINED BY PLDA USING THE ORIGINAL DATASET

		Predicted class				
		F	**L**	**H**	**M**	**O**
True class	**F**	84.7	0.7	8.0	6.7	0.0
	L	7.0	80.7	8.7	3.7	0.0
	H	22.3	5.0	67.7	4.3	0.7
	M	6.0	5.0	0.3	82.0	6.7
	O	9.0	6.7	13.0	3.0	68.3

TABLE III. CONFUSION MATRIX OBTAINED BY INCEPTION-RESNET-V2 USING THE ORIGINAL DATASET

		Predicted class				
		F	**L**	**H**	**M**	**O**
True class	**F**	86.7	6.7	0.0	3.3	3.3
	L	3.3	83.3	10.0	0.0	3.3
	H	0.0	3.3	86.7	3.3	6.7
	M	3.3	3.3	6.7	73.3	13.3
	O	0.0	0.0	3.3	16.7	80.0

TABLE IV. CONFUSION MATRIX OBTAINED BY INCEPTION-RESNET-V2 USING THE ENLARGED DATASET

		Predicted class				
		F	**L**	**H**	**M**	**O**
True class	**F**	90.0	3.3	0.0	3.3	3.3
	L	3.3	90.0	6.7	0.0	0.0
	H	0.0	3.3	90.0	3.3	3.3
	M	6.7	3.3	6.7	76.7	6.7
	O	0.0	6.7	10.0	6.7	76.7

It can be observed from the confusion matrices in the three tables that for most classes (except Melasma), the proposed method shows significant improvement compared to the V-PLDA method. It is worth noting that for two classes that are poorly classified by V-PLDA, Hori's Nevus and Nevus of Ota, the proposed method improves the accuracy by about 10% percent for each class. This improvement is remarkable since it indicates that the pre-trained Inception-ResNet-v2 and the generate images by SMOTE tend to contain more general features of different levels of abstraction and make full use of the images at hand to deal with the large within-class variance despite the limited number of domain-specific images. Therefore, the proposed method can fit distinct textures and patterns of multiple classes and produce better results in general.

IV. CONCLUSION

In this paper, we proposed a new method by combining deep learning and SMOTE for the classification of five most common non-tumorous facial pigmentation disorders in Asia. Pre-trained network Inception-ResNet-v2 is used as the deep neural network for our method and a real-world clinical dataset is used for training and testing. Compared to the state-of-the-art method V-PLDA [5], the overall classification accuracy improves by more than 7%. In particular, for the two classes that had relatively low classification accuracies compared to other classes, Hori's Nevus and Nevus of Ota, the performance is greatly improved. This improvement is of vital importance since it provides an opportunity to enlarge a relatively small domain-specific dataset and make full use of limited information at hand. Transfer learning with the deep neural network is applied to provide more training information from a general dataset. This is significant because this combination of deep learning and data augmentation with domain-specific information also paves the way for further applications in similar fields, which can facilitate early diagnosis and help reduce mistreatment by charlatans or non-professionals. Future work includes further improving the SMOTE to generate more diverse synthesized data, and exploiting other advanced methods for data augmentation, for example, Generative Adversarial Network (GAN).

ACKNOWLEDGEMENTS

We wish to acknowledge the funding support for this project from Nanyang Technological University under the Undergraduate Research Experience on Campus (URECA) program.

References

[1] G. Dougherty, "Image analysis in medical imaging: recent advances in selected examples," Biomedical Imaging and Intervention Journal, vol. 6, no. 3, p. e32, Jul-Sep 2010.

[2] M. J. McAuliffe, F. M. Lalonde, D. McGarry, W. Gandler, K. Csaky and B. L. Trus, "Medical Image Processing, Analysis and Visualization in clinical research," in Proceedings 15th IEEE Symposium on Comuputer-Based Medical Systems. CBMS 2001, Bethesda, MD, USA, 2001.

[3] M. Filho, Z. Ma and J. M. R. Tavares, "A Review of the Quantification and Classification of Pigmented Skin Lesions: From Dedicated to Hand-Held Devices," Journal of Medical Systems, vol. 39, no. 11, pp. 1-12, November 2015.

[4] G. Capdehourat, A. Corez, A. Bazzano and P. Musé, "Pigmented Skin Lesions Classification Using Dermatoscopic Images," in Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 14th Iberoamerican Conference on Pattern Recognition, CIARP 2009, Guadalajara, Jalisco, Mexico, 2009.

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

[6] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition," Pattern Analysis and Applications, vol. 9, no. 2-3, pp. 273-292, October 2006.

[7] F. Bianconi and A. Fernández, "Evaluation of the effects of Gabor filter parameters on texture classification," Pattern Recognition, vol. 40, no. 12, pp. 3325-3335, December 2007.

[8] Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning," Nature, no. 521, pp. 436-444, 27 May 2015.

[9] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115-118, 25 February 2017.

[10] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo and Z. Zhao, "Deep Transfer Learning for Modality Classification of Medical Images," Information, vol. 8, no. 3, p. 91, 2017.

[11] L. D. Nguyen, D. Lin, Z. Lin and J. Cao, "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation," in 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 2018.

[12] V. Sze, Y.-H. Chen, J. Emer, A. Suleiman and Z. Zhang, "Hardware for Machine Learning: Challenges and Opportunities," in 2017 IEEE Custom Integrated Circuits Conference, Austin, Texas, 2017.

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

[14] G. Douzas, F. Bacao and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE," Information Sciences, no. 465, pp. 1-20, 2018.

[15] E. Ramenol, Y. Caballero, R. Bello and F. Herrera, "SMOTE-RSB *: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory," Knowledge and Information Systems, vol. 33, no. 2, pp. 245-265, November 2012.

[16] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," in the Thirty-First AAAI Conference on Artificial Intelligence, San Franciscon, 2016.

[17] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya and R. Wald, "Deep learning applications and challenges in big data analytics," Journal of Big Data, vol. 2, no. 1, pp. 1-21, December 2015.

[18] S. H. Chang, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura and R. M. Summers, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Trans Med Imaging, vol. 35, no. 5, pp. 1285-1298, May 2016.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," in CVPR, Las Vegas, Nevada, 2016.

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