BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES
DOI: 10.24425/bpasts.2024.151675

Melanoma recognition using an ensemble of deep CNN structures

F. GIL[2], S. OSOWSKI[1,2]
[1]Warsaw University of Technology, Warsaw, Poland
[2]Military University of Technology, Warsaw, Poland

Abstract.
The paper proposes a deep-learning approach to the recognition of melanoma images. It relies on the application of many different architectures of CNN combined in the form of an ensemble. The units of the highest efficiency are selected as the potential members of the ensemble. Different methods of arrangement of the ensemble members are studied and the limited number of the best units are included in the final form of an ensemble. The results of numerical experiments performed on the ISIC2017 database have shown the very high efficiency of the proposed ensemble system. The best accuracy in recognition of melanoma from non-melanoma cases obtained by the ensemble was 96.54% at AUC=0.9909, sensitivity 94.71%, and specificity 97.67%. These values are superior to the results presented for this ISIC2017 database.

Keywords: CNN structures, ensemble of classifiers, melanoma recognition.

1. INTRODUCTION

The image recognition is an important subject in supporting the medical diagnosis. It belongs to very difficult classification problems, of special interest to researchers.

To such problems belongs the recognition of dermoscopic images in melanoma. It is a very demanding task, because of a large variety of images in the same class and the close similarity of samples representing opposite classes. Former approach typically used by medical experts was based on application of ABCDE rules [1],[2],[3] assessing such factors as asymmetry (different shape of the image from the left and right as well as from bottom and upper side), border (irregular, blurry or ragged lesions), color (great changes of shades from brown to black, inconsistent pigmentation), diameter (usually greater than 6mm, and with progressive changes in size) and evolution representing history of changes over time.

The progress in information technology has allowed computer methods to support the recognition process. The earliest solutions relied on many preprocessing steps of images, including segmentation, definition of numerical descriptors of the image (such as color distribution statistics, wavelet analysis, color texture descriptors, global and dynamic thresholding), selection of diagnostic features, and the final step of recognition of lesions using classifiers. Different types of classification units, like K-nearest neighbors (KNN), naive Bayes, a random forest of decision trees, neural networks, fuzzy and neuro-fuzzy systems, support vector machines, etc., have been proposed [2],[3],[4],[5]. However, the results of such approaches are of limited accuracy and need further improvements.

Nowadays, the most effective classification systems use the idea of computational intelligence, especially deep architectures [5]. The convolutional neural networks (CNN) [6] play the most significant way in classification tasks of image recognition. Thanks to combining the automatic generation of features and final classification stage in one common structure they are very efficient in image analysis.

The single CNN structures, as well as different arrangements of the ensemble, are regarded as the most perspective [7],[8],[9]. This is due to the fact, that CNN networks integrate in a single architecture the automatic generation/selection of features and recognition of classes. Such an approach simplifies the classification task and leads to the improvement of the results in image recognition.

In the work of Esteva et al. [10] the CNN network results were presented for a very large number of samples of skin lesions (129450 clinical images used for training) obtained from 18 different physician-curated, open-access online repositories (including ISIC) and Stanford University Medical Center. Reported test results for 3 classes showed an average accuracy of 93.3%, which is still better than the results obtained by 21 board-certified dermatologists for the investigated database.

Codella et al. [9] have proposed an ensemble of CNN for recognizing melanoma from other samples representing the second class. They obtained the average accuracy measured on the additional 100 test images equal to 76%, sensitivity 82%, and specificity 62%.

The paper of Yuexiang and Linlin [11] presented a deep learning framework consisting of two fully convolutional residual networks to simultaneously produce the segmentation and classification results for the ISIC2017 dataset.

Different deep models of image preprocessing and classification have been presented in [12], [13], [14], [15], [16]. The best results have been obtained by using such deep CNN structures as Resnet152 [12], LCnet [13], Efficientnet, Inception., Resnet50, VGG, Mobilnet, Densenet [14], [15], [16].

Alenezi et al. in their work [7] have proposed a multi-stage recognition framework with a deep residual neural network and hyperparameter optimization-based decision model to recognize melanoma from non-melanoma. The declared efficiency highly depends on the size of the datasets. The larger the dataset the better accuracy. The best results corresponded to ISIC2020 of the largest number of samples.

*e-mai: stanislaw.osowski@pw.edu.pl

El-Khatib et al. have proposed in [17] a system composed of a few CNN networks using transfer learning. The reported accuracy and F1 score for the ISIC2020 database was equal to 93.50%.

