How to Enhance Classification Results of the Trained Models using Invariant Dataset Augmentation

Piotr Milczarski[1][0000−0002−0095−6796] and
Norbert Borowski[2][0000−0002−3861−2411]

[1] University of Lodz, Faculty of Mathematics and Computer Science, Łódź, Poland
piotr.milczarski@uni.lodz.pl
[2] Lodz University of Technology, Institute of Information Technology, Łódź, Poland
norbert.borowski@p.lodz.pl

Abstract. In the paper, we show how to enhance the results of the classification of the images by already trained models without retraining. We use the rotation invariance of the images and Invariant Dataset Augmentation (IDA) method instead. The results of enhancement using the IDA method are shown using the dermoscopic PH2 dataset with 200 images, the dermoscopic Derm7pt dataset with 1011 images, and the X-Ray COVID-19 dataset with 3175 images. The authors method allows us to do the test not only on the single image but also on its invariant copies. Depending on the image symmetry, there can even be an eightfold increase in the number of test images. In the presented research, we used CNN networks VGG19, XN, and Inception-ResNet-v2 for dermoscopic datasets and the three classification features of skin lesions defined by well-known dermoscopic criteria. VGG19, Xception, and Inceptionv3 networks are used for COVID-19 dataset with three classes. The confusion matrix parameters shown in the paper improved significantly comparing the tests on the original dataset versus IDA increased one and the worst case scenario. It is achieved on the original model without retraining it on the augmented dataset using IDA method. An example for dermoscopic PH2 and Derm7pt datasets resulted in precision between 98-100%, true positive rate 98-100%, false positive rate $<$2.3%, F1 test = 0.95 and MCC test = 0.95. The results for the COVID-19 X-ray resulted in maximum precision for 3 classes $>$98.1%, F1 score almost 1.0 and MCC $>$0.97. The advantage of the method is that the trained model is smaller and more efficient than the new model trained on the eightfold increased dataset.

Keywords: Invariant Dataset Augmentation · Increasing classification rates · Deep learning · COVID-19 · Dermoscopic images.

1 Introduction

The motivation of the paper is to enhance the classification rates provided by already trained and used models in the e.g. health-care systems. In the screening

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_24

methods e.g. dermoscopic and X-Ray ones, we increase the efficiency by improving the confusion matrix parameters of the feature classifications that point to possible health problems. The advantage of the method is that the trained network is smaller and more efficient than the new network trained on the eightfold increased data set.

That general approach can provide better results while using CNN networks in other research disciplines, not only dermatology or pulmonology. This method is integrated during the classification stage, where the model processes both the original image and its pixel-invariant copies. Although standard CNNs do not possess inherent rotation invariance, evaluating them against a collection of pixel-invariant images improves classification performance, leading to higher metrics such as recall (TPR) and test F1 rates.

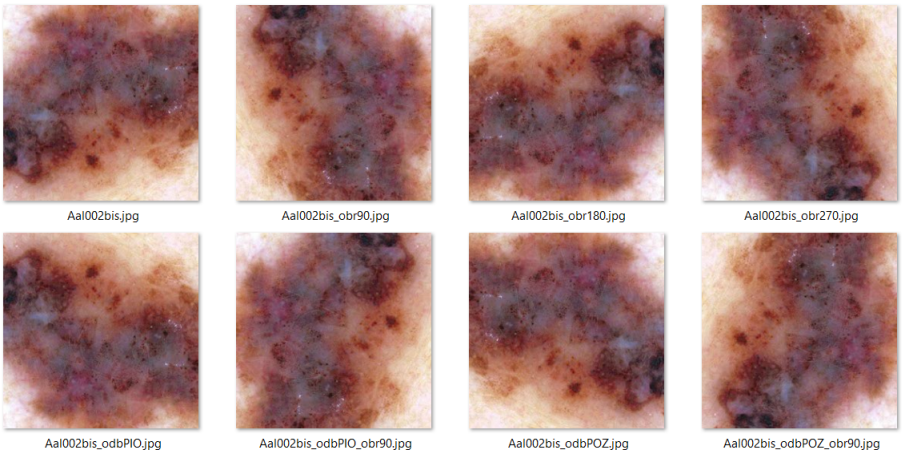

Fig. 1. The image (top-left) Aaal002.jpg from Derm7pt [1] and its invariant copies.

