Deep Learning Ensemble for Melanoma Recognition

Stanislaw Osowski
Military University of Technology
Warsaw University of Technology, Warsaw, POLAND
stanislaw.osowski@ee.pw.edu.pl

Tomasz Les
Warsaw University of Technology, Warsaw, POLAND
tomasz.les@ee.pw.edu.pl

***Abstract*—The paper presents the deep learning ensemble of classifiers in recognition of melanoma on the basis of dermoscopy image analysis. The ensemble is based on 9 units supplied by the activation signals from the convolutional neural network. To provide the independence of unit operation few different feature selection methods combined with three types of classification networks have been used. The pre-trained Alexnet CNN structure has been used in this application. The experiments have been performed using two data bases in recognition of melanoma and non-melanoma cases. One of them is very well known large ISIC base and the second smaller data base collected in Warsaw Memorial Cancer Center and Institute of Oncology. The results have shown advantage of the ensemble over individually running classifiers. The accuracy was increased by few percentage points.**

***Keywords*—CNN, ensemble of classifiers, melanoma recognition.**

I. INTRODUCTION

Melanoma is the most dangerous skin cancer in world-wide extent [1]. When not cured in proper time it may spread and lead finally to death. The recognition of melanoma is usually done on the basis of dermoscopic images of the lesions. However, recognition of the melanoma from non-melanoma changes on the basis of such image is still very difficult task, due to many reasons, such as visual similarity between melanoma and non-melanoma lesions, large differences between images representing samples of the same class, differences in color, shape, size of images, etc. Therefore, the automatic systems, which are able to recognize between melanoma and non-melanoma changes are very useful in medical practice.

The typical aspects of image characterization, proposed by medical experts form the so called ABCDE rules [2],[3],[4]. They consider: 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), evolution representing history of changes over time.

In conventional computer aided melanoma recognition the additional features, that characterize the images, such as percolation descriptors and maximum subregions measures based on Kolmogorov-Smirnov distance [7], as well as textural and color features [5],[6] are also added to enhance the image description.

Many works devoted to automatic melanoma recognition have proposed various solutions, differing by the methods of defining structures of the classifiers. They include clustering approach, linear discriminant analysis, neural networks, fuzzy and neuro-fuzzy systems, support vector machines (SVM), K-nearest neighbors (KNN), naïve Bayes, random forest, convolutional neural networks, etc. [8], [9],[10].

The reported sensitivity and accuracy vary from case to case and depend greatly on the applied data base. For example, in [9] Sabouri et al., have shown a sensitivity of 83.06% and specificity of 90.05% in the solution employing cascade classifiers in recognition of melanoma from non-melanoma. The results presented in [6] show an accuracy of 91.26% on the set of 289 dermoscopic images (114 malignant, 175 benign), partitioned into train, validation and test image subsets. In [10] authors have shown application of combination of ABCDE parameters, pigment distribution and texture features combined with the ensemble of 5 classifiers. The declared accuracy was 96.8% in recognition of melanoma from non-melanoma. It should be stressed, that the results of classification are dependent not only on the applied method of image processing, but first of all on the data base used in the investigations.

Especially difficult is the International Skin Imaging Collaboration (ISIC) data base available in Internet [15],[16], which was investigated in some papers [18],[19]. The results presented in these works differ significantly, depending on the chosen subset of data and the applied method. However, the best results have been achieved by using deep learning in the form of CNN and ensemble of such networks. In [18] Esteva et al., have presented results of application of CNN networks for large number of samples (129450 clinical images used in training), which come from 18 different clinician-curated, open-access online repositories (including ISIC), as well as from Stanford University Medical Center. The validation results referring to 3 class recognition problem have shown an average accuracy of $72.1 \pm 0.9\%$, which was better than the results of 21 board-certified dermatologists. In [11] authors declared the best validation score of 76% for recognition of six classes of data from ISIC 2018 data base using ensemble of CNN networks. In [20] Yang et al., declared the area under ROC curve AUC = 0.880 for two classes of images (melanoma versus non-melanoma) taken from ISIC 2017 data base. In [21] authors have proposed an ensemble of CNN networks, which combined the whole image context with the cropped region context from the ground true segmentation of images in the data base containing 900 training and 379 testing images

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

representing two classes. The declared AUC in recognition of 2 classes of samples was equal 0.843 and precision 64.9%. The average accuracy measured on 100 test images was equal 76%, sensitivity 82% and specificity 62%.

This paper presents a new ensemble approach organized in a different way. The main tool is the pre-trained CNN Alex network [14]. The activations obtained from fc7 (fully connected layer) are subject to selection using four selection methods (stepwise fit, nearest neighbor analysis, reliefF and Fisher discriminant). Each of the selection methods create different set of features, which are next supplied as the input attributes to two types of classifiers: support vector machine (SVM) and random forest (RF). These 8 members of ensemble are supplemented by the classical softmax classifier based on randomly selected activation signals of CNN. In this way the ensemble is built from 9 members. Final decision of ensemble is taken by majority voting.