This paper proposes a different approach to the recognition of medical images representing melanoma. It applies many different architectures of CNN combined in the ensemble form. The most important difference to the existing methods is the way, in which the members of the ensemble are selected. In the first step, the individual CNN structures are assessed and the best units according to the chosen quality measure are selected as the potential members of the ensemble. In the second step, different compositions of such members are studied and assessed by using the validation datasets. In the final step of experiments, most perspective sets create the ensembles of the final form subjected to checking.

The important step applied in this paper is also the proper segmentation of the original images included in the ISIC database. By applying the modified flood fill method [18], the region of interest containing only the lesion region has been extracted from the images. Thanks to this the diversified background region (sometimes occupying half of the image) representing the noise has been eliminated and did not take part in the recognition process.

The results of numerical experiments performed on the ISIC2017 database have shown very good performance of the proposed system. The best accuracy in recognition of melanoma from non-melanoma cases obtained by the ensemble was 96.54%, at AUC=0.9909, sensitivity 94.71%, and specificity 97.67%. All of them exceed the presented results in the other papers for this ISIC2017 database.

Very good results of melanoma recognition are due to the novel approach to the ensemble creation. The proposed system is formed from many different CNN architectures, precisely selected for the task. The differences in the structures allow them to provide high independence in their operation, leading to the improvement of the generalization ability of the system.

The novelty of the paper is also included in the efficient preprocessing of the original ISIC images, which leads to the accurate extraction of the regions of interest (ROI) containing only lesion regions, which are the most important in the recognition process.

2. DATABASE OF MELANOMA IMAGES

ISIC database ISIC2017 of melanoma images is used in this paper [19],[20]. It is an open-source public access archive of skin images to test and validate the methods in automated diagnostic systems. Among different versions of the available ISIC datasets (2017, 2019, 2020) the ISIC2017 seems to be the most demanding, since it contains the smallest population of data, hence the most difficult in class recognition using an automatic system.

Two classes of images have been considered in the experiments.

- Class 1 representing melanoma (945 images)
- Class 2 represents other, non-melanoma cases (1543 images).

Both datasets are only slightly unbalanced. The results of experiments have shown that this imbalance is not a problem for the proposed classification system. Therefore, no specialized methods (like GAN, variational autoencoder, or introducing noise to the images) have been used to enrich the learning data.

Figure 1 presents some examples of original images from this database representing melanoma and non-melanoma cases.

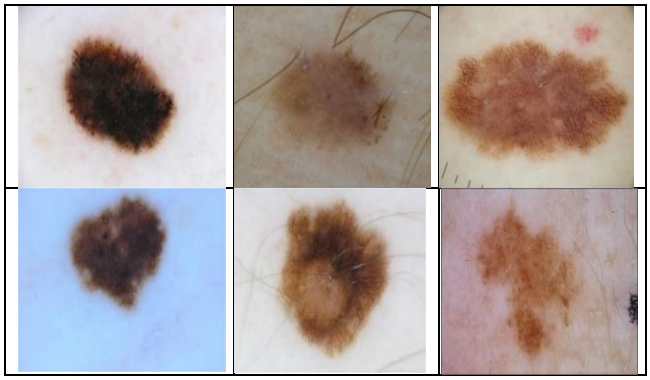

Fig. 1. The representative original images belonging to melanoma (upper row) and non-melanoma (bottom row) cases. The background occupies a large part of the images

The first and second rows of the figure represent melanoma and non-melanoma samples, respectively. The images contain not only lesion region, but also the background of different colors and diversified structures in each image. This part of the images is not important in the recognition of melanoma.

The regions of interest (ROI) corresponding to the true lesions differ significantly in the sample images belonging to the same class. The differences are visible in the color distribution, structure, and size of the ROI. The proportion of the area of ROI and the total area of the image is significantly different in particular samples. Moreover, some similarity of the ROIs corresponding to the opposite classes is also visible (see for example the ROI of melanoma and non-melanoma in the first or third column). To compare the melanoma and non-melanoma sets of images the statistical characterization of them has been done.

Table 1 shows values of the statistical parameters describing the pixel intensity in the total populations of melanoma and non-melanoma samples. They include mean value, standard deviation, energy, skewness, and kurtosis. The high similarity of these parameters characterizing both classes is evident. For example, the mean value for the melanoma class is 88.04 ± 22.82 and 82.75 ± 21.08 for the non-melanoma class. Even more similar are the values of energy: 12818 ± 5018 for melanoma and 12000 ± 5159 for non-melanoma.

