



Classification of Dermatological Asymmetry of the Skin Lesions Using Pretrained Convolutional Neural Networks

Michał Beczkowski¹ , Norbert Borowski² , and Piotr Milczarski¹  

¹ Faculty of Physics and Applied Informatics, Department of Computer Science, University of Lodz, Pomorska Street 149/153, 90-236 Lodz, Poland

{michal.beczowski, piotr.milczarski}@uni.lodz.pl

² Faculty of Physics and Applied Informatics, Department of Nuclear Physics, and Radiation Safety, University of Lodz, Pomorska Street 149/153, 90-236 Lodz, Poland

norbert.borowski@uni.lodz.pl

Abstract. In dermatology, malignant melanoma is one the most deadly forms of skin cancer. It is extremely important to detect it at an early stage. One of the methods of detecting it is an evaluation based on dermoscopy combined with one of the criteria for assessing a skin lesion. Such an evaluation method is the Three-Point Checklist of Dermoscopy which is considered a sufficient screening method for the assessment of skin lesions. The proposed method, founded on the convolutional neural networks, is aimed at improving diagnostics and enabling the preliminary assessment of skin lesions by a family doctor. The current paper presents the results of the application of convolutional neural networks: VGG19, Xception, Inception-ResNet-v2, for the assessment of skin lesions asymmetry, along with various variations of the PH2 database. For the best CNN network, we achieved the following results: true positive rate for the asymmetry 92.31%, weighted accuracy 67.41%, F1 score 0.646 and Matthews correlation coefficient 0.533.

Keywords: CNN · Dermoscopy · Asymmetry · Pretrained convolutional neural networks · Invariant dataset augmentation

1 Introduction

1.1 Dermatological Asymmetry of Skin Lesions and Screening Methods

According to the statistics compiled by European Cancer Information System (ECIS) [1] and the American Cancer Society (ACS) [2] life-threatening melanoma can be completely cured if removed in the early stages [3, 4]. According to ECIS, the estimated risk for 2018 of melanoma varies from 38.2 new cases in Germany to 13.6 cases in Iceland per 100K age and gender standardized population [1]. Also, ACS reports in 2018 that the risk of Americans developing cancer over their lifetime is 37.6% for females and 39.7% for males, where the melanoma risk is 1 in 42 cases for females and 1 in 27

cases for males [2]. It is necessary to develop a quick and effective diagnostic method: to minimize the excision of benign lesions and increase the detection of melanoma. Dermatology experts use various screening methods such as the Three-Point Checklist of Dermoscopy (3PCLD) [5–7], The Seven-Point Checklist (7PCL) and the ABCD rule [9, 10]. All of them are considered effective in skin lesion assessments.

3PCLD methodology is based on the criteria of asymmetry in shape, hue, and structure distribution within the lesion defined and it can have a value of either 0 for symmetry in two axes, 1 for symmetry in one axis, or 2 for asymmetry. In this method, the pigmented network and blue-white veil are either present or absent. Another example of the screening method used in dermatology is the ABCD's of melanoma. In this rule ABCD stands for asymmetry, border (not well-defined, irregular), color (more than one shade), diameter (usually larger than 6 mm) and evolution (changing features over time). All those features are characteristics of melanoma that general physicians or dermatologists check while diagnosing. Like in 3PCLD methodology, this method focuses on the asymmetry of the lesion [11–13] which is one of the common characteristics of skin damage that can be noticed visually. These examples show the importance of symmetry/asymmetry in various screening methods of detecting melanoma. In the paper, we show the results of the CNN application to the problem of the asymmetry within the skin lesion in the dermoscopic images.

There are a few publications about the symmetry/asymmetry of the skin lesion using machine learning/AI methods. In the paper [16], there is only shape asymmetry discussed and the authors tested several ML methods on the PH2 dataset. The result showed 95.8% of accuracy, with true positive rates for the asymmetry 92.5%, 95.7% for the 1-axis symmetry and 100% for the symmetric lesions while using the SVM with the radial basis kernel function.

This research paper presents the results of the application of convolutional neural networks for the diagnosis of skin lesions asymmetry. The neural networks were based on available, pre-trained networks such as Xception (XN), VGG19 [14], and Inception-ResNet-v2 (IRN2). Those networks provide promising results even with a relatively small but well-described PH2 dataset [15].

1.2 Dermatological Datasets

From the available databases, we have chosen the PH2 dataset [15] to conduct our research. This database consists of dermoscopic images obtained at the Dermatology Service of Hospital Pedro Hispano (Matosinhos, Portugal) under the same conditions through the Tuebinger Mole Analyzer system using a magnification of 20 times. Images in the dataset are 8-bit RGB color images with a resolution of 768×560 pixels.