The second issue that the method solves is the acquisition and analysis of the data, as well as the extraction of the features. The classification of features should not depend on the orientation and position of the object in the flat 2D image (see Fig. 1). The optical sensor must exhibit rotational invariance to ensure that feature extraction remains independent of the object's spatial orientation. Consequently, we propose a methodology for CNN networks evaluation that incorporates both the primary image and its corresponding invariant representations.

As a case-study we use different medical dataset. One of them asre dermoscopic datasets of skin lesion and their assessment methodologies the three-point checklist of dermoscopy (3PCLD) [18, 20] and the seven-point checklist (7PCL) [10]. The next case-study dataset contains X-ray images of the lung with their assessment methods. The problem of COVID-19 and X-Ray classification is also widely presented and several different approaches are used as in the dermoscopy case [14, 17]. The usual classification is based on X-Ray lungs images with normal lungs, with viral pneumonia, and COVID-19 ones.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_24

Convolutional Neural Networks (CNNs) inherently lack rotational invariance, a limitation extensively documented in the literature [2, 3, 9, 19]. To address this deficiency, various architectural enhancements and methodologies have been proposed to integrate invariant properties into deep learning frameworks [3, 9]. This paper delineates a strategy that leverages geometric symmetry to optimize classification performance; comprehensive technical details and preliminary findings are further elaborated in the authors' previous works [2, 12, 13, 15].

In the paper, we show that using Invariant Dataset Augmentation (IDA) method in a test phase provides similar results of classification the images as pretraining the models using IDA augmented models.

2 Invariant Dataset Augmentation in Trained Models

In the papers [2, 12–15], the IDA method is used to increase the number of images using invariant transformations to tackle the lack of the rotation invariant property in CNN networks. It can be explained by the properties of the convolution given by the example equation:

$I(x,y) = \sum_{i=0}^{n} \sum_{j=0}^{m} k(i,j) I(x+i,\ y+j)$,

where the kernel of the operation is $k(I,j)$ is of size NxM. The image size is NxM, where $N \geq n$ and $M \geq m$. The classification probability of invariant images can be different for each image copy. The problem of rotation invariance is presented in, e.g. Tab. 1, where the classification probability of the image IMD168 from PH2 dataset [11] and its copies differs greatly for each invariant copy. The asymmetry classification (0 - no symmetry, 1 axis symmetry, 2-axis symmetry) sometimes results in opposite result.

Table 1. The classification probability of symmetry by VGG19 on image IMD168 from PH2 dataset [11]

Image IMD168 and its copies	**Symmetry probability**		
	0	**1**	**2**
Original	0.0131	0.9593	0.0276
90 rotation	0.5501	0.4017	0.0482
180 rotation	0.9420	0.0407	0.0173
270 rotation	0.3044	0.6713	0.0243
Vertical reflection	0.3781	0.6170	0.0050
Vert. reflection and 90 rotation	0.4937	0.4964	0.0099
Horizontal reflection	0.0543	0.9403	0.0054
Hor. reflection and 90 rotation	0.6154	0.3441	0.0405

The explanation of the IDA method and funding is provided in our other articles [2, 13, 15]. It is based on natural pixel invariant transformations of the images. In some cases, it is only mirroring reflection in other rotation by 90 degrees and mirroring.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_24

Let us briefly explain how to apply the Invariant Dataset Augmentation method for testing only after preprocessing the test image, i.e. cropping and scaling, to the CNN networks requirements in the following steps:

1. Find the image symmetry invariance features e.g. rotation, mirroring.
2. Define the number of folds of the image.
3. Create the invariant copy/-ies from the original image.
4. Classify images.
5. Use IDA's worst-case scenario to assign the class to the image.