TABLE 1.
The statistical parameters of the pixel intensity of images included in the melanoma and non-melanoma classes of the ISIC2017 database

	Mean	Standard deviation	Energy	Kurtosis	Skewness
Melanoma	88,04 ± 22,82	66,21 ± 12,73	12818 ± 5018	2,17 ± 1,01	-0,08 ± 0,68
Non-melanoma	82,75 ± 21,08	67,65 ± 11,5	12000 ± 5159	1,91 ± 0,49	0,14 ± 0,46

Very characteristic is the high value of standard deviation of all parameters in both classes. The typical is for example the standard deviation of the skewness, which exceeds the corresponding mean values a few times. The values presented in Table 1 are evidence of high differences among the images forming the same class and large similarity of images representing opposite classes.

In the first step of building an efficient classification system, we should eliminate the influence of the not important factors like the background and concentrate only on the ROI representing the true lesions. Therefore, the segmentation of the images aimed to extract the true ROI should be done in the first step.

3. SEGMENTATION STEP OF THE IMAGES

The original ISIC database was created by many institutions around the world and contains images of different proportions between the ROI corresponding to lesions and the background containing many undesirable factors. In some images, the background occupies more than half of the image size. In the recognition process, only the ROI plays an important role. Therefore, reducing the influence of the background in the image recognition process is very important. After doing it the extracted ROIs of the images are saved and used as the input data in further experiments.

The extraction of the ROI is the part of segmentation process, aimed at finding the regions of pixels representing the lesions. This step is done here by using the modified region growing procedure [18], called flood fill algorithm (FF). It is directed to create the mask covering the lesions region of the image. The input to the procedure is the original RGB image (Img_{RGB}) and the output – the mask Img_{mask} representing the pixels forming the recognized ROI.

In the first step, the RGB image is converted to a grayscale image Img_{gs}. The FF algorithm assumes that the neighboring pixels of Img_{gs} are characterized by similar levels of intensity. The flooding procedure starts from two different reference regions Ref_a and Ref_b of the grayscale image. They aim in two opposite directions. The Ref_a represents the region of the image outside a circle of the radius R_a defined by

$$R_a = \frac{\max(iw, ih)}{2} \cdot 0.8 \qquad (1)$$

where iw and ih represent the width and the height of the image. The starting point corresponds to the highest mean intensity level of the area. The flooding process is directed toward the center of the image.

On the other side, the Ref_b covers the inside of the region of the lowest mean intensity level with a constant value R_b=50 pixels and is directed outside the center. The parameter values used in the definition of R_a and R_b have been obtained in the introductory experiments.

Both FF processes starting from regions Ref_a and Ref_b are applied simultaneously. The similarity measure $K(x,y)$ based on the neighboring pixel intensity values in the reference areas is calculated using the expression.

$$K(x,y) = \frac{\mathrm{Im}\,g_{gs}(x,y)}{avg(\mathrm{Ref}(\mathrm{Im}\,g_{gs}))} \cdot 255 \qquad (2)$$

The pixels of the similar values of this measure are merged in both reference areas Ref_a and Ref_b, respectively.

The FF processes in both regions are executed until their areas meet. The border points of both FF areas, define the boundary of the neoplastic lesions corresponding to ROI. In the next step, the image is cropped from four sides (up, down, left, right) until the final size of the mask, covering ROI is obtained. In the last step, the final mask Img_{mask} is filled by the area of pixels existing in the original image Img_{RGB}.

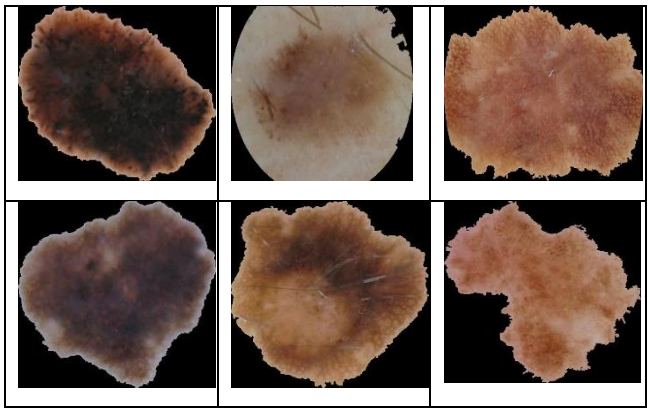

Fig.2. The sample images representing the cropped original images obtained by using the presented algorithm: the first row presents the melanoma and the second one – is the non-melanoma cases. They correspond to the original ISIC images depicted in Fig, 1

Figure 2 presents examples of the cropped segmented images representing different neoplastic lesions corresponding to melanoma and non-melanoma cases. They correspond exactly to the ISIC original images represented by Fig. 1. The first row represents melanoma and the second the non-melanoma. It is evident that the ROIs representing now lesion regions occupy the maximum part of the images, and the background areas are limited to a minimum.