This image database contains a total of 200 dermoscopic images of melanocytic lesions, including 80 common nevi, 80 atypical nevi, and 40 melanomas. The PH2 database 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) [8, 15, 16].

One of the alternatives for the PH2 database is the ISIC Archive which contains the largest publicly available collection of quality controlled dermoscopic images of skin

lesions [17]. The ISIC Archive contains over 24,000 dermoscopic images, which were collected from leading clinical centers internationally and acquired from a variety of devices within each center. The ISIC dataset metadata does not provide information about the asymmetry of lesions. The other examples of the dermatological datasets can be found in Interactive Atlas of Dermoscopy or An Atlas of Surface Microscopy of Pigmented Skin Lesions [18, 19].

1.3 Pretrained Convolutional Neural Network and Their Features

The pretrained Convolutional Neural Networks have different features that should be taken into account when choosing a network to apply to a given problem. The most important characteristics are network accuracy, true positive and negative rate, speed of classification, and size. While selecting a network these features should be taken into account. Currently, we can choose within several pretrained networks. The chosen three networks' characteristics are given in Table 1. The network depth is defined as the largest number of sequential convolutional or fully connected layers on a path from the input layer to the output layer. The inputs to all networks are RGB images.

Table 1. Pretrained convolutional neural networks parameters

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

2 Data Preparation for the Research

2.1 Augmentation and Preparation of the Database

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, as well as other features present in both 3PCLD and 7 PCL, were used. For the transformation of images, we chose three rotations by 90°, 180° and 270°, mirroring on the vertical and horizontal axis, and a 90° rotation of the images after mirroring (Fig. 1). In total, we got seven transformations for each image that did not change the pixels, shape, or color distribution. These transformations allowed us to increase the PH2 database from 200 to 1600 images.

To show the idea of the author method, to classify not only using the original image but as well its invariant copies we provide in Table 2 the classification probabilities of

the exemplary image classification and its invariant copies (the image IMD168 from the PH2 dataset) by the chosen VGG19 CNN network trained on the images from PH2 and their seven copies. The probability of the classification for the asymmetry (column with value '0') networks varies from 0.013 to 0.94. The same variance of probability occurs for other CNN networks. It can be concluded that the same image and its invariant versions can provide us with opposite classification results due to convolutional network operations.

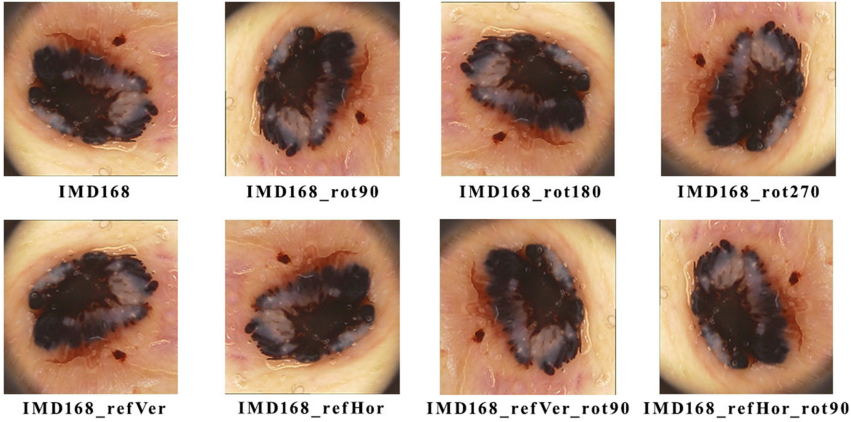


Fig. 1. A sample of image IMD168 from the PH2 [15] dataset and its invariant augmentation.

Table 2. VGG19 CNN classification probability of the image IMD168 from the PH2 dataset.

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

For example, during a convolution each of the eight images with the filters and their weights give a different output result due to the convolution properties that can be derived from the formula:

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

where the kernel k is of size n by m . The image is size $N \times M$, where $N \geq n$ and $M \geq m$. As it is shown in the final section the probability of the classification of each of the invariant images can vary from 0.0 to 1.0.

