Increasing Skin Lesions Classification Rates using Convolutional Neural Networks with Invariant Dataset Augmentation and the Three-Point Checklist of Dermoscopy

Piotr Milczarski[1][0000−0002−0095−6796],
Norbert Borowski[1][0000−0002−3861−2411],
Michał Beczkowski[2][0000−0001−5220−138X]

[1]*Lodz University of Technology*
Institute of Information Technology
Politechniki 8, 93-590 Łódź, Poland
{piotr.milczarski, norbert.borowski}@p.lodz.pl
[2]*University of Lodz*
Faculty of Physics and Applied Informatics
Pomorska str. 149/153, 90-236 Łódź, Poland
michal.beczkowski@uni.lodz.pl

DOI:10.34658/9788366741928.52

Abstract. *In the paper, we show how to tackle the problem of lack of the rotation invariance in CNN networks using the authors' Invariant Dataset Augmentation (IDA) method. The IDA method allows to increase the classification rates taking into account as an example the classification of the skin lesions using a small image set. In order to solve the problem of the lack of rotation invariance, IDA method was used and the dataset was increased in an eightfold and invariant way. In the research, we applied the IDA methods and compared the results of VGG19, XN and Inception-ResNet-v2 CNN networks in three skin lesions features classification defined by well-known dermoscopic criterions e.g. the Three-Point Checklist of Dermoscopy or the Seven-Point Checklist. Due to Invariant Dataset Augmentation, the classification rate parameters like true positive rate by almost 20%, false positive rate as well as the F1 score and Matthews correlation coefficient have been significantly increased opposite to type II error that has significantly decreased. In the paper, the confusion matrix parameters result in: 98-100% accuracy, 98-100% true positive rate, 0.0-2.3% false positive rate, tests F1=0.95 and MCC=0.95. That general approach can provide higher results while using CNN networks in other applications.*

Keywords: *invariant dataset augmentation, dermoscopic images, blue-white veil, lesion symmetry, convolutional neural networks, artificial intelligence*

1. Introduction

The authors' motivation is to enhance screening methods by raising the classification probability of the positive features that point to the possible health problems. That general approach can provide higher results while using CNN networks in other disciplines not only dermatology. The results vary for different CCN networks but they can be used in a support system for the general practitioners.

The problem of melanoma and other skin illnesses is the fourth vital death problem in society according to European Cancer Information System (ECIS) [1] and American Cancer Society (ACS) [2].

In dermoscopy, the Three-Point Checklist of Dermoscopy (3PCLD)[3] and Seven-Point Checklist (7PCL) [4] is defined and it is proved to be a sufficient screening method in the skin lesions assessments during the checking by dermatology expert. In the 3PCLD method, there are three criteria: asymmetry of shape, hue and structure distribution within the lesion with discrete value either 0, 1 or 2; blue-white structures and atypical pigment network. Apart from 3PCLD and 7PCL there is the ABCD rule [5] that has its own methodology of the skin lesion assessment.

In the problem of data acquisition and data analysis, the feature extraction or classification should not depend on the object position within the image. Let us imagine that our sensor e.g. camera matrix can acquire the image with the features depending on the object rotation. That is why, we propose to use while testing by machine learning methods e.g. CNN networks, the original image and its invariant copies.

The rotation-invariance property is the network's capability to predict the class regardless of the object's orientation [6, 7]. Generally the CNN networks have problem with the rotation invariance and the lack of rotation-invariance is regarded as their weakness [7]. There are some solutions proposed to tackle the problem and provide the CNN networks and deep learning applications this property [8, 9, 10].

In computer science, there are feature-based and appearance-based approaches [4, 11, 12, 13, 14, 15, 16, 17, 18] that show around 70-95% accuracy in feature accuracy of the skin lesions feature assessments.

