# Ensembled Skin Cancer Classification (ISIC 2019 Challenge Submission)

Tat Dat Tô, Dinh Thi Lan, Thi Thu Hang Nguyen, Thi Thuy Nga Nguyen, Hoang-Phuong Nguyen, Le Bich Phuong, Tien Zung Nguyen

# Ensembled Skin Cancer Classification (ISIC 2019 Challenge Submission)

To Tat Dat*, Dinh Thi Lan, Nguyen Thi Thu Hang,
Nguyen Thi Thuy Nga, Le Bich Phuong, Nguyen Hoang Phuong, Nguyen Tien Zung

October 28, 2019

**Abstract**

We describe our combined approach to the problem of classication of dermoscopic (and clinical) images of skin lesions for skin cancer detection and classification. This work is part of a research project of Torus Actions, which is supported by fundings from Deep Clinics, a new generation clinic in Vietnam specialized in skin care and skin disorders.

## 1 Introduction

In this paper we briefly discuss our colmmbined approach to the problem of detection and classification of skin cancer based on dermoscopic (and also clinical and macroscopic) images of skin lesion. We include this description in our submission to the ISIC 2019 Challenge (https://challenge2019.isic-archive.com/).

The main ingredients of our approach are:

1. *Divide and conquer*: we divide the n-ary classication problem (n=9, with 9 classes: AK, BCC, BKL, DF, MEL, NV, SCC, VASC, and UNKnown) into many smaller binary and 3-ary problems which are easier to treat.

2. *Data normalization*: we normalize the data (via automatic preprocessing) before feeding them to the classification models.

3. Using recent convolutional neural networks (*CNNs*) and fighting *data imbalance* and *overfitting*: Current state-of-the-art CNNs, e.g. EfficientNet (cf. [4]) and Inception Resnet (cf. [6]) may achieve very high accuracy in training, while the validation results remain poor, due to fast overfitting ('rote learning'). We fought overfitting by a combination of methods, including: using our own loss functions, modifying the off-the-shelf models, and training from the beginning (not using pretrained weights). To fight data imbalance among the classes (when training), we use standard up-sampling techniques, and also a new class of loss functions invented by us.

4. Our new *asymmetric focal binary loss function*, which seems to better fight data imbalance problems than the usual symmetric loss functions. Our loss function as an asymmetry parameter which can be tuned depending on the level of data imbalance.

5. We also build direct 8-class and 9-class classifcation models

6. Ensembling and assembling algorithms.

7. We use metadata to mainly identify images which correspond to a same lesion and which therefor should belong to a same class, via a similarity search algorithm which uses the images and the metadata. This allow us to make the results more consistent.

