Supplementary Online Content

Han SS, Moon IJ, Lim W, et al. Keratinocytic skin cancer detection on the face using region-based convolutional neural network. *JAMA Dermatol*. Published online December 4, 2019. doi:10.1001/jamadermatol.2019.3807

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Model Training

1-1) Training the Blob Detector

The blob detector was trained with 21,421 images of malignant and benign nodular disorders from the photograph of Asan Medical Center and 103,627 general object images from ImageNet[1]. We trained the blob detector using faster-RCNN (model = VGG-16; https://github.com/rbgirshick/py-faster-rcnn)[2] to suggest possible lesions of interest from clinical images and generate numerous blobs of 256×256 resolution (NMS_THRESHOLD=0.3 and CONF_THRESHOLD=0.1).

1-2) Training the Fine Image Selector

Although the blobs generated by the blob detector mostly contained lesional blobs with adequate quality, a number of inadequate blobs were also included. In addition, some blobs were too blurry to analyze correctly. To define the safe limits for analysis given the input[3,4], we created a fine image selector to classify the generated blobs into 1) fine lesional blob, 2) inadequate blob, 3) nonspecific or normal blob, and 4) general object blob (Figure 1A). Adequate blobs were defined as the ones with a detectable, well-focused skin lesion at the center of the image without any general object. To create the fine image selector, we generated hundreds of thousands of raw blobs from facial images of the entire training dataset using the blob detector. Based on image findings, we manually classified these blobs to 1) 81,030 fine blobs, 2) 59,319 inadequate blobs, 3) 270,224 normal or nonspecific blobs, and 4) 152,493 general object blobs. With these 4-class training images, we fine-tuned SE-ResNet-50 end-to-end.

1-3) Training the Disease Classifier

To reduce false positives for common benign disorders, we needed to acquire as many images of normal and benign disorders as possible. We used the 81,030 fine lesional blobs. After manual annotation process, we obtained 8,280 acne blobs, 10,965 lentigo blobs, 31,469 melanocytic nevus blobs, and 4,705 seborrheic keratosis blobs, all of which comprised the secondary training dataset. The diagnosis of the secondary training dataset was tagged based on image findings.

We trained the temporal disease classifier with primary and secondary datasets, and then, the temporal classifier was used to annotate the tertiary dataset. First, using the blob detector and fine image selector, we generated numerous blobs from all clinical images. The temporal classifier detected possible false-positive blobs and automatically annotated the blobs to 178 classes, which was the tertiary training dataset. We repeated the process of creating the tertiary dataset two times while removing similar or duplicated crops.

For training the final version of disease classifier, we used the primary, secondary, and tertiary training datasets. The disease classifier was trained with 721,749 image crops of 178 disease classes. We fine-tuned SENet-154[5] and SE-ResNeXt-50 end-to-end. The final output was obtained by arithmetically averaging the outputs of the two models.

With NVIDIA Caffe (https://github.com/nvidia/caffe; version 0.17.2, CUDA 10.0, cuDNN 7.6.2), we trained our CNN models using a transfer learning method. We adapted the ImageNet pre-trained model and replaced the last layer with 178 disease class outputs, and all layers of the network were fine-tuned. For data augmentation, original images were cropped

near the central area of the lesion, after which they were rotated, magnified, and shifted. With reference of previous reports[3,5-7], the hyper-parameters were set as follows: learning_rate=0.001, gamma=0.1, weight_decay=0.00001, mini_batch_size=32, solver=SGD, momentum=0.9, total_iteration=90 epoch, and step_iteration=30 epoch for the disease classifier. All images were resized to 224×224, and histogram normalization was performed as a preprocessing step before training the models. For image augmentation, we applied crop, shift, and rotation with the original images, and built a training LMDB (lightning memory-mapped database) which comprised 3,027,198 images. The output values of SE-Net[5] and SE-ResNeXt-50 were arithmetically averaged to obtain a final model output.

