Łódź, 22.06.2021

# Classification of Dermatological Asymmetry of the Skin Lesions using Pretrained Convolutional Neural Networks

Michał Beczkowski, Norbert Borowski, Piotr Milczarski

FACULTY OF PHYSICS
AND APPLIED INFORMATICS
University of Lodz

# Agenda

1. Dermoscopy Methodologies
2. Dermoscopic Datasets
3. Invariant Dataset Augmentation
4. PH2 image set preparation
5. Pretrained Convolutional Neural Networks features
6. Results
7. Conclusions

## Three-Point Checklist of Dermoscopy (3PLCD)

The 3PLCD algorithm relies on the following criteria:

- **Asymmetry in structure and/or in color in one or two axis of the lesion**. The contour shape of the lesion does not impact on the symmetry.
- Atypical network, defined as pigmented network with thickened lines and irregular distribution.
- Blue-white structures, namely any white and/or blue color visible in the lesion, including blue-white veil, scar-like depigmentation, and regression structures such as peppering.

## ABCD rule

ABCD stands for:

- **asymmetry**,
- border (not well-defined, irregular),
- color (more than one shade),
- diameter (usually larger than 6mm)
- evolution (changing features over time).



**Asymmetry of the lesion is one of the common characteristics of skin damage that can be noticed visually**

# Dermoscopic Datasets

- PH2 – (Mendonca et. al.) 200 images, bmp, 768x568
- Derm CS, Derm7pt – (Argenziano et. al.) 1011 cases, jpg, 768x512
- ISIC – more than 24000 cases

**PH2 dataset - contains a total of 200 dermoscopic images of melanocytic lesions,** including 80 common nevi, 80 atypical nevi, and 40 melanomas. **The PH2 data-base includes medical annotation of all the images** namely medical segmentation of the lesion, clinical and histological diagnosis and the assessment of several dermoscopic criteria (colors; pigment network; dots/globules; streaks; regression areas; blue-whitish veil)

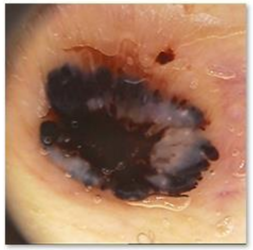
# Invariant Dataset Augmentation

The PH2 database contains 117 fully symmetric, 31 symmetric in one axis and 52 fully asymmetric images of skin lesions. In order to use this database in our research, we had to increase the number of images while minimalizing possible influence on the pixel distribution.

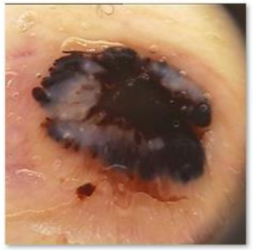
To create new images, various geometric transformations that do not change the asymmetry of shape, shade and structure distribution were used :

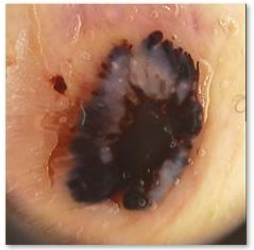
- rotations by 90°, 180° and 270°,
- mirroring on the vertical and horizontal axis,
- 90° rotation of the images after mirroring.

**In total, we got seven transformations for each image that did not change the pixels, shape, or color distribution**.


IMD168


IMD168_obr90


IMD168_obr180


IMD168_obr270


IMD168_odbPIO


IMD168_odbPOZ

## PH2 image set preparation

The next step in preparing the database was to scale the images to the input sizes required by the selected networks:

- scaling the shorter dimension of images (in our case, height) to the input size, e.g. 224px using the Bicubic Sharper algorithm in Photoshop
- Cropping all images to square shape

**The dataset prepared in this way contains 936 fully symmetric, 248 symmetric in one axis and 416 fully asymmetric images of skin lesions**

## PH2 division

To selected networks, two databases were added: training and testing. Both databases were created by dividing the augmentation PH2 dataset into two sets in the following proportions 75% training and validation and 25% testing.
The division was carried out so that the original images and their copies were in one set.