4. ENSEMBLE SYSTEMS BASED ON DIFFERENT CNN ARCHITECTURES

The ROI images extracted in the segmentation procedure create the set of data used in the recognition process. The images of melanoma represent class 1 and non-melanoma samples the opposite class 2.

The classification system proposed in the paper will apply different architectures of convolutional neural

networks organized in the form of an ensemble. It is well known [21], that many classifiers combined in an ensemble and properly aggregated, may generate improved results, even with respect to the best individual member. However, the most important condition to achieve it is to provide the independent operation of its members. This can be obtained in different ways, for example by applying different types of classifiers, different sets of learning data used in the training process, diversified set of input attributes, etc.

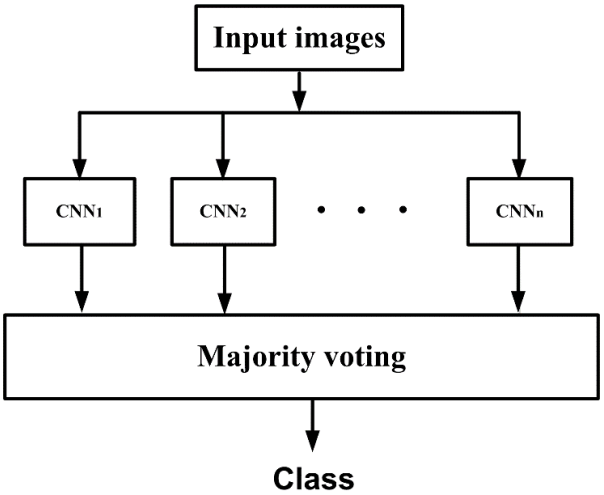

Fig. 3. The general structure of the proposed ensemble. The number of units and their composition are chosen by considering their achievements

In our solution, we have applied the set of CNN classifiers of different architectures and used transfer learning. The pre-trained structures available in Matlab [22] have been adjusted to the actual problem. The first 30% of the hidden, locally connected layers have been left unchanged. The other layers, including the fully connected structure of the softnet, have been subjected to adaptation using the actual learning data in a 5-fold cross-validation technique. The time of adaptation of the particular CNN units forming an ensemble was changing a lot, from a few minutes in Alexnet to a few hours in the case of Nasnetlarge. These values correspond to the typical laptop containing GPU.

The general structure of the proposed ensemble is presented in Fig. 3. The choice of the members and their number is based on their achievements in the class recognition.

To achieve the highest efficiency of the ensemble the set of CNN classifiers was carefully selected based on the achievements of the individual members. Since training the CNN from scratch needs a huge population of learning data, we have used the pre-trained units in the transfer learning mode available in Matlab [22]. The following CNN architectures were included in the pool and considered as the potential members of an ensemble: Squeezenet, Googlenet, Inceptionv3, Densenet201, Mobilnetv2, Resnet18, Resnet50, Resnet101, Xception, Inceptionresnetv2, Shufflenet, Nasnetmobile, Nasnetlarge, Darknet19, Darknet53, Efficientnet, Alexnet, VGG16 and VGG19. The CNN classifiers considered as the potential members of the ensemble differ in the way the data are processed, the number of layers, the types of applied filters, etc. Therefore, their independence in the class recognition process is relatively high. The important point is to estimate the proper size of their population to obtain the best performance of the ensemble.

5. STATISTICAL RESULTS OF NUMERICAL EXPERIMENTS

In the first phase of experiments, the individual classifiers have been retrained on the database set described in the previous section. In the first stage, the samples of the dataset were randomized applying the uniform distribution. Next, the 5-fold strategy has been applied. It means the set of the sample images has been divided into 5 subsets. Four of them have been used for learning and the fifth one in testing. In each fold, the testing subset was changing. Only the results of testing are considered in the assessment of the classification results. All CNN architectures have been trained using the ADAM algorithm implemented in the Matlab platform [22].

The assessment of the quality of classifiers was based on the following parameters: accuracy (ACC), area under the ROC curve (AUC), true positive rate (TPR), true negative rate (TNR), positive precision value (PPV), and negative precision value (NPV) [23],[24]. Table 2 presents the average values of these parameters for the testing data achieved by the considered candidates for the ensemble. The standard deviations between the succeeding folds were very small (below 1%), hence their values are omitted in further presentation of results in the tables. Additionally, the confusion matrix in the last column for each classifier is also included.