Weighting Premalignant Outputs in the Formula of the Malignancy Output

The malignancy output was defined as : “the sum of outputs of malignant disorders” + 0.2 × “the sum of outputs of premalignant disorders”.

Premalignant lesions such as actinic keratosis (AK) need particular attention and close observation. Therefore, lesions highly suspicious of AK (AK output = 1.0) were taken into account and were weighted because they would be otherwise neglected for the end user.

Like physicians evaluated each lesion as “biopsy or not” / “malignancy or not”, the algorithm generated reports in two different values, namely T_{80} / T_{90}. The threshold for T_{80} and T_{90} without adding “0.2 × premalignant” were 0.2278 and 0.4451, respectively. We chose “0.2” as a weight value because “0.2” is close to the threshold of T_{80} (=0.2278) for lesions very much likely to be premalignant. For instance, lesions suspicious of AK with 100% certainty are given

a weighted factor of 0.2, which gave warning message to the user although it is not reported as being malignant (0.2 < 0.2278).

The final threshold for T_{80} and T_{90} when a weighted calculation of 0.2 × premalignant is performed turned out to be 0.2545 and 0.4687, respectively.

eTable 1. Summary of the Primary, Secondary, and Tertiary Training Dataset

	Primary	Secondary	Tertiary
No. of images	182,348	563,066	361,472
No. of disorders	178 disorders	178 disorders and general objects	178 disorders
Diagnosis method	Clinical diagnosis and manual annotation based on image findings	Manual annotation by dermatologist based on image finding	Automatic annotation by algorithm

The primary training dataset consisted of the clinical photographs from the Asan Medical Center, MED-NODE[8], Seven-Point Checklist Dermatology Dataset[9], and images on the internet. We used search engines (google.com and bing.com) to gather 51,459 website images. Just as it was done in the ImageNet study, two dermatologists manually tagged each image with the corresponding diagnosis one by one. In the primary dataset, there were numerous benign lesions. We cropped and annotated them to make a secondary dataset. The tertiary dataset was automatically generated and annotated using the algorithm trained with the primary and secondary datasets.

eTable 2. Locations of malignant lesions in the PS and DER validation datasets.

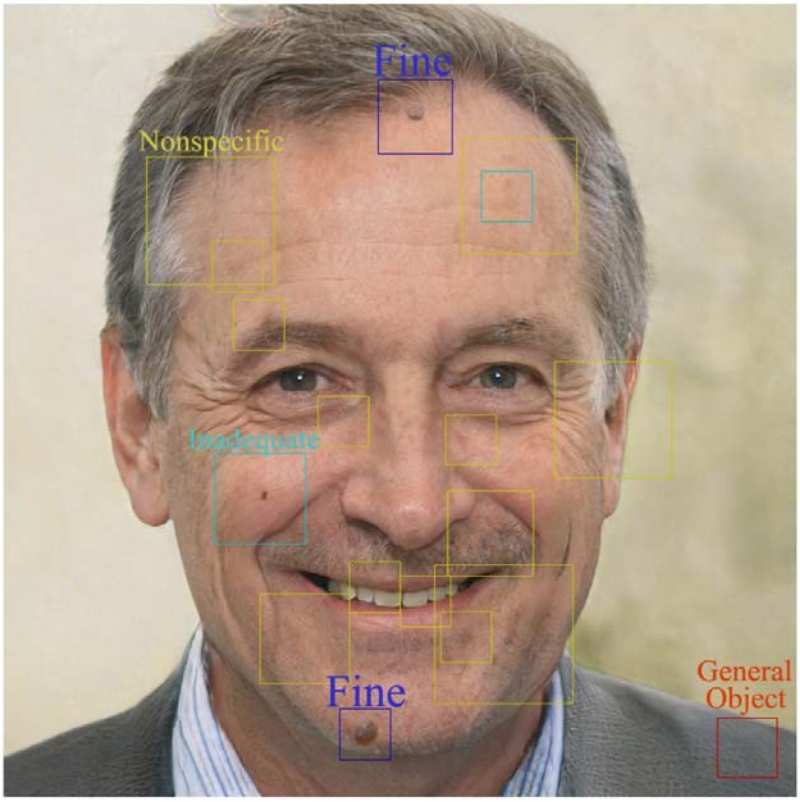
	Basal cell carcinoma			Squamous cell carcinoma			Intraepithelial carcinoma		
	Asan	Hallym	Chonnam	Asan	Hallym	Chonnam	Asan	Hallym	Chonnam
Scalp	6	0	0	1	0	0	1	0	0
Forehead	2	0	0	0	0	3	0	0	0
Periocular	6	4	12	0	0	1	0	0	0
Nose	20	9	7	0	2	9	1	0	0
Lip & chin	3	2	5	1	0	8	0	0	0
Ear	3	0	0	3	0	0	1	0	0
Cheek & temple	12	11	8	10	5	18	11	0	0
Sum	52	26	32	15	7	39	14	0	0