The next step in preparing the database was to scale the images to the input sizes required by the selected networks Table 1. First, we scaled the shorter dimension of images (in our case, height) to the input size, e.g. 224 px (see Table 1) using the Bicubic Sharper algorithm in Photoshop. Then, all images were cropped to a square shape. As a result, we obtained a set of images scaled to the sizes required by each of the networks, e.g. 224×224 px, see Table 1. The dataset prepared in this way contains 936 fully symmetric, 248 symmetric in one axis and 416 fully asymmetric images of skin lesions and met the requirements of our research and could be used for network tests.

2.2 CNN Network Setting and Configuration

We used pre-trained networks in our research because they are trained on the ImageNet database [20]. Moreover, those networks use as starting point to learn a new task previous abilities to extract informative features from natural images. Since in each pre-trained 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.

To achieve the highest classification rates we have conducted the initial research testing the wide variety of the following parameters for all three networks:

- 30–60 epochs;
- learning rate from $1e - 4$ to $1e - 2$.

After the initial research we choose for:

- VGG19 – learning rate $1e - 4$ and 40 epoch;
- XN - learning rate $5e - 4$ and 30 epoch;
- IRN2 - learning rate $5e - 4$ and 30 epoch.

The time of training depends on the network and number of the training images and the machine specification. The times for the machine 1 specification have varied from around:

- 18 min for VGG19;
- 30 min for XN;
- 60 min for IRN2.

2.3 Hardware Description

To ensure the credibility of the results, the research was conducted independently on two computers with the same operating system (Microsoft Windows 10 Pro) and different configurations:

- Set 1. Processor: Intel(R) Core(TM) i7-8700K CPU @ 3.70 GHz (12 CPUs), Memory: 64 GB RAM, Graphics Card: NVIDIA GTX 1080Ti with 11 GB of Graphics RAM.
- Set 2. Processor: Intel(R) Core(TM) i7-9700K CPU @ 3.60 GHz (8 CPUs), Memory: 16 GB RAM, Graphics Card: NVIDIA GeForce RTX 2070 with 8 GB of Graphics RAM.

On both machines, the research was conducted using Matlab 2019b with up-to-date versions of Deep Learning Toolbox™ (v. 12). Deep Learning Toolbox allows to transfer learning with pretrained deep network models, see Table 1. The second hardware set was used to test the procedure to see whether the classification parameters depend on their hardware. Different configurations affected only the time of execution in CNN networks training and at the end training. When working on both machines the calculated average accuracy, as well as their maximum and minimum, showed results close to each other. It proved that the procedure was not hardware-dependent.

3 Research Method Description

The first step in our research method is database preparation. 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. The division was carried out 4 times so that each image was included in the test set. The training, validation and testing image cases for the three chosen networks were the same, although the image sizes were different. This allows us to assess and compare the results more thoroughly. Also, 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, which was also divided in the previously mentioned way into training and testing set, see Table 3. All steps were repeated on different image sizes to make it possible to research different networks, see Table 1.

Table 3. The number of the images in the original and augmented PH2 set.

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

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 mentioned in Table 1 parameters such as accuracy, true positive rate were defined and calculated according to Eqs. (1)–(6). Next, their average values with the variance, minimum and maximum values were calculated for twenty-five (5 rounds \times 5). Correct classification plus overestimation which is the Accuracy + Error Type I were considered best for purpose of our research: it is better if the screening method overestimates the diagnosis than the opposite (Underestimation Error Type II - False Negative) as the final diagnosis of malignant melanoma takes place after histopathological research.

The confusion matrix parameters are defined as follows:

$$ACC = (TP + TN)/N \quad (1)$$

$$TPR = TP/(TP + FN) \quad (2)$$

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

$$FPR = FN/(FP + TN) \quad (4)$$

$$F1 = 2TP/(2TP + FP + FN) \quad (5)$$

$$MCC = (TP * TN - FP * FN) / \sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)} \quad (6)$$

where:

- N - a number of all cases;
- true positive, TP – number of positive results i.e. correctly classified cases;
- true negative, TN – number of negative results i.e. correctly classified cases;
- false positive, FP – number of negative results i.e. wrongly classified cases as positive ones;
- false negative, FN – number of positive results i.e. wrongly classified cases as negative ones, also called Type II error;
- accuracy, ACC ; weighted accuracy, $w.ACC$;
- true positive rate, TPR , also called Recall; TPR_i – stands for true positive rate for the symmetry values $i = 0, 1$ and 2 ;
- false positive rate, FPR ; FPR_i – stands for false positive rate for the symmetry values $i = 0, 1$ and 2 ;
- score test $F1$;
- Matthews correlation coefficient, MCC .

In Table 4 weighted F1 and MCC are calculated as for weighted accuracy, Eq. (3).