To check whether increasing the database with image copies obtained after rotations and mirroring gives better results, the tests were carried out on the original PH2 database file.

| Number of images | Original PH2 dataset | | | Invariant dataset augmentation | | |
|---|---|---|---|---|---|---|
| | Total | Train | Test | Total | Train | Test |
| Fully Symmetric | 117 | 88 | 29 | 936 | 704 | 232 |
| Symmetric in 1 axes | 31 | 23 | 8 | 248 | 184 | 64 |
| Fully Asymmetric | 52 | 39 | 13 | 416 | 312 | 104 |
| Total | 200 | 150 | 50 | 1600 | 1200 | 400 |

# Pretrained Convolutional Neural Networks features

We used pretrained networks in our research because they are trained on the ImageNet database.
Since in each pretrained network the last three layers are configured to classify 1000 classes, we separated all but the last three layers and replaced them so that the networks would classify images into 3 classes. Due to this method and 3PCLD, **the networks classified the images as symmetrical, symmetrical in one axis, and asymmetrical**.

| Network | Depth | Size [MB] | Parameters [Millions] | Image input size | Average accuracy [ %] |
|---|---|---|---|---|---|
| VGG19 | 19 | 535 | 144 | 224x224 | 70 |
| Xception | 71 | 85 | 22.9 | 299x299 | 80< |
| Inception-ResNet-v2 | 164 | 209 | 55.9 | 299x299 | 80 |

## Confusion matrix

The networks were tested 5 times on each pair of training, validation and testing sets. The resulting networks are saved for future testing and analysis of the results.
For each CNN parameters such as accuracy, true positive rate were defined and calculated according to:

$$ACC = (TP + TN) / N$$

$$TPR = TP / (TP + FN)$$

$$w.\ ACC = (TPR_0 + TPR_1 + TPR_2)/ 3$$

$$FPR = FN / (FP + TN)$$

$$F1 = 2TP / (2TP+FP+FN)$$

$$MCC = (TP*TN - FP*FN) / \sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}$$

# Results

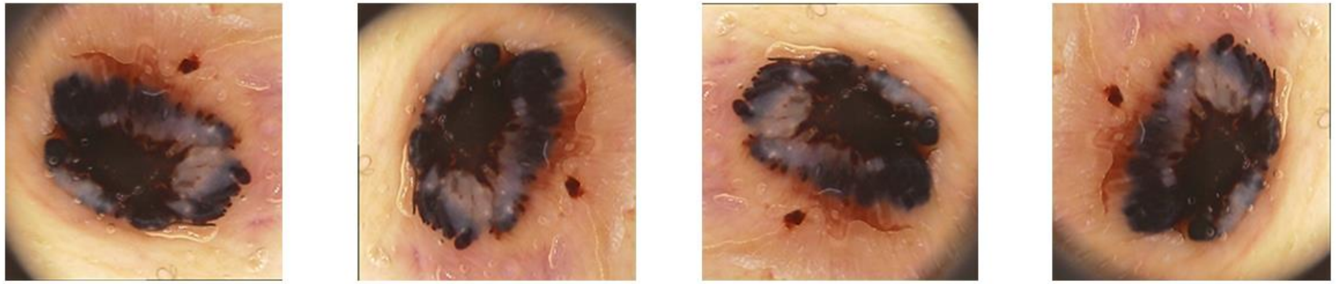
The results were analyzed in three ways:

1. T1 - networks tested on a subset of original images;
2. T8 - networks tested on the original set and its seven copies;
3. IDA - networks tested on the original set and its seven copies but in the worst-case scenario, i.e. if one of the 8 copies of the images has been recognized as asymmetric, all its copies have been classified as asymmetrical.

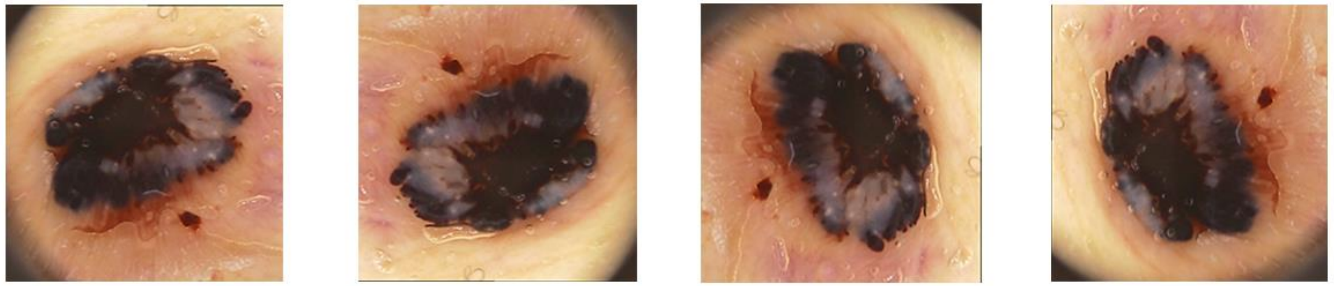
## Results

**The classifications results for the asymmetry**. The chosen confusion matrix factors true positive rate for full asymmetry (TPR0), true positive rate for symmetry in one axis (TPR1), true positive rate for full symmetry (TPR2), false positive rate for full asymmetry (FPR0), false positive rate for symmetry in one axis (FPR1), false positive rate for full asymmetry full symmetry (FPR2) with their average (AVG), variance (VAR), minimum (MIN) and maximum (MAX) values for the chosen CNN network.