The results show high differences in the efficiency of the particular solutions of CNN classifiers. The best results correspond to the Resnet101 and the worst to Alexnet. For example, the average accuracy ACC in the test changed from 79.26% (Alexnet) and 80.47 (Squeezenet) to 95.78% for Resnet101 and 94.69% for Nasnetlarge.

The CNN architectures of the lowest efficiency are excluded from the consideration (for example Alexnet and Squeezenet). Only the members of comparable results are considered for the set. The aggregation of the results of the ensemble members was based on the majority voting. In this process, the probability values of class membership pointed out by the members have been used instead of their binary translation. The ith class probability $p(i)$ is calculated by summing the proper pointing $p(i, j)$ of the M members forming the ensemble.

$$p(i) = \sum_{j=1}^{M} p(i, j) \tag{3}$$

The process of choosing the best composition of the ensemble members may consider different parameters of the quality. In this solution, we have studied three of the most important factors: ACC, AUC, and TPR.

TABLE 2.

The results of the experiments show the mean values of the quality measures obtained by the candidates for the ensemble. The best results corresponding to the Resnet101 are depicted in bold

CNN	AUC	ACC	TPR	TNR	PPV	NPV	Confusion. matrix	
1) Squeezenet	0.8726	0.8047	0.6857	0.8775	0.7742	0.8201	648	297
							189	1354
2) Googlenet	0.9451	0.8585	0.8455	0.8665	0.7950	0.9016	799	146
							206	1337
3) Inceptionv3	0.9861	0.9421	0.9312	0.9488	0.9176	0.9575	880	65
							79	1464
4) Densenet201	0.9813	0.9401	0.9164	0.9546	0.9252	0.9491	866	79
							70	1473
5) Mobilenetv2	0.9743	0.9244	0.8730	0.9559	0.9239	0.9248	825	120
							68	1475
6) Resnet18	0.9655	0.8983	0.8889	0.9041	0.8502	0.9300	840	105
							148	1395
7) Resnet50	0.9817	0.9317	0.9037	0.9488	0.9153	0.9415	854	91
							79	1464
8) Resnet101	**0.9865**	**0.9578**	**0.9238**	**0.9786**	**0.9636**	**0.9545**	**873**	**72**
							33	**1510**
9) Xception	0.9777	0.9196	0.8720	0.9488	0.9125	0.9237	824	121
							79	1464
10) Inceptionresnetv2	0.9819	0.9349	0.8868	0.9644	0.9384	0.9329	838	107
							55	1488
11) Shufflenet	0.9731	0.9184	0.9132	0.9216	0.8770	0.9455	863	82
							121	1422
12) Nasnetmobile	0.9767	0.9172	0.8857	0.9365	0.8952	0.9305	837	108
							98	1445
13) Nasnetlarge	0.9909	0.9469	0.9280	0.9585	0.9320	0.9560	877	68
							64	1479
14) Darknet19	0.9702	0.9043	0.9048	0.9041	0.8524	0.9394	855	90
							148	1395
15) Darknet53	0.9745	0.9232	0.8825	0.9482	0.9125	0.9295	834	111
							80	1463
16) Efficientnetb0	0.9572	0.8947	0.7968	0.9546	0.9149	0.8847	753	192
							70	1473
17) Alexnet	0.8539	0.7926	0.6349	0.8892	0.7782	0.7991	600	345
							171	1372
18) VGG16	0.9403	0.8473	0.8709	0.8328	0.7613	0.9133	823	122
							258	1285
19) VGG19	0.9498	0.8830	0.8265	0.9177	0.8601	0.8962	781	164
							127	1416

TABLE 3.
The CNN architectures form different compositions of the ensemble.

No	Quality	Number of units	CNN architectures forming the ensemble
1	AUC	3	[3,8,13]
	ACC	3	
	TPR	3	
2	AUC	5	[3,8,13,10,7]
3	ACC	5	[3,8,13,4,10]
4	TPR	5	[3,8,13,4,11]
5	AUC	10	[3,7,8,10,13,4,9,12,15,5]
6	**ACC**	**10**	**[3,8,13,4,10,7,5,15,9,11]**
7	TPR	10	[3,8,13,4,11,14,7,6,10,12]
8	AUC	15	[3,7,8,10,13,4,9,12,15,5,11,14,6,16,19]
9	ACC	15	[3,8,13,4,10,7,5,15,9,11,12,14,6,16,19]
10	TPR	15	[3,8,13,4,11,14,7,6,10,12,15,5,9,18,2]
11	-	19	All CNN architectures