4 Results

The research method described above allowed us to obtain 60 neural networks. Results from those networks were recorded and analyzed. 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.

The advantage of the IDA procedure is the increased value of the true positive rate (TPR) for the positive cases i.e. asymmetric ones. Asymmetric lesions according to 3PCLD and ABCD rule are more prone to be melanocytic. On the other hand, this procedure increases the false-positive value (FPR) (see Table 4) which can be considered as its biggest disadvantage. However, this procedure finds more melanoma cases than the T1 or T8 methods. The procedure of IDA is also used in blue-white veil classification by CNN in [21].

When comparing the results of the networks, we also took into account such classification characteristics as weighted accuracy (ACC), F1 score and Matthews correlation coefficient (MCC), see Table 5. Within the network, the results were similar regardless of the method used (T1, T8, IDA). However, the results of each network differed. The best results for these classification characteristics were shown by the Xception (XN) network with an accuracy score of 78.9%.

Table 4. The classifications results for the asymmetry. The chosen confusion matrix factors true positive rate for full asymmetry (TPR_0), true positive rate for symmetry in one axis (TPR_1), true positive rate for full symmetry (TPR_2), false positive rate for full asymmetry (FPR_0), false positive rate for symmetry in one axis (FPR_1), false positive rate for full asymmetry full symmetry (FPR_2) 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

(continued)

Table 4. (continued)

CM factor		VGG19			Xception			Inception-ResNet-v2		
		T1	T8	IDA	T1	T8	IDA	T1	T8	IDA
	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

Additionally, to compare the networks, the area under curve (AUC) value was used. VGG19 turned out to be the best network and obtained a result of 0.9652. Figure 2 shows the best receiver operating characteristic curve (ROC) with the highest value of the area under curve (AUC).

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.

Table 5. 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 coefficient (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
w.MCC	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

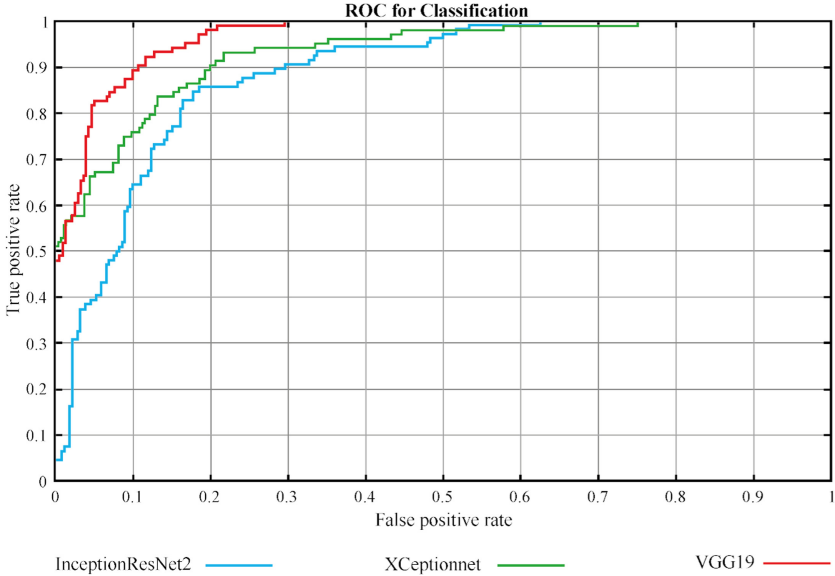


Fig. 2. The examples of the best receiver operating characteristic curve (ROC) with the highest value of the area under curve (AUC) for the three chosen CNNs.

5 Conclusions

Asymmetry plays an important role in the assessment of skin lesions, which is evident in dermatological diagnostic methods such as The Three-Point Checklist of Dermoscopy (3PCLD). Melanoma diagnosis based on asymmetry can also be made with the use of properly trained CNN networks. Such networks can serve as a helpful tool in the preliminary diagnosis of dangerous skin lesions.

In our research, we used three pretrained networks (Xception, VGG19, Inception-ResNet-v2) and trained them on our enlarged PH2 database. The method developed by us (using different forms of augmentation) turned out to be in many cases more effective than training the network only on the original images, even by 20% higher. In the studies we achieved a maximum of 68.56% weighted accuracy, 92.31% true positive rate, 66% false positive rate with tests $F1 = 0.74$, $MCC = 0.58$ and $AUC = 0.97$.