The results obtained for both ISIC and our own data base have shown high advantage of the proposed solution over individual results. The accuracy of this 2-class recognition problem was much better than the results of individual classifiers forming the same ensemble.

The rest of the paper is organized as follows. Section 2 describes two data bases used in experiments. Section 3 is devoted to the segmentation of images. Section 4 presents the ensemble of classifiers used in image recognition problem. Section 5 is devoted to presentation of the results of numerical experiments. The concluding section summarizes the presented considerations.

II. DATA BASES OF MELANOMA IMAGES

Two data bases have been used in experiments. The first one is ISIC data base of melanoma images [15]. It is an open source public access archive of skin images to test and validate the proposed methods in automated diagnostic systems.

The data base used in experiments contained images representing two classes:

- Melanoma – 945 images
- non-melanoma – 1543 images

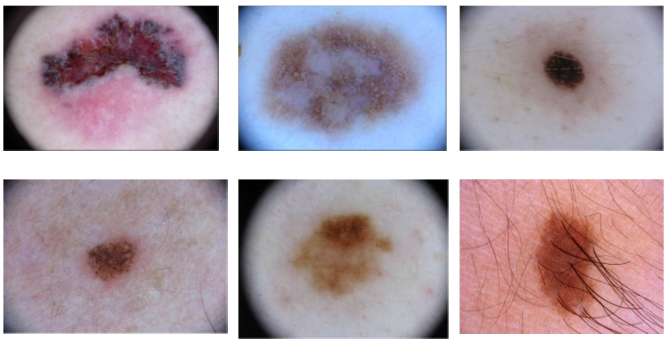

Fig. 1. The representative images belonging to melanoma (upper row) and non-melanoma (bottom row) classes.

The original exemplary images taken from ISIC data base are presented in Fig. 1. They cover not only lesion region, but also the background of no importance in recognition process. The first and second row represent melanoma and non-melanoma samples, respectively. The variety of shapes and colors of the lesions are visible. They are differently distributed within the image, the background in each image is different and the size and shape of region of interest representing the skin lesions change from sample to sample. Moreover, we can see high similarity of the samples representing two opposite classes.

In this particular data base the first task is to segment the regions of lesions (so called ROI), replacing the background pixels by zero values and leaving only the most important parts of the images, which carry the information of skin changes. The segmented regions will be saved in separate files and used in recognition of melanoma.

The second data base was collected in Warsaw Memorial Cancer Center and Institute of Oncology, Department of Soft Tissue/Bone Sarcoma and Melanoma. The images are also acquired in the form of dermoscopic images, coded in RGB format. The size of images was varying from 465×599 to 1077×1899 pixels. All of them have been stored in JPEG format. The acquired dermoscopic image base contains 174 samples representing the melanoma (lentigo maligna melanoma and nodular melanoma) and 186 of non-melanoma (seborrheic keratosis, angioma, pigmented nevus and atypical nevus). This way the total number of images was 360. All cases have been qualified to the proper class by experienced dermatologists at application of ABCDE criteria [3]. Their results have been confirmed by pathomorphological analysis and histological diagnosis.

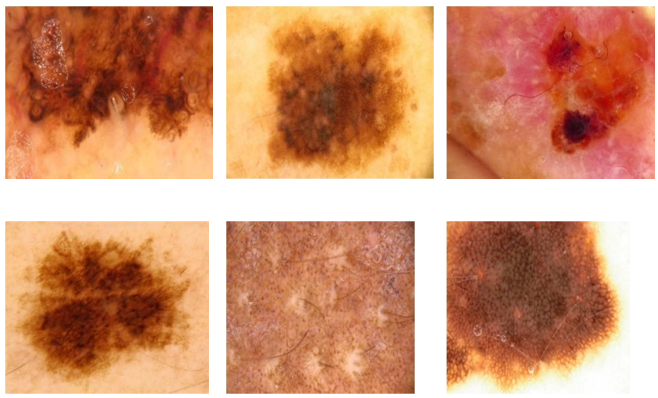

Fig. 2. The chosen examples of melanoma and non-melanoma cases: the first row images represent melanoma and second ones – the non-melanoma cases.

Fig. 2 presents some chosen examples of images, representing both classes. The samples show the lesions acquired from both adults and children. They come from different parts of the body: face, back, hand, etc. Some cases representing nevus were innate and come into existence at later years of life of the patients. Once again, large differences among the images belonging to the same class can be observed.