In the paper, we present how to tackle the problem of rotation invariance with the proposed Invariant Dataset Augmentation (IDA) [14, 15, 19] method. The IDA approach is used not only for data augmentation but also as a proliferation of the validation and test datasets and that results in the eightfold classification of the same image as a set of bytes but different due to its topology. The IDA method shows that the classification results and their confusion matrix parameters e.g. accuracy and true positive rate as well as the F1 test and MCC can be much higher than in the case when only original images are used as a test set. To show the advantages of the IDA method we have prepared the images from the PH2 dataset [20]. After preparing the dataset, several neural networks have been built with the help of the available VGG19, Xception, and Inception-ResNet-v2 pretrained CNN networks. The networks are trained, validated and tested on the original images taken from the PH2 [20] datasets.

In the research, we achieved for the blue-white veil/symmetry classification using 3PCLD approach around 78-87% for the original PH2 dataset. But, the true positive rate (the blue-white veil is present) was between 65-80% in comparison to the true negative rate (the blue-white veil is absent) that varied from 95 to 98%. The reason for that is the fact that there are only around 20% of the images with the blue-white veil feature in the PH2 [20] dataset. However, after using the IDA approach, the achieved in the research average accuracy is around 94 % but the weighted accuracy is around 90% for the blue-white veil, while the sensitivity is even 20% higher than in a standard approach and reached even 98-100% for the images from PH2 dataset.

The paper is organized as follows. In Section 2 the Invariant Dataset Augmentation (IDA) is shown. The screening methods and dermoscopic datasets are discussed in Section 3. The research and its results are presented in Section 4. Finally, Section 5 shows the conclusions.

2. Invariant Dataset Augmentation

The Invariant Dataset Augmentation has been described in [14, 15, 19] and its foundation are based on the idea of how to make the available image dataset bigger as much as possible not changing the data within (i.e. pixels) and its order and relation to other data saved in pixels using the available geometrical invariant transformations. These transformations do not change the asymmetry of shape, hue, and structure distributions, as well as other features that are in the original copy of the image that is taken into account. Fig.1 shows the example of the original image IMD168 from PH2 set.

While data augmentation authors often use a rotation of scaling the image, but while looking for very fragile features in the images, these features can vanish after that procedure e.g. blue-whitish veil. We have chosen seven image transformations: rotation by 90°, 180° and 270°, mirror reflection by a vertical and horizontal axis of the images and their rotations by 90°. In the previous publications [14, 15, 19] we have shown that pretrained CNN networks have not rotation invariant property.

3. IDA and PH2 Dataset

In dermoscopy, the Three-Point Checklist of Dermoscopy (3PCLD) [3] is defined and regarded as a sufficient screening method. The dermatology experts proved it is a sufficient screening method of skin lesions assessments. There are

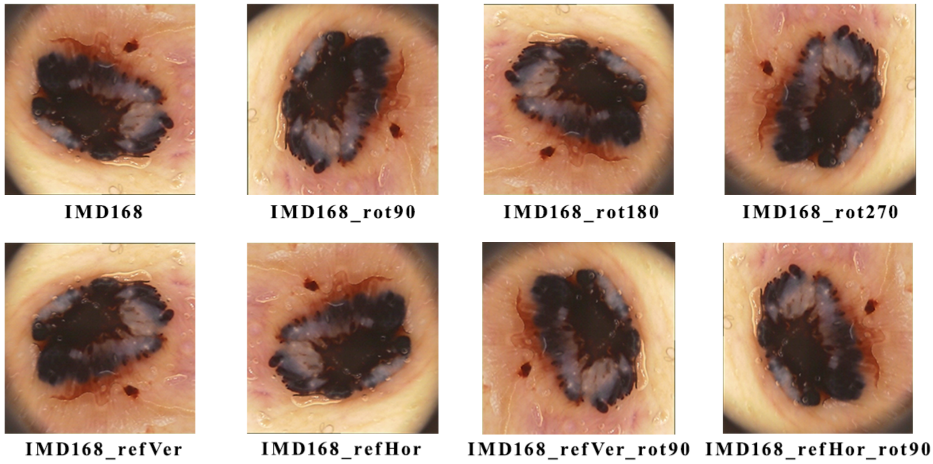

Figure 1. The image IMD168 from PH2 [20] and its invariant copies. Source: own work.