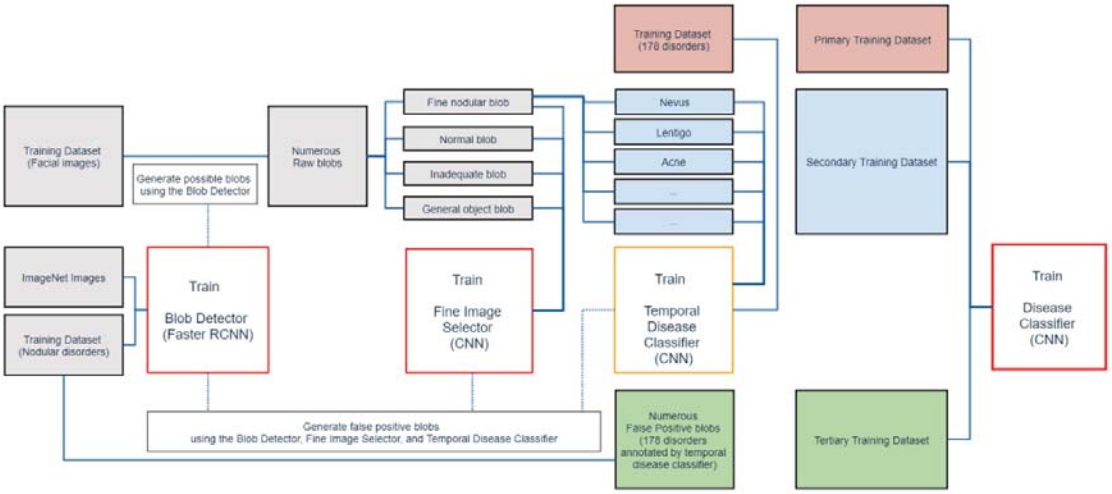
DER dataset consisted of clinical photographs from Asan Medical Center. PS dataset consisted of photographs from Hallym University Hospital and Chonnam National University Hospital.

eTable 3. Devices used to take clinical photographs

	Camera Body	Camera Lens	Light
Asan (room 1)	Nikon D7100	Nikon AF-S DX 18-55mm f/3.5-5.6G ED II	-
Asan (room 2)	Nikon D7000	Nikon AF-S DX 18-55mm f/3.5-5.6G VR	Softbox (Prospot p400 gold)
Hallym	Nikon D5600	Nikon AF-P 18-55mm	Ring flash (Horusbennu MRF-995)
Chonnam (main studio)	Canon EOS-80D	Canon EF24-105mm F4L II	Ring flash (CMR-14X III)
Chonnam (sub studio)	Canon EOS-80D	Canon EF24-70mm F2.8L II	Ring flash (CMR-14X III)

eFigure 1. Representative examples of fine lesional, normal, inadequate, and general object blobs and training procedures of the algorithm