| CM factor | | VGG19 | | | Xception | | | Inception-ResNet-v2 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | T1 | T8 | IDA | T1 | T8 | IDA | T1 | T8 | IDA |
| $TPR_0$ [%] | AVG | 58.8 | 60.2 | 71.9 | 65.4 | 67.1 | 80.4 | 53.1 | 55.9 | 70.4 |
| | VAR | 8.4 | 6.8 | 10.4 | 13.1 | 9.6 | 12.3 | 7.9 | 4.4 | 5.7 |
| | Min | 46.2 | 49.0 | 53.8 | 38.5 | 52.9 | 61.5 | 38.5 | 49.0 | 61.5 |
| | Max | 69.2 | 74.0 | 92.3 | 84.6 | 80.8 | 92.3 | 69.2 | 67.3 | 84.6 |
| $TPR_1$ [%] | AVG | 26.9 | 26.7 | 33.1 | 25.6 | 24.4 | 21.9 | 7.5 | 10.3 | 13.1 |
| | VAR | 15.3 | 12.9 | 9.3 | 22.0 | 16.3 | 12.7 | 8.5 | 10.0 | 9.5 |
| | Min | 12.5 | 12.5 | 12.5 | 0.0 | 6.3 | 0.0 | 0.0 | 0.0 | 0.0 |
| | Max | 62.5 | 57.8 | 50.0 | 62.5 | 53.1 | 37.5 | 25.0 | 31.3 | 25.0 |
| $TPR_2$ [%] | AVG | 80.0 | 80.9 | 70.7 | 82.2 | 83.0 | 66.4 | 83.6 | 82.5 | 62.8 |
| | VAR | 11.8 | 10.6 | 13.0 | 7.1 | 5.7 | 7.2 | 5.7 | 4.7 | 9.9 |
| | Min | 58.6 | 61.2 | 44.8 | 72.4 | 75.9 | 55.2 | 75.9 | 73.3 | 44.8 |
| | Max | 96.6 | 95.3 | 86.2 | 93.1 | 91.4 | 75.9 | 89.7 | 89.2 | 75.9 |
| $FPR_0$ [%] | AVG | 9.2 | 8.9 | 15.9 | 11.6 | 11.1 | 25.0 | 10.8 | 10.9 | 23.4 |
| | VAR | 4.7 | 3.8 | 6.3 | 3.2 | 4.2 | 5.8 | 3.3 | 2.1 | 5.8 |
| | Min | 0.0 | 2.0 | 2.7 | 5.4 | 6.1 | 16.2 | 5.4 | 8.1 | 13.5 |
| | Max | 16.2 | 14.5 | 2.4 | 18.9 | 17.2 | 2.4 | 16.2 | 15.5 | 4.8 |
| $FPR_1$ [%] | AVG | 11.1 | 11.4 | 14.9 | 10.4 | 8.7 | 12.6 | 9.6 | 8.7 | 16.5 |
| | VAR | 9.8 | 8.7 | 11.6 | 7.8 | 6.4 | 9.3 | 7.3 | 5.3 | 9.9 |
| | Min | 0.0 | 2.1 | 2.4 | 0.0 | 3.3 | 2.4 | 0.0 | 2.1 | 4.8 |
| | Max | 33.3 | 29.5 | 35.7 | 23.8 | 20.2 | 28.6 | 23.8 | 18.8 | 33.3 |
| $FPR_2$ [%] | AVG | 42.6 | 40.5 | 25.5 | 33.1 | 35.7 | 19.0 | 48.6 | 49.0 | 28.6 |
| | VAR | 9.6 | 8.4 | 9.4 | 8.2 | 11.7 | 11.7 | 6.7 | 7.2 | 9.9 |
| | Min | 23.8 | 25.0 | 9.5 | 14.3 | 16.1 | 0.0 | 38.1 | 37.5 | 9.5 |
| | Max | 61.9 | 54.8 | 38.1 | 42.9 | 48.2 | 33.3 | 66.7 | 63.7 | 42.9 |