In our corresponding research [21, 22] in the field of a dermatological image processing using the Invariant Dataset of Augmentation is used with the PH2 [15] and the Atlas of Dermoscopy (Derm7pt) [8, 18] datasets to increase the classification rates e.g. true positive rate, test F1 and MCC using CNNs in comparison to the feature based methods used in [12, 16, 23].

References

1. European Cancer Information System (ECIS). <https://ecis.jrc.ec.europa.eu>. Accessed 05 Jan 2021
2. ACS – American Cancer Society. <https://www.cancer.org/research/cancer-facts-statistics.html>. Accessed 05 Jan 2021
3. Was, L., Milczarski, P., Stawska, Z., Wiak, S., Maslanka, P., Kot, M.: Verification of results in the acquiring knowledge process based on IBL methodology. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds.) ICAISC 2018. LNCS (LNAI), vol. 10841, pp. 750–760. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91253-0_69
4. Celebi, M.E., Kingravi, H.A., Uddin, B.: A methodological approach to the classification of dermoscopy images. *Comput Med. Imaging Graph.* **31**(6), 362–373 (2007)
5. Soyer, H.P., Argenziano, G., Zalaudek, I., et al.: Three-point checklist of dermoscopy. A new screening method for early detection of melanoma. *Dermatology* **208**(1), 27–31 (2004)
6. Argenziano, G., Soyer, H.P., et al.: Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. *J. Am. Acad. Dermatol.* **48**(9), 679–693 (2003)
7. Milczarski, P.: Symmetry of hue distribution in the images. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds.) ICAISC 2018. LNCS (LNAI), vol. 10842, pp. 48–61. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91262-2_5
8. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE J. Biomed. Health Inform.* **23**(2), 538–546 (2019)
9. Argenziano, G., Fabbrocini, G., et al.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermoscopy and a new 7-point checklist based on pattern analysis. *Arch. Dermatol.* **134**, 1563–1570 (1998)

10. Carrera, C., Marchetti, M.A., Dusza, S.W., Argenziano, G., et al.: Validity and reliability of dermoscopic criteria used to differentiate nevi from melanoma: a web-based international dermoscopy society study. *JAMA Dermatol.* **152**(7), 798–806 (2016)
11. Nachbar, F., Stolz, W., Merkle, T., et al.: The ABCD rule of dermatoscopy. High prospective value in the diagnosis of doubtful melanocytic skin lesions. *J. Am. Acad. Dermatol.* **30**(4), 551–559 (1994)
12. Milczarski, P., Stawska, Z., Maslanka, P.: Skin lesions dermatological shape asymmetry measures. In: *Proceedings of the IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS*, pp. 1056–1062 (2017)
13. Menzies, S.W., Zalaudek, I.: Why perform Dermoscopy? The evidence for its role in the routine management of pigmented skin lesions. *Arch Dermatol.* **142**, 1211–1222 (2006)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *Conference Track Proceedings of 3rd International Conference on Learning Representations (ICRL)*, San Diego, USA (2015)
15. Mendonca, T., Ferreira, P.M., Marques, J.S., Marcal, A.R.S., Rozeira, J.: PH2 – a dermoscopic image database for research and benchmarking. In: *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, pp. 5437–5440 (2013)
16. Milczarski, P., Stawska, Z.: Classification of skin lesions shape asymmetry using machine learning methods. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds.) *WAINA 2020. AISC*, vol. 1150, pp. 1274–1286. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44038-1_116
17. The International Skin Imaging Collaboration: Melanoma Project. <http://isdis.net/isic-project/>. Accessed 21 Mar 2020
18. Argenziano, G., Soyer, H.P., De Giorgi, V., et al.: *Interactive Atlas of Dermoscopy*. EDRA Medical Publishing & New Media, Milan (2002)
19. Menzies, S.W., Crotty, K.A., Ingvar, C., McCarthy, W.H.: *An atlas of surface microscopy of pigmented skin lesions*. Dermoscopy. McGraw-Hill, Australia (2003)
20. ImageNet. <http://www.image-net.org>. Accessed 07 Jan 2021
21. 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.) *AINA 2021. LNNS*, vol. 226, pp. 421–432. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75075-6_34
22. Milczarski, P., Was, Ł.: Blue-White Veil classification in dermoscopy images of the skin lesions using convolutional neural networks. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds.) *ICAISC 2020. LNCS (LNAI)*, vol. 12415, pp. 636–645. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-61401-0_59
23. Milczarski, P., Stawska, Z., Was, L., Wiak, S., Kot, M.: New dermatological asymmetry measure of skin lesions. *Int. Journal of Neural Networks and Advanced Applications*, Prague, pp. 32–38 (2017)