*Contact author: tat-dat.to@torus-actions.fr

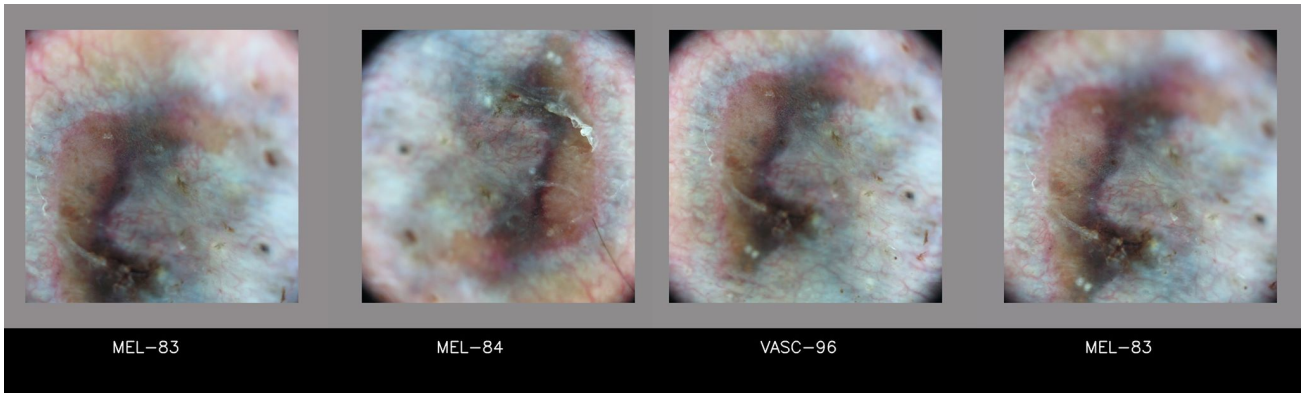


Figure 1: Example of inconsistency: Images ISIC0057830, 0068605, 0072884, 0059197 (normalized here) are apparently of the same lesion, according to our similarity search (which makes use of metadata to simplify the task), but are predicted to belong to different classes (MEL and VASC) by a model.

## 2 Classification submodules and the assembling process

We make two kinds of models:

- Direct 8-class or 9-class classification models (In the case of 8 classes, we use sigmoid activation function and allow the image to not belong to any class)

- Step by step classiciation, and individual binary models (a single class vs. the rest)

Then we combine all of our results together by an assembling and voting algorithm:

Results from a cross-validated 8-class classification serve as the base. Then these results are modified by the results from the oher models, and from the similiraty search and majority voting for images which have more than one class. We give somwhat more priority in voting to under-represented classes (classes with fewer images)

## 3 Data preparation and preprocessing

### 3.1 Datasets

The main daata set used in our traing are provided by the organizers of ISIC 2019 Challange: the HAM10000 dataset [10], the BCN dataset [12], and the MSK dataset [11].

Besides these datasets, we also downloaded from many different sources on the internet additional clinical and dermoscopic images of skin lesions, most notably: DermnetNZ

In total, we added approximatively 3600 dermoscopic images (of 8 classes) to the ISIC2019 datasets for training. We also added a few thousand clinical images to our database, not for the ISIC 2019 Challenge, but for the models that we are developing which can make probabilistic predictions on clinical images as well.

We also created a dataset of a few thousand images for the UNK class by using images of other skin diseases. (This could be and should be improved by adding more diversified pictures).

### 3.2 Data normalization

Each image (for training, validation, and prediction) undergoes the following normalization process before being fed into the classification models:

First, we apply a skin lesion segmentation to it, based on a previous segmentation module that we developed for ISIC 2018 challenge (cf. [1]). This segmentation module is good enough for what we do in the next steps. Based on this segmentation, we create the smallest ractangular box which contains the lesion, determine its center and its size.

Second, by cutting and padding, we transform each image into a square image, whose center is the center of the lesion and whose size (height and width) satisfies several properties: not too small (at least 256 pixels) and not too big (contains the lesion, plus some margin), and is a multiple of 32 (?). We use constant padding for the missing area (if any), by using the average color of the image after cutting but before padding.

Last, we apply a color normalization function (based on whiteness) to our square images.

After this normalization process, each images becomes square, of approriate size, whiteness-balanced, and has the lesion centered. We will call them *normalized images* (see Figure **??**).