TABLE 4.
The results of the efficiency of different compositions of the ensemble. They are represented by AUC, ACC, TPR, TNR PPV, NPV, and the confusion matrix

Ensemble	**AUC**	**ACC**	**TPR**	**TNR**	**PPV**	**NPV**	**Confusion Matrix**	
1	0.9811	0.9582	0.9312	0.9747	0.9576	0.9586	880	65
							39	1504
2	0.9866	0.9634	0.9376	0.9793	0.9651	0.9624	886	59
							32	1511
3	0.9866	0.9626	0.9354	0.9793	0.9651	0.9612	884	61
							32	1511
4	0.9867	0.9602	0.9397	0.9728	0.9548	0.9634	888	57
							42	1501
5	0.9926	0.9626	0.9429	0.9747	0.9581	0.9653	891	54
							39	1504
6	**0.9909**	**0.9654**	**0.9471**	**0.9767**	**0.9613**	**0.9679**	**895**	**50**
							36	**1507**
7	0.9915	0.9590	0.9460	0.9669	0.9460	0.9669	894	51
							51	1492
8	0.9924	0.9598	0.9323	0.9767	0.9607	0.9593	881	64
							36	1507
9	0.9924	0.9598	0.9323	0.9767	0.9607	0.9593	881	64
							36	1507
10	0.9922	0.9598	0.9386	0.9728	0.9548	0.9628	887	58
							42	1501
11	0.9908	0.9546	0.9249	0.9728	0.9541	0.9548	874	71
							42	1501

The introductory experiments have shown that too small or too large sets did not improve the results at a sufficient rate. Different populations of ensemble members have been investigated. The choice of units was based on their mentioned quality measures: ACC, AUC, and TPR.

After many introductory simulations, 11 combinations of the best units have been selected in the study. Their compositions are presented in Table 3.

The CNN structures are represented here by the numbers: Squeezenet (1), Googlenet (2), Inceptionv3 (3), Densenet201 (4), Mobilnetv2 (5), Resnet18 (6), Resnet50 (7), Resnet101 (8), Xception (9), Inceptionresnetv2 (10), Shufflenet (11), Nasnetmobile (12), Nasnetlarge (13), Darknet19 (14), Darknet53 (15), Efficientnet (16), Alexnet (17), VGG16 (18) and VGG19 (19). The sequence of their appearance in the set corresponds to their quality measures. Three different quality measures corresponding to the best units have been considered in creating the ensemble (column 2). One considered ensemble was formed from all 19 units (the last row in the table). All proposed compositions of the ensemble used in experiments are presented in Table 3.

The results of the performance of different compositions of ensemble represented by all quality values AUC, ACC, TPR, TNR PPV, NPV, and confusion matrices are shown in Table 4. They correspond to the testing data only and have been obtained in the 5-fold cross-validation approach. The best solution, pointed out in the table by the bold numbers, corresponds to the ensemble number 6, composed of 10 units: Resnet101, Inceptionv3, Nasnetlarge, Densenet201, Inceptionresnetv2, Resnet50, Mobilnetv2, Darknet53, Xception and Shufflenet. Such a combination resulted in the top results for accuracy and sensitivity.

Figure 4 illustrates the performance of different compositions of the ensemble concerning the accuracy ACC, the area under the ROC curve AUC), and the sensitivity of melanoma recognition TPR. The horizontal axis represents the succeeding compositions (from 1 to 11) mentioned in Table 4. Irrespective of the composition of the ensemble the results are on a very high level (all above 92%)..

Observe, that all compositions of the ensemble have improved the best individual result of AUC corresponding to Resnet101. The accuracy value obtained by all ensembles (except one including all CNN architectures) also outperformed the best individual result of ACC=0.9578.

The best result of ACC of the ensemble number 6 is equal to ACC=0.9654. It outperformed the average value of its members ACC_{av}=0.9339±0.1292. The least efficient was the ensemble composed of all 19 units. However, even in this case the obtained accuracy ACC=0.9546 was higher than the average of its all members, ACC_{av}=90.08%±4.67%.

The improved values of other quality measures (TPR, NPR, PPV, and NPV) have been observed for all compositions of an ensemble. A very important advantage is the high reduction of critical errors (recognition of melanoma as non-melanoma). The best individual unit (Resnet101) has committed 72 such misclassifications. The best ensemble number 6 has reduced this number to only 50.