The images prepared by dermatologists were cropped by them to the most important lesion regions. Therefore, there was no need to apply the additional segmentation procedure. As a result the whole images from this base have been used in further experiments.

III. SEGMENTATION OF IMAGES

The original melanoma ISIC images are given in a general form with possible presence of many undesirable factors and the background area representing no information connected with the skin changes. The first step is to crop the image to only lesion region, and then train the system using only this area of the images. Therefore, segmentation is one of the important stages for computer aided diagnosis of melanoma with dermoscopic images. This step may be regarded as elimination of the background "noise". Choosing the right segmentation technique is a very important factor in cropping the image.

The process of an automatic localization of neoplastic lesion in the analyzed image is based on the modified region growing procedure called flood fill algorithm [22]. The assumption of the algorithm is the following: the input is a digital RGB image – $\mathrm{Img}_{\mathrm{rgb}}$ and the output is a white mask imposed on the pixels of this image, which belong to the recognized neoplastic lesion – $\mathrm{Img}_{\mathrm{mask}}$. This mask will be used to segment proper region of the original color image.

The applied algorithm exploits the fact that neighboring pixels have a similar gray level. The algorithm is applied for the gray scale version of the image. The RGB image is first converted to the gray scale $\mathrm{Img}_{\mathrm{gsi}}$ according to

$$\mathrm{Img}_{gsi}(x,y) = \mathrm{Img}_{r}(x,y) + \mathrm{Img}_{g}(x,y) + \mathrm{Img}_{b}(x,y) \qquad (1)$$

Two parallel flood fill processes are applied (towards outside and inside) on the gray scale image, starting from two chosen reference areas Ref_a and Ref_b. The algorithm calculates the similarity measure $K(x,y)$ based on the pixel values of reference areas for all neighboring pixels. The value $K(x,y)$ of the (x,y) pixel is calculated according to

$$K(x,y) = \frac{\mathrm{Img}_{gsi}(x,y)}{avg(\mathrm{Ref}(\mathrm{Img}_{gsi}))} \cdot 255 \qquad (2)$$

The areas of similar value of $K(x,y)$ are merged together in both reference areas, respectively. The areas Ref_a and Ref_b have been defined in the following way. The reference Ref_a has been assumed as a region outside a circle of the radius R_a defined by

$$R_a = \frac{\max(iw, ih)}{2} \cdot 0.8 \qquad (3)$$

where iw is the image width and ih is the image height. It starts from the region of the highest mean intensity level of the image. The Ref_b represents the area of pixels within a circle of a constant radius $R_b = 50\ px$, starting from the region of the lowest mean intensity level of the image. These values were determined based on the analysis of some chosen images (around 10% of the available data base).

The flood fill processes are executed in parallel (alternately) until both areas meet together. At the border points of both flood fill areas, a created boundary defines the found object (neoplastic lesion). Next, the image is pruned from four sides (up, down, left, right) until proper size of image is obtained. The final Img_{mask} is filled by the new area of a texture taken from original image Img_{rgb}. The scheme of the algorithm for finding the mask Img_{mask}, which represents the lesion, is depicted in Fig. 3.

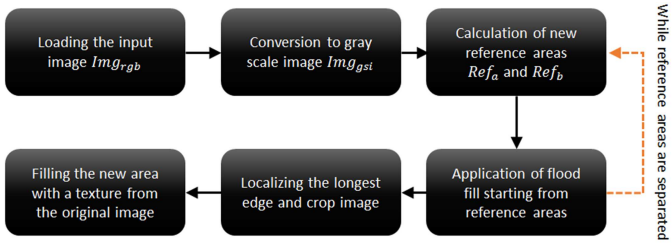

Fig. 3. The scheme of algorithm for finding Img_{rgb} mask.

The illustration of the consecutive steps of the algorithm is depicted in Fig. 4. Fig. 4a presents the original RGB image. The next image in Fig. 4b is its version in gray scale. The following pictures (from 4c to 4f) illustrate the process of growing both areas: Ref_b starting from red circle (Fig. 4c) and Ref_a representing the green area. Both areas walk toward themselves and the meeting border is presented in Fig. 4g.