three criteria in the 3PLCD method: asymmetry of shape, hue, and structure distribution within the lesion with discrete value either 0, 1, or 2; blue-white structures and atypical pigment network. The Seven-Point Checklist (7PCL) [4] is another screening method used in skin lesions assessment. In the 7PCL case, dermatology experts examine pigment network, blue-white veil, streaks, pigmentation, regression structures, dots and globules, and vascular structures as the important assessment factors. The ABCD and ABCDE rules [5] are examples of the screening methods that take into account the symmetry/asymmetry, border, color, diameter, and evolving of the lesion.

These transformations used by IDA do not change the pixels, they are pixel invariant, mutually unambiguous, and reversible. Seven new copies for each image are achieved. Altogether, we have 1600 images out of 200 based on the PH2 dataset with:

- 288 images with present blue-white veil and 1312 without a blue-white veil;
- 936 fully symmetric images, 248 symmetric in 1 axis and 416 fully asymmetric;
- 928 images with atypical pigment network and 672 with typical one.

4. Results

In order to show the advantages of IDA method we have compared this methodology with training and testing on the same image sets several times using the IDA transformed images. The set was split into: 75% of the images for the train and validation set; 25% for the test one which. The train set has IDA copies only. Some of the results were presented i.e. symmetry property in [15] and blue-white veil in [14, 19].

The research has been conducted using Matlab 2022b with up-to-date versions of Deep Learning Toolbox™. To achieve a high validation accuracy i.e. 100% or close and a low validation loss i.e. around 0.5 or less we have trained the networks using 30 epochs. The higher values of the epochs have not changed the validation accuracy and the loss as well as the testing values of the accuracy, true positive rate, etc. To achieve the highest classification rates we have tested the wide variety of the following parameters: 30-100 epochs; learning rate from 5e-5 to 1e-2. The times of training have varied from around 12 minutes for VGG19, 40 minutes for XN, and 60 minutes for IRN2 for PH2 dataset.

The confusion matrix factors are used for the chosen CNN network and they are calculated according to the well-known equations [10, 19]. In the T1 and T8 columns, the results of the tests are shown using each image obtained from the IDA procedure as a separate one. In Tables 1 and 2 the classification results for the blue-white veil training on PH2 dataset using IDA transformations with the test results for a single original image set (T1), augmented dataset with 8 copies of the original file (T8), and using the worst-case scenario (IDA) classification are presented. The chosen confusion matrix factors: true positive rate (TPR), false positive rate (FPR) with their average (AVG), variance (VAR), minimum (MIN) and maximum (MAX) values and F1 score and Matthews correlation coefficient (MCC) for the chosen CNN network. 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).

The original authors' approach 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. In the paper for the reaching 98-100% accuracy, 98-100% true positive rate, 0.0-2.3% for a minimum of a false positive rate with tests F1=0.95 and MCC=0.95 as well as AUC=1. For the blue-white veil the area under curve (AUC) for the receiver operating characteristic curve (ROC) the average values are 0.966, 0.969 and 0.957, as well as the maximum values 0.996 for VGG19, 0.985 for Xception and 0.971 for Inception-ResNet-v2.

In the case of the symmetry/asymmetry research, the IDA method turned out to be in many cases more effective than training the network only on the original images, even by 20% higher. In the dermatological asymmetry studies, we achieved for VGG19, the best CCN network, a maximum of 68.56% weighted accuracy, 92.3% true positive rate, tests F1 = 0.56 and MCC = 0.62, AUC = 0.965, and a