Table 2. Classification probability of X-ray images using IDA approach for the chosen three images and their invariant copy. Where C19= COVID19(194), C19_m = COVID19(194)_mir, N = NORMAL (964), N_m = NORMAL (964)_mir, VP=Viral Pneumonia (314), VP_m = Viral Pneumonia (314)_mir.

CNN	**Classification probability by CNN / True Class**								
	VGG19			**Xception**			**Inception_v3**		
Img ID	**C**	**N**	**V**	**C**	**N**	**V**	**C**	**N**	**V**
C19	0.813	0.184	0.003	0.011	0.965	0.025	0.238	0.611	0.151
C19_m	0.007	0.993	3e-4	0.013	0.267	0.720	0.207	0.716	0.077
N	0.005	0.995	2e-4	0.285	0.641	0.074	0.856	0.135	0.009
N_m	0.001	0.999	1e-4	0.740	0.156	0.104	0.984	0.014	0.002
VP	0.000	0.000	1.000	0.0008	0.000	0.999	0.001	3e-4	0.999
VP_m	2e-4	7e-4	0.999	0.001	1e-4	0.999	0.003	2e-4	0.997

In Tables 1 and 2 detailed classification probabilities are presented to highlight the ambiguous result of the classification. Tab. 1 shows the classification probabilities for 3 chosen models with 8 IDA augmented images as a test. Tab. 2 shows the classification probabilities for 3 images and its invariant copy for each class of the X-Ray set. Tab. 2 shows that the classification probability for 3 classes may differ as in the dermoscopic feature case.

Rotations and mirror reflection are used for the dermoscopic images and blue-white veil feature, asymmetry, pigment network distribution. Seven new copies are achieved for each image. For X-Ray images only mirroring was applied. These transformations used by IDA do not change pixels; they are pixel invariant, mutually unambiguous, and reversible.

For the PH2 dermoscopic dataset [11], in total, 1600 images out of 200 were achieved. For Derm7pt [1, 10] 8088 out of 1011 images were prepared.

Similar situation applies to X-ray lungs dataset that was built using 3175 cases from the patients with viral pneumonia (1296 images) and COVID-19 (540 images) as well as healthy ones (1339 images). A subset of the RYDLS-20

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_24

dataset [16], consisting of 1144 chest X-ray (CXR) cases, was incorporated into the analysis.

The COVID-19 dataset of 540 images was collected and chosen [14]:

- 32 from Actualmed COVID-19 chest X-ray data initiative [5].
- 162 from the North America, R. S. COVID-19 radiography database [4, 17].
- 321 from Cohen, Morrison, and Dao. COVID-19 image data collection [7, 8].
- 25 from COVID-19 chest X-ray data initiative [6].

3 Results

Tables 3 and 4 present a detailed comparison of the classification results obtained on the original model and the model built using the IDA augmented train set. We show for both approaches the results of testing on the original images (T1), IDA augmented images (T8), and a worst-case scenario (IDA) i.e. if one positive the case is positive. We have used several models and studied their behavior, e.g. InceptionResNetv2 (IRN2), Inception-v3, Xception, VGG19.

Table 3. The classification of the blue-white veil classification using original images from PH2 as a train set. T1 - test on original images, T8 - test on 8 copies independently, IDA - worst-case scenario.

CM factor		VGG19			Xception			IRN2		
		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 Max	0.94	0.95	1.0	0.79	0.74	0.94	0.86	0.76	0.86

The research has been conducted for original and IDA augmented models with the following conditions:

- the dataset has been divided into 4 equal parts;
- the train and validation dataset consisted of 3 parts (75%) images and the test of the remaining 1 part (25%);
- the experiments are repeated 5 times for each model and dataset, resulting in 20 models for each CNN network.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_24

In the result, different models used the same images as train and validation sets although their resolutions were different. Detailed results are shown for the blue-white feature classification because that feature can be easily lost by rotating the image at a random angle.