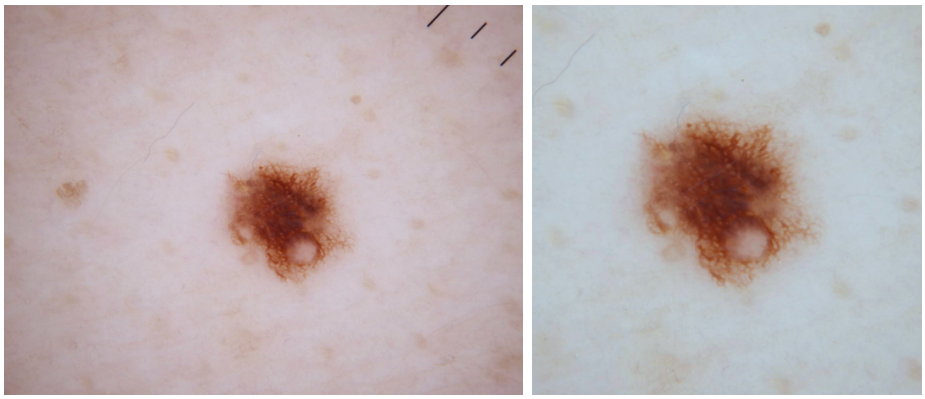

Figure 2: Before (left) and after (right) normalization

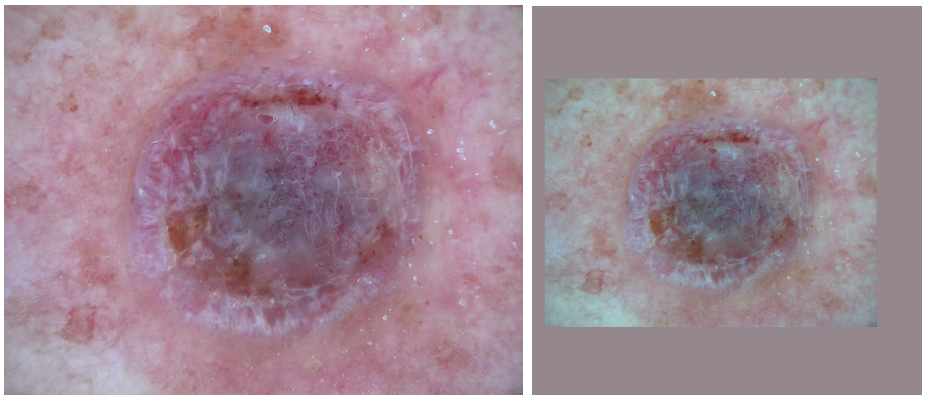

Figure 3: Before (left) and after (right) normalization

### 3.3 Data augmentations

For training, we use some standard data augmentations, including random resizes, cuts, flips, rotations, nonlinear distorsions, noises, and mixups.

For validation and prediction, we use only rotations by angles which are multiple of $\pi/2$, and small translations (to create 16 images out of each image, and then use average voting on them). All normalized images for validation and testing are resized to $380 \times 380$.

The random mixup process [9] (used in some part of our training) creates an image $img = \alpha * img_1 + (1-\alpha) * img_2$ out of a couple of images $(img_1, img_2)$, where $\alpha$ is a small number, say between 0 and 0.1. The class of $img$ is that of $img_2$.

### 3.4 Data balancing

Since the data is extremely imbalanced, for training we use the up-sampling data balancing technique by multiplying the data with different ratios on different classes.

### 3.5 Use of metadata

Since each single lesion may appear many times in the training datasets provided by ISIC2019 (up to 15 images for a same lesion?), there is a real danger of inflated validation scores if there is an intersection between the validation set and the training set on the level of lesions. So we had to use metadata to make sure that there is no lesion-level intersection between these sets.

### 3.6 Validation sets for cross-validation

We create three non-intersecting validation sets for the purpose of cross-validation. (Each model will be run three times, one for each validation set). This cross validation is necessary in order to improve the results, otherwise we would miss a large chunk of data for training.

## 4 The loss functions

For n-ary classification (sub)modules, we used categorical cross entropy. This entropy function is symmetric among the class, and while it is good enough in many cases, does not contained any parameter which can be tuned.

Our new binary loss function, which is used in most of our submodules, is focal (it is mostly concentrated on the wrongly predicted cases), and is asymmetric, with a paparemter which can be tuned for optimization, depending on the level of data imbalance (to be used together with data rebalancing).

We are also doing theoretical research on our (these and other) new loss functions. We believe that the choice of loss functions is very important for the training of models.

## 5 Remarks

Some classes are not really unknown (e.g., Bowen disease), but the organizers did not give any instructions on where t put them among the 9 classes.

Due to a teachnical problem, we didn't have time for cross validation in our first submission. We will do it in our second submission (using metadata)

## 6 Other approach

For only-image task, we also use a direct 8-classes classification model based on the EfficientNet-B4 (cf. [4]) to treat the problem. We use only internal data to train this model.

**Model.** We use the EfficientNet-B4 (cf. [4]) which is trained on 5-fold.

**Preprocessing.** We first randomly flipped images along both dimensions with a probability of $p = 0.5$. Furthermore, we distorted the images with random changes in brightness and saturation. We normalized images by subtracting dataset's image mean. Last, we resized a fixed size of 380x380 for EfficientNet-B4 model.