Table 1. The results for the blue-white veil classification

CM factor		VGG19			Xception			Inception-ResNet-v2		
		T1	**T8**	**IDA**	**T1**	**T8**	**IDA**	**T1**	**T8**	**IDA**
w.ACC [%]	**AVG**	87.2	88.1	92.2	82.2	80.2	88.0	79.8	78.2	83.9
	MAX	98.8	97.5	100	87.7	85.3	94.4	88.9	86.1	93.2
TPR [%]	**AVG**	77.8	79.6	89.4	66.7	61.8	81.7	62.2	58.4	73.9
	VAR	11.7	10.2	8.2	7.9	7.1	9.5	13.3	16.1	10.1
	Min	55.6	66.7	77.8	55.6	51.4	66.7	33.3	23.6	55.6
	Max	100	97.2	100	77.8	73.6	88.9	77.8	76.4	88.9
FPR [%]	**AVG**	3.4	3.3	5.0	2.3	1.5	5.6	2.7	2.1	6.1
	VAR	2.2	2.0	2.2	1.6	1.1	3.9	1.7	1.3	2.7
	Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	2.4
	Max	7.3	7.0	7.3	4.9	3.0	14.6	4.9	5.2	12.2
Test	**F1 Max**	0.95	0.96	1.0	0.82	0.79	0.94	0.88	0.80	0.89
	MCC	0.94	0.95	1.0	0.79	0.74	0.94	0.86	0.76	0.86

minimum of 2.7% a false positive rate for the asymmetric cases.

The time of training for the BW approach has varied from around 12 minutes (VGG19), 40 minutes (Xception), and 60 minutes (IRN2) for the PH2 dataset. The times of the training of the CNNs in the BW2 case varied from 5 minutes (VGG19 with 60 epochs) to 14 minutes (XN with 100 epochs), and 21 minutes (IRN2 with 60 epochs) for the PH2 dataset.

5. Conclusions

The correct selection of specific features, in particular, the blue-whitish veil has an impact on the analysis of specific disease fragments. In the research, we have used the three pretrained CNN networks: Xception, VGG19 [11], and Inception-ResNet-v2 as well as PH2 dermoscopic image dataset [20]. The results achieved are quite promising. The average accuracy was above 90 %. The networks usually quite well classified the images of the lesions. VGG19 appears to show the best results from three CNNs.

In the previous papers we have shown that the classifications results are higher when the networks have been trained, validated and tested on the image datasets achieved using the Invariant Dataset Augmentation (IDA). That general approach can also provide higher results while using CNN networks in other disciplines not only dermatology, chest X-ray images of the patients with different lung illnesses e.g. viral pneumonia and COVID-19 [21].

Table 2. The results for the asymmetry classification

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
w.ACC [%]	AVG	55.2	56.0	58.6	57.8	58.1	56.2	48.1	49.6	48.8
	Max	69.2	68.9	68.3	78.9	73.3	68.6	54.8	56.8	57.0
Test	F1 Max	0.543	0.548	0.555	0.560	0.560	0.511	0.446	0.465	0.439
	MCC	0.578	0.524	0.621	0.732	0.650	0.643	0.429	0.424	0.449

References

[1] ECIS, *European cancer information system*, 2023, (access: 14-02-2023). https://ecis.jrc.ec.europa.eu

[2] ACS, *American cancer society*, 2023, (access: 14-02-2023). https://www.cancer.org/research/cancer-facts-statistics.html

[3] Soyer H.P., Argenziano G., Zalaudek I., Corona R., Sera F., Talamini R., Barbato F., Baroni A., Cicale L., Di Stefani A., Farro P., Rossiello L., Ruocco E., Chimenti S., *Three-Point Checklist of Dermoscopy: A New Screening Method for Early Detection of Melanoma*, *Dermatology*, 2004, vol. 208, no 1, pp. 27–31, doi: 10.1159/000075042.

[4] Kawahara J., Daneshvar S., Argenziano G., Hamarneh G., *Seven-point checklist and skin lesion classification using multitask multimodal neural nets*, *IEEE Journal of Biomedical and Health Informatics*, 2019, vol. 23, no 2, pp. 538–546, doi: 10.1109/JBHI.2018.2824327.

[5] Nachbar F., Stolz W., Merkle T., Cognetta A.B., Vogt T., Landthaler M., Bilek P., Braun-Falco O., Plewig G., *The abcd rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions*, *Journal of the American Academy of Dermatology*, 1994, vol. 30, no 4, pp. 551–559, doi: https://doi.org/10.1016/S0190-9622(94)70061-3.