Table 4. The classification of the blue-white veil classification using images from PH2 set and its 7 copies as a train set. T1 - test on original images, T8 - test on 8 copies independently, IDA - worst-case scenario.

CM factor		VGG19			Xception			IRN2		
		T1	**T8**	**IDA**	**T1**	**T8**	**IDA**	**T1**	**T8**	**IDA**
w.ACC [%]	**AVG**	87.4	87.7	89.5	84.9	84.1	89.5	82.9	82.9	85.5
	MAX	94.4	94.3	98.8	92.0	91.3	93.2	88.9	88.0	93.2
TPR [%]	**AVG**	76.7	77.6	82.7	72.2	69.9	84.4	68.3	67.8	76.1
	VAR	12.1	9.6	10.23	7.5	7.9	6.5	8.1	9.3	8.8
	Min	55.6	62.5	66.7	55.6	58.3	66.7	44.4	47.2	66.7
	Max	88.9	88.9	100	88.9	83.3	88.9	77.8	79.2	88.9
FPR [%]	**AVG**	2.0	2.3	3.8	2.4	1.7	5.4	2.6	1.9	5.1
	VAR	2.3	1.8	1.6	1.7	1.0	1.7	1.4	1.4	3.2
	Min	0.0	0.3	2.4	0.0	0.3	2.4	0.0	0.3	2.4
	Max	7.3	5.2	7.3	4.9	3.4	7.3	7.3	5.8	12.2
Test	**F1 Max**	0.94	0.93	0.95	0.84	0.90	0.89	0.88	0.81	0.89
	MCC Max	0.93	0.91	0.94	0.81	0.88	0.86	0.86	0.78	0.86

Tab. 3 shows results for the blue-white veil for models that were trained on a small PH2 dataset (case BW1), where T1 stands for test on original images, T8 - test on 8 copies independently, IDA - worst-case scenario (case BW8). Tab. 4 also shows the classification results of the blue-white veil for the same models as in Tab. 3 but trained on the augmented datasets. It can be derived from the research on 20 models that:

- accuracy and recall values are higher in IDA, BW1 and BW2 cases;
- VGG19 shows that in BW1 case the recall is higher by 6.7%, but FPR is also higher by 1.2%; F1 test and MCC are higher in BW1 case;
- Xception shows in BW8 case the recall higher by 2.7% with almost the same FPR as in BW1 case; F1 test and MCC are higher in BW1 case; F1 test and MCC are higher in BW1 case;
- IRN2 shows that in BW8 case the recall is higher by 2.2%, but FPR is lower by 1%; F1 test and MCC are the same in BW1 and BW8 cases;

4 Conclusions

The results of the classification should not depend on the image acquisition conditions, such as angle. Retraining the models on a new augmented dataset is not often needed. The better results than on the original model and one picture test approach is also achieved while using the same model but IDA augmented

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_24

images for the test taking into account their rotation importance. The results for the dermoscopic datasets show the precision between 98- 100%, the true positive rate 98- 100%, the false positive rate 0.0-2.3%, the F1 test 0.95 and the MCC test 0.95. The results of the COVID-19 and X-Ray image classification resulted in precision>97.5% for the COVID-19, balanced ACC>98.5%, the recall (TPR) around 97%; F1 test = 1.0 and MCC test>0.97 [12].

The advantage of the method is that the trained network is smaller and more efficient than the new network trained on the eightfold increased data set. We have also trained the same models using the images and their copies [2, 14, 15]. The results were slightly better than those presented in this paper, but the models were 8 times larger.