IMD168 IMD168_rot90 IMD168_rot180 IMD168_rot270



IMD168_refVer IMD168_refHor IMD168_refVer_rot90 IMD168_refHor_rot90

| IMD168 image version | Classification probability by VGG19 | | | Classification by IDA method |
|---|---|---|---|---|
| | 0 | 1 | 2 | |
| original | 0.0131 | 0.9593 | 0.0276 | 0 |
| rot 90 | 0.5501 | 0.4017 | 0.0482 | 0 |
| rot 180 | 0.9420 | 0.0407 | 0.0173 | 0 |
| rot 270 | 0.3044 | 0.6713 | 0.0243 | 0 |
| mir Vert | 0.3781 | 0.6170 | 0.0050 | 0 |
| mir Vert rot 90 | 0.4937 | 0.4964 | 0.0099 | 0 |
| mir Hor | 0.0543 | 0.9403 | 0.0054 | 0 |
| mir Hor rot 90 | 0.6154 | 0.3441 | 0.0405 | 0 |

## Results

**The classifications results for the asymmetry.** The chosen confusion matrix factors weighted accuracy (w.ACC) with their average (AVG), variance (VAR), minimum (Min) and maximum (Max) values and weighted F1 score (w.F1), weighted Matthews correlation coeffi-cient (w.MCC) for the chosen CNN network.

| CM factor | | VGG19 | | | Xception | | | Inception-ResNet-v2 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | T1 | T8 | IDA | T1 | T8 | IDA | T1 | T8 | IDA |
| w.ACC [%] | AVG | 55.2 | 56.0 | 58.6 | 57.8 | 58.1 | 56.2 | 48.1 | 49.6 | 48.8 |
| | VAR | 8.3 | 7.1 | 6.2 | 11.6 | 8.7 | 8.4 | 4.1 | 4.2 | 5.7 |
| | Min | 39.1 | 41.9 | 44.7 | 41.1 | 48.5 | 43.1 | 40.7 | 43.4 | 38.0 |
| | Max | 69.2 | 68.9 | 68.3 | 78.9 | 73.3 | 68.6 | 54.8 | 56.8 | 57.0 |
| w.F1 | AVG | 0.543 | 0.548 | 0.555 | 0.560 | 0.560 | 0.511 | 0.446 | 0.465 | 0.439 |
| | VAR | 0.093 | 0.078 | 0.071 | 0.126 | 0.098 | 0.085 | 0.047 | 0.059 | 0.063 |
| | Min | 0.370 | 0.402 | 0.423 | 0.403 | 0.465 | 0.397 | 0.389 | 0.392 | 0.323 |
| | Max | 0.673 | 0.682 | 0.678 | 0.801 | 0.739 | 0.634 | 0.522 | 0.580 | 0.547 |
| wMCC | AVG | 0.400 | 0.324 | 0.438 | 0.463 | 0.363 | 0.424 | 0.318 | 0.249 | 0.309 |
| | VAR | 0.136 | 0.114 | 0.099 | 0.130 | 0.121 | 0.114 | 0.073 | 0.071 | 0.086 |
| | Min | 0.061 | 0.101 | 0.264 | 0.231 | 0.223 | 0.280 | 0.135 | 0.158 | 0.170 |
| | Max | 0.578 | 0.524 | 0.621 | 0.732 | 0.650 | 0.643 | 0.429 | 0.424 | 0.449 |

## Conclusions

From our research we have chosen the best CNN networks:

- VGG19 - true positive rate for the asymmetry 84.62%, weighted accuracy 68.29%, F1 score 0.682 and Matthews correlation coefficient 0.581;
- Xception - true positive rate for the asymmetry 92.31%, weighted accuracy 67.41%, F1 score 0.646 and Matthews correlation coefficient 0.533;
- Inception-ResNet-v2 - true positive rate for the asymmetry 53.85%, weighted accuracy 51.57%, F1 score 0.528 and Matthews correlation coefficient 0.295.

Our original approach using the defined Invariant Dataset Augmentation shows that the classification characteristics like accuracy and true positive rate as well as the F1 and MCC tests can be much higher (5-20%) than using only original images.

Łódź, 22.06.2021

# Thank you for your attention!

## Classification of Dermatological Asymmetry of the Skin Lesions using Pretrained Convolutional Neural Networks

Michał Beczkowski, Norbert Borowski, Piotr Milczarski

FACULTY OF PHYSICS
AND APPLIED INFORMATICS
University of Lodz