[6] Rodriguez Salas R., Dokladal P., Dokladalova E., *Rotation invariant networks for image classification for hpc and embedded systems*, *Electronics*, 2021, vol. 10, no 2, doi: 10.3390/electronics10020139.

[7] Tarasiuk P., Szczepaniak P.S., *Novel convolutional neural networks for efficient classification of rotated and scaled images*, *Neural Computing and Applications*, 2022, vol. 34, no 13, pp. 10519–10532.

[8] Cheng G., Zhou P., Han J., *Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images*, *IEEE Transactions on Geoscience and Remote Sensing*, 2016, vol. 54, no 12, pp. 7405–7415, doi: 10.1109/TGRS.2016.2601622.

[9] Dieleman S., Willett K.W., Dambre J., *Rotation-invariant convolutional neural networks for galaxy morphology prediction*, *Monthly Notices of the Royal Astronomical Society*, 2015, vol. 450, no 2, pp. 1441–1459, doi: 10.1093/mnras/stv632.

[10] Weiler M., Hamprecht F.A., Storath M., *Learning steerable filters for rotation equivariant cnns*, [In:] *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 849–858, doi: 10.1109/CVPR.2018.00095.

[11] Simonyan K., Zisserman A., *Very deep convolutional networks for large-scale image recognition*, [In:] *International Conference on Learning Representations*.
https://www.robots.ox.ac.uk/~vgg/publications/2015/Simonyan15/

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

[13] Milczarski P., Stawska Z., *Classification of skin lesions shape asymmetry using machine learning methods*, [In:] L. Barolli, F. Amato, F. Moscato, T. Enokido, M. Takizawa (eds.), *Web, Artificial Intelligence and Network Applications*, Springer International Publishing, Cham, pp. 1274–1286, doi: 10.1007/978-3-030-44038-1_116.

[14] Milczarski P., Beczkowski M., Borowski N., *Blue-white veil classification of dermoscopy images using convolutional neural networks and invariant dataset augmentation*, [In:] L. Barolli, I. Woungang, T. Enokido (eds.), *Advanced Information Networking and Applications*, Springer International Publishing, Cham, pp. 421–432, doi: 10.1007/978-3-030-75075-6_34.

[15] Beczkowski M., Borowski N., Milczarski P., *Classification of dermatological asymmetry of the skin lesions using pretrained convolutional neural networks*, [In:] *Artificial Intelligence and Soft Computing*, Springer International Publishing, 2021, pp. 3–14, doi: 10.1007/978-3-030-87897-9_1.

[16] Menzies S.W., Menzies S.W., Crotty K., Ingvar C., McCarthy W., *Dermoscopy: An Atlas*, McGraw-Hill Australia, 2009.

[17] He K., Zhang X., Ren S., Sun J., *Deep residual learning for image recognition*, [In:] *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[18] Celebi M.E., Iyatomi H., Stoecker W.V., Moss R.H., Rabinovitz H.S., Argenziano G., Soyer H.P., *Automatic detection of blue-white veil and related structures in dermoscopy images*, *Computerized Medical Imaging and Graphics*, 2008, vol. 32, no 8, pp. 670–677, doi: 10.1016/j.compmedimag.2008.08.003.

[19] Milczarski P., Beczkowski M., Borowski N., *Enhancing dermoscopic features classification in images using invariant dataset augmentation and convolutional neural networks*, [In:] *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28*, Springer, pp. 403–417, doi: 10.1007/978-3-030-92238-2_34.

[20] Mendonça T., Ferreira P.M., Marques J.S., Marcal A.R., Rozeira J., *Ph 2-a dermoscopic image database for research and benchmarking*, [In:] *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, pp. 5437–5440, doi: 10.1109/EMBC.2013.6610779.

[21] Milczarski P., Beczkowski M., Borowski N., *Covid-19 lungs assessment in chest x-ray images using convolutional neural networks*, [In:] *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 2, pp. 1062–1067, doi: 10.1109/IDAACS53288.2021.9661046.