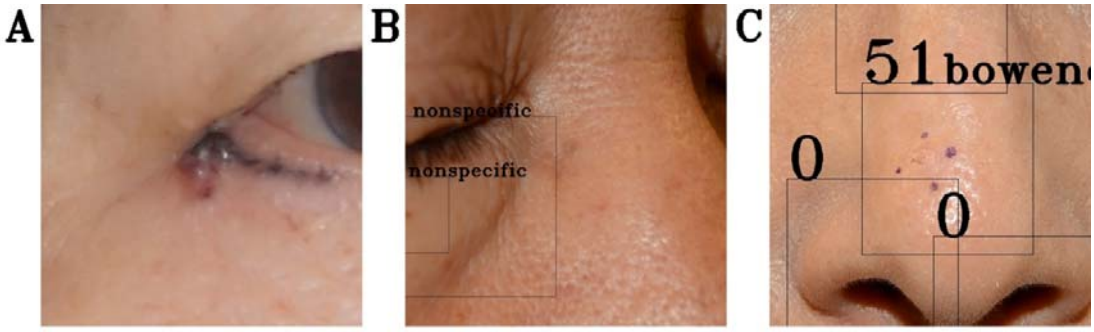




(Upper figure) We generated numerous raw blobs from the entire clinical photograph using the blob detector. And then, we classified those raw blobs as fine lesional, normal, inadequate, and general object blobs.
The image of the face of this man is a fake image generated by style-GAN[10].

(Lower figure) Training Procedures of the Blob Detector, Fine Image Selector, and Disease Classifier.

* Manually annotating the generated blobs among nodular blobs, normal blobs, inadequate blobs, and general object blobs, and then, training a fine image selector using those annotated blobs that amounts to a total of 563,066 blobs.
** Manually annotating common benign disorders (i.e. nevus, seborrheic keratosis, lentigo, and acne) based on image findings from the fine lesional blobs, which will be the secondary training dataset. We trained a temporal disease classifier with both the primary training and the secondary training datasets.
*** Running the blob detector, fine image selector, and temporal disease classifier to generate a tertiary training dataset from all clinical photographs. At this stage, the annotation process was done automatically using the temporal disease classifier. Lastly, training a final version of disease classifier with all training datasets (721,749 images; primary + secondary + tertiary training dataset).



eFigure 2. Representative examples of incorrectly diagnosed cases: “miss”, ”inherent”, and “over”

(A) Basal cell carcinoma (BCC) on the lower eyelid of a 80-year-old female from the PS dataset. Our model did not detect this lesion as well as some of the small lesions close to the lid margin. Overall, 2 cases out of 185 cancer patients were missed because the algorithm could not localize the malignant lesion.

(B) BCC on the right sidewall of the nose of a 70-year-old male from the DER dataset. This morpheaform BCC was clinically inconspicuous and could not be detected inherently by the model.

(C) Seborrheic keratosis on the nasal tip of a 61-year-old female from the DER dataset. Clinically mimicking an actinic keratosis, this lesion was overdiagnosed as Bowen’s disease (Intraepithelial carcinoma) by the model with a malignancy output of 51. This could have resulted from both the high incidence of malignancy on the nose and the skin markings that were not removed at the time of photography[11].

eReferences

1. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision.* 2015;115(3):211-252.
2. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. Paper presented at: Advances in neural information processing systems2015.
3. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PloS one.* 2018;13(1):e0191493.
4. Narla A, Kuprel B, Sarin K, Novoa R, Ko J. Automated Classification of Skin Lesions: From Pixels to Practice. *Journal of Investigative Dermatology.* 2018;138(10):2108-2110.
5. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition2018.
6. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition2016.
7. Keskar NS, Mudigere D, Nocedal J, Smelyanskiy M, Tang PTP. On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836.* 2016.
8. Giotis I, Molders N, Land S, Biehl M, Jonkman MF, Petkov N. MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert systems with applications.* 2015;42(19):6578-6585.
9. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. 7-Point Checklist and Skin Lesion Classification using Multi-Task Multi-Modal Neural Nets. *IEEE Journal of Biomedical and Health Informatics.* 2018.
10. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. *arXiv preprint arXiv:1812.04948.* 2018.
11. Winkler JK, Fink C, Toberer F, et al. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA dermatology.* 2019.