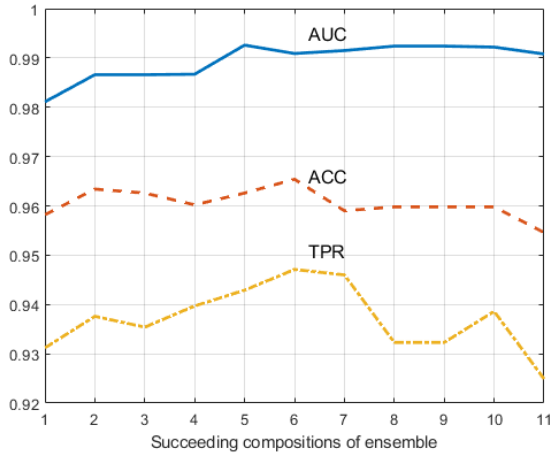

Fig. 4. The illustration of the performance of different compositions of the ensemble concerning AUC, ACC, and TPR. The horizontal axis represents the succeeding ensembles (from 1 to 10)

A very important advantage of an ensemble is the high reduction of critical errors (recognition of melanoma as non-melanoma). The best individual unit (Resnet101) has committed 72 such misclassifications. The best ensemble number 6 has reduced this number to only 50.

6. DISCUSSION OF RESULTS

It is interesting to compare our best results (accuracy ACC=96.54%, sensitivity TPR=94.71%, specificity TNR=97.67%, positive precision value PPV=94.16%, negative precision value NPV=96.79% and AUC=0.9909) with these presented in the international publications for the same database ISIC2017.

The paper of Yang et al. [8] has declared only the area under the ROC curve AUC = 0.880 for the ISIC2017 database.

The paper of Yuexiang and Linlin [11] presented the results of the deep learning framework for the ISIC2017 dataset, declaring an accuracy of 85.7%, sensitivity of 49%, specificity of 85.7%, and AUC of 0.912.

The results of the Acosta et al. paper [12] obtained for the ISIC2017 dataset by applying Resnet152 were as follows: ACC=90.4%, TPR=82%, and TNR= 92.5%.

The interesting is the comparison of results for different versions of ISIC presented by Kaur et al. in the paper [13] by using deep CNN LCnet. The best recognition accuracy values obtained for these sets were: 81.41% (ISIC 2016), 88.23% (ISIC 2017), and 90.42% (ISIC2020). The succeeding version of ISIC is of a larger population.

The recent results of the Dutta et al. work [15] obtained for the ISIC2017 were as follows: AUC=0.87, sensitivity 73%, class precision 76%, and F1=74%.

Our results obtained by the best ensemble in all considered quality measures are superior to these presented for this ISIC2017 database in the mentioned papers.

7. CONCLUSIONS

The paper has proposed a novel architecture of an ensemble composed of many different deep CNN structures. The applied ensemble members differ in many aspects of signal processing (organization of layers, the width and depth of the network, number and size of filters, different types of activation functions, etc.). Therefore, in the recognition process, they are concentrated on different aspects of the analyzed images. Thus, the units are highly independent in their assessment of the input image.

As a result, the classification verdicts of the members are diverse, which provides a good perspective for improving the generalization ability of the system, by applying the procedure of fusing their results.

The numerical experiments performed using the ISIC2017 database have shown very high efficiency of the system created from the precisely selected CNN architectures. Moreover, the ensemble reduces significantly the most severe misclassification cases (melanoma recognized as non-melanoma). For example, the best individual unit (Resnet101) has committed 72 such misclassifications, while the ensemble has shown only 50 errors.

The presented procedure of creating the optimal ensemble is universal and applicable to any problems of class recognition. It may be useful in different areas of research, not limited to medical problems. Moreover, it is not restricted to the two class recognition tasks.

REFERENCES

[1]. S. E. Yagerman, L. Chen, N. Jaimes, S. W. Dusza, A. C. Halpern and A. Marghoob, "Do UC the melanoma?' Recognizing the importance of different lesions displaying unevenness or having a history of change for early melanoma detection*", Australas J. Dermatol, vol.* 55, pp. 119–24, 2014.

[2]. A. G. Manousaki, A. G. Manios, E. I. Tsompanak and A. D. Tosca, "Use of color texture in determining the nature of melanocytic skin lesions—a qualitative and quantitative approach", *Computers in Biology and Medicine, vol.* 36, pp. 419–427, 2006.