References

1. Argenziano, G., Soyer, P., De Giorgi, V., Piccolo, D., Carli, P., Delfino, M., Ferrari, A., Hofmann-Wellenhof, R., Massi, D., Mazzocchetti, G., Scalvenzi, M., Wolf, I.: Interactive atlas of dermoscopy. Dermoscopy: a tutorial (Book) and CD-ROM. Edra - Medical Publishing & New Media Milan, Italy (02 2000)
2. 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, pp. 3–14. Springer International Publishing (2021). https://doi.org/10.1007/978-3-030-87897-9_1
3. 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 **54**(12), 7405–7415 (Dec 2016). https://doi.org/10.1109/TGRS.2016.2601622
4. Chowdhury, M.E.H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Emadi, N.A., Reaz, M.B.I., Islam, M.T.: Can ai help in screening viral and covid-19 pneumonia? IEEE Access **8**, 132665–132676 (2020). https://doi.org/10.1109/ACCESS.2020.3010287
5. Chung, A.: Actualmed covid-19 chest x-ray dataset initiative (2020), https://github.com/agchung/Actualmed-COVID-chestxray-dataset, last accessed on July 28, 2025
6. Chung, A.: Covid-19 chest x-ray dataset initiative (2020), https://github.com/agchung/Figure1-COVID-chestxray-dataset, last accessed on March 27, 2026
7. Cohen, J.P., Morrison, P., Dao, L.: Covid-19 image data collection. arXiv 2003.11597 (2020), https://github.com/ieee8023/covid-chestxray-dataset, last accessed on March 27, 2026
8. Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., Ghassemi, M.: Covid-19 image data collection: Prospective predictions are the future. arXiv 2006.11988 (2020), https://github.com/ieee8023/covid-chestxray-dataset, last accessed on March 27, 2026
9. Dieleman, S., Willett, K.W., Dambre, J.: Rotation-invariant convolutional neural networks for galaxy morphology prediction. Monthly Notices of the Royal Astronomical Society **450**(2), 1441–1459 (04 2015). https://doi.org/10.1093/mnras/stv632

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_24

10. 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 **23**(2), 538–546 (2019). https://doi.org/10.1109/JBHI.2018.2824327
11. 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). pp. 5437–5440. IEEE (2013). https://doi.org/10.1109/EMBC.2013.6610779
12. Milczarski, P.: Increasing the classification rates of the trained models using invariant dataset augmentations. In: 2025 IEEE/CVF International Conference on Computer Vision (ICCV) Workshops. pp. 1083–1092 (2025). https://doi.org/10.1109/ICCVW69036.2025.00117
13. Milczarski, P., Beczkowski, M., Borowski, N.: Blue-white veil classification of dermoscopy images using convolutional neural networks and invariant dataset augmentation. In: Barolli, L., Woungang, I., Enokido, T. (eds.) Advanced Information Networking and Applications. pp. 421–432. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-75075-6_34
14. 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 (2021). https://doi.org/10.1109/IDAACS53288.2021.9661046
15. 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. pp. 403–417. Springer (2021). https://doi.org/10.1007/978-3-030-92238-2_34
16. Pereira, R.M., Bertolini, D., Teixeira, L.O., Silla, C.N., Costa, Y.M.: Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios. Computer Methods and Programs in Biomedicine **194**, 105532 (2020). https://doi.org/10.1016/j.cmpb.2020.105532
17. Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S.B., Islam, M.T., Al Maadeed, S., Zughaier, S.M., Khan, M.S., Chowdhury, M.E.: Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. Computers in Biology and Medicine **132**, 104319 (2021). https://doi.org/10.1016/j.compbiomed.2021.104319
18. 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 **208**(1), 27–31 (02 2004). https://doi.org/10.1159/000075042
19. Tarasiuk, P., Szczepaniak, P.S.: Novel convolutional neural networks for efficient classification of rotated and scaled images. Neural Computing and Applications **34**(13), 10519–10532 (2022)
20. Was, L., Wiak, S., Milczarski, P., Szymanski, L.: Evaluation of dermatological asymmetry measure of shape by expectation-maximization. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds.) Artificial Intelligence and Soft Computing. pp. 201–211. Springer Nature Switzerland, Cham (2025)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_24