We tested random rotations, mix-up ([9]), cutout ([13]) without an improvement in performance in this approach.

**Training.** Class-weighted cross entropy loss are selected as the loss function so as to punish harder on the false predictions on those classes with smaller datasets.

We use a batch size of 128 by using the Accumulated Gradients. Adam, Nadam (Adam + Nesterov momentum) and AdamW (Adam with weight) are tested in terms of optimizers in this approach. Finally, we chose AdamW which got the best convergence. In terms of learning rate schedule, we follow a stepwise approach. We first chose a starting learning rate of $lr = 0.0005$ and started reducing it with a factor of $\lambda = 0.9$ after 2 epochs. Then we continued Cyclical Learning Rates in last 20 epochs (cf. [14]).

**Post-processing.** From the fact that we do not have any image of 'UNK' class in the training dataset, we then valided 0-valued for "UNK"-class except image of ID: **"ISIC0035068"**.

For metadata task, we use a direct 9-classes classification model based on Inception-Resnet with external data for the class UNK.

# References


[1] To Tat Dat, Nguyen Thi Thu Hang, Lu Quoc Hung, Nguyen Thi Thuy Nga, Nguyen Tien Zung, *Skin Lesion Analysis Towards Melanoma Detection for ISIC 2018*, Toulouse Symposium on Deep Learning (October 18-19 2018), hal-01847743,

[2] Kingma, Diederik P. and Ba, Jimmy. Adam, *A Method for Stochastic Optimization*, arXiv:1412.6980 [cs.LG], December 2014.

[3] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern, *Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)*, 2017; arXiv:1710.05006.

[4] Mingxing Tan and Quoc V. Le. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. ICML 2019.

[5] Philipp Tschandl, Cliff Rosendahl, Harald Kittler, *The HAM10000 Dataset: A Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions*, 2018; arXiv:1803.10417.

[6] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A. Alemi: *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*, AAAI 2017: 4278-4284

[7] Olaf Ronneberger, Philipp Fischer, Thomas Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234–241, 2015, arXiv:1505.04597 [cs.CV]

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, *Deep Residual Learning for Image Recognition*, arXiv:1512.03385 [cs.CV]

[9] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz, *mixup: Beyond Empirical Risk Minimization*, arXiv:1710.09412 [cs.LG]

[10] Tschandl P., Rosendahl C. Kittler H. *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions.* Sci. Data 5, 180161 doi.10.1038/sdata.2018.161 (2018)

[11] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, Allan Halpern, *Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)*, 2017; arXiv:1710.05006.

[12] Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, Josep Malvehy *BCN20000: Dermoscopic Lesions in the Wild*, 2019; arXiv:1908.02288.

[13] Terrance DeVries, Graham W. Taylor, *Improved Regularization of Convolutional Neural Networks with Cutout*, 2017; arXiv:1708.04552.

[14] Leslie N. Smith *Cyclical Learning Rates for Training Neural Networks*, 2017; arXiv:1506.01186.
Search

Hang Nguyen and Thi Thuy Nga Nguyen: *LAAS-CNRS Laboratoire d'analyse et d'architecture des systèmes, 7 Avenue du Colonel Roche, 31400 Toulouse*, {nttnga, hangntt}@torus-action.fr

LE Bich Phuong: *Department of Mathematics, Faculty of Basic Sciences, Hanoi University of Mining and Geology, Hanoi, Vietnam.* ltbphuong@torus-action.fr

Tien Zung Nguyen, Hoang Phuong Nguyen and Tat Dat Tô : *Institut Mathématiques de Toulouse, Université Paul Sabatier, 31062 Toulouse cedex 09, France*, {ntzung,nhphuong, tat-dat.to}@torus-action.fr

Tat Dat Tô: *Ecole Nationale de l'Aviation Civile, Unversite Federale de Toulouse 7, Avenue Edouard Belin, FR-31055 Toulouse, France.*

Dinh Thi Lan: *Torus Actions SAS, 24 rue Delmas, 31400 Toulouse, France*, dtlan@torus-actions.fr