[3]. A. Zakeri and A. Hokmabadi, "Improvement in the diagnosis of melanoma and dysplastic lesions by introducing ABCD-PDT features and a hybrid classifier", *Biocybernetics and Biomedical Engineering, vol.* 38, pp. 456–466, 2018.

[4]. R. Garnavi, M. Aldeen and J. Bailey, "Computer-aided diagnosis of melanoma using border and wavelet-based texture analysis", *IEEE Transactions on Information Technology in Biomedicine*, vol.16, no. 6, pp. 1-13, 2012.

[5]. S. Osowski, B. Sawicki and A. Cichocki, "Computational intelligence in engineering practice", *Bulletin of the Polish Academy of Sciences, Technical Sciences, vol.* 69, no. 3, pp. 1-5, doi: 10.24425/bpast.2021.137052 (2021).

[6]. I. Goodfellow, Y. Bengio and A. Courville, "*Deep Learning*", MIT Press, Massachusetts, 2016.

[7]. F. Alenezi, A. Armghan and K. Polat, "A multi-stage melanoma recognition framework with deep residual neural network and hyperparameter optimization-based decision support in dermoscopy images", *Expert Systems with Applications*, vol. 215, pp. 1-10, 2022, 119352, https://doi.org/10.1016/j.eswa.2022.119352.

[8]. X. Yang, Z. Zen, S. Y. Yeo, C. Tan, H. L. Tey and I. Su, "A novel multitask deep learning model for skin lesion segmentation and classification", *http://arXiv.org.abs/1610.04662*, 2017.

[9]. N. C. Codella, Q. B. Nguyen, S. Pankanti, D. Gutman, B. Helba, A. Halpern and J. R. Smith, "Deep learning ensembles for melanoma recognition in dermoscopy images", *IBM J. of Research and Development, vol.* 61, no. 4, 2017.

[10]. A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. Swetter, H. Blu and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks*", Nature*, vol. 542, pp. 115–118, 2017.

[11]. L. Yuexiang and S. Linlin., "Skin lesion analysis towards melanoma detection using deep learning network". *Sensors* (Basel), vol. 18, pp. 1-8, 2018.

[12]. M.F.J Acosta., L.Y.C Tovar., M.B., Garcia-Zapirain and W. Percybrooks, "Melanoma diagnosis using deep learning techniques on dermatoscopic images", *BMC Med Imaging*, vol. 21, pp. 213-224, 2021.

[13]. R. Kaur, H. Gholamhosseini, R. Sinha and M. Linden M., "Melanoma classification using a novel deep convolutional neural network with dermoscopic images", *Sensors, vol.* 22, no. 3, 2022, doi.org/10.3390/s22031134.

[14]. N. Joson N and M. S. Nair, "On the performance of CNN and GAN models for melanoma classification", in Proc. *International Conference on Artificial Intelligence and Signal Processing* (AISP), 2022, doi: 10.1109/AISP53593.2022.9760626.

[15]. A. Dutta, M. Kamrul Hasan and M. Ahmad, "Skin lesion classification using convolutional neural network for melanoma recognition", in M. S. Uddin, and J. C. Bansal, Eds. *Algorithms for Intelligent Systems*,. Springer, Singapore, 2021.

[16]. G. Alwakid, W. Gouda and M. Humayun, "Melanoma detection using deep learning-based classifications", *Healthcare, vol.* 10, no. 12, pp. 234-240, 2022.

[17]. H. El-Khatib, A. M. Stefan and D. Popescu, "Performance improvement of melanoma detection using a multi-network system based on decision fusion", *Applied Sciences*, vol. *13, no.* 18, 2023, 10536; doi.org/10.3390/app131810536.

[18]. S. Osowski and T. Les, "Deep Learning Ensemble for Melanoma Recognition", in Proc. *Intern. Joint Conf. on Neural Networks* (*IJCNN*) , Glasgow, pp. 1-7, 2020.

[19]. https://isisc-archive.com.

[20]. https://challenge.kitware.com/#challenge/n/ISIC_2017%3A_Skin_Lesion_Analysis_Towards_Melanoma_Detection.
[21]. L. Kuncheva, "*Combining Pattern Classifiers: Methods and Algorithms"*, Wiley, 2014.
[22]. Matlab2023, Mathworks, Natick, USA, 2023.
[23]. P. N. Tan, M. Steinbach and V. Kumar, "*Introduction to Data Mining"*, Pearson Education Inc., Boston, 2014.
[24]. S. Osowski and R. Szmurło, "*Matematyczne modele uczenia maszynowego w językach Matlab i Python*", Oficyna Wydawnicza PW, Warszawa, 2023.