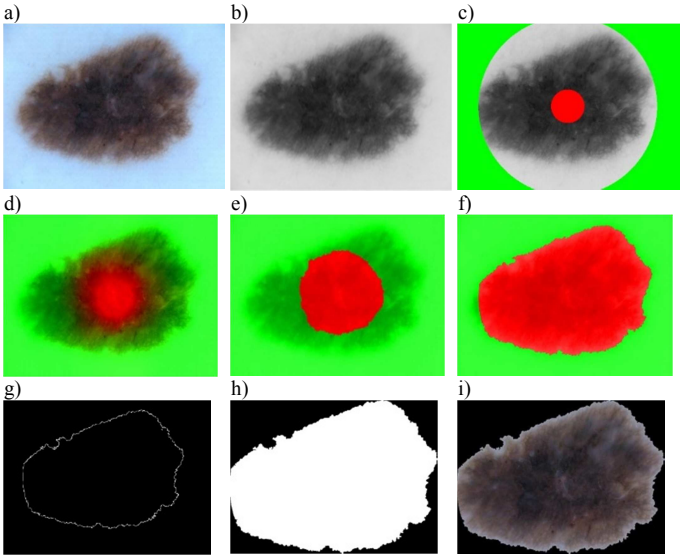

Fig. 4. The consecutive steps of the algorithm for finding the cropped image representing the lesion.

This border defines the limitation of lesion. Fig. 4h represents the white mask and Fig. 4i – the final ROI of the image in RGB scale, extracted by the algorithm. The pixels in the area outside the contour of lesion have been transformed to zero values.

Fig. 5 presents some examples of the segmented images representing different neoplastic lesions. They characterize both: melanoma and non-melanoma cases (the first row – melanoma and the second row – non-melanoma). The lesion region in each image occupies now the maximum part of the image. The background area, carrying no diagnostic information, is limited to minimum.

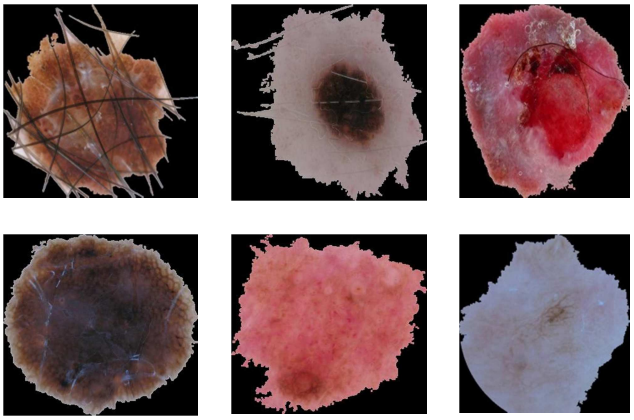

Fig. 5. The sample images representing the cropped original images obtained by using the presented algorithm: the first row presents the melanoma and the second one – the non-melanoma cases.

IV. Ensemble of Classifiers

The ensemble of classifiers has been arranged on the basis of 9 members, all strictly connected with deep convolutional neural network built on the basis of pre-trained AlexNet [12],[14]. It was trained on more than a million images from the ImageNet database. The crucial condition in arranging the ensemble of classifiers is to provide the independent operation of its members. This was achieved in various ways:

- Each classifier was trained on a set of diagnostic features selected in different ways.
- The ensemble members have applied three different types of classifiers: softmax, support vector machine (SVM) and random forest of decision trees (RF).
- Each classifier was trained on different, randomly selected sets of learning images.
- In the case of softmax the structure of fully connected network was varying, by applying different number of hidden neurons in fc7 layer (changing from 300 to 1500), as well as using different dropout ratio (from 0.3 to 0.6)

The ensemble was composed of 9 units, based on the CNN structure as the starting fundamental building block for all classifiers. The data, processed by the few local convolutional layers of CNN, were used for creating the input attributes to particular members of ensemble. The first classifier was defined in a classical form as the softmax built in the structure of CNN. Different drop-out ratio values have been used. The other 8 units built on the basis of SVM and RF are supplied by the selected signals of the activation vector created by fully connected layer (fc7 containing 4096 elements). The selection process was arranged in 4 different ways: stepwise fit (SWF), neighborhood component analysis (NCA), reliefF (RelF) and Fisher discriminant analysis (FD).

Stepwise fit (SWF) selection [24] is based on adding or removing variables from the general set of features. After a new variable is added or removed, a test is made to check, if some variables from the set can be deleted without significant increasing the error of classification. The procedure terminates when the actual quality measure is maximized, or when the available improvement falls below some critical value.

Nearest neighbor analysis (NCA) method of selection [27] uses the KNN classifier. However the distances between vectors are subject to scaling. For N-dimensional feature vectors $\mathbf{x}$ the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined in the form

$$D(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^{N} w_l^2 \left| x_{il} - x_{jl} \right| \tag{4}$$

The parameter w_l is associated with the lth feature. The higher this value the more important the feature is.

In reliefF (RelF) method of selection [26] each feature vector $\mathbf{x}$ looks for the closest instances from each class in the data base. The closest same-class instance is called 'nearHit', and the closest different-class instance is called 'nearMiss'. Each ith component of feature vector $\mathbf{x}$ is associated with the weight w_i, which is subject to adaptation according to formula [12],[26]

$$w_i := w_i - \left(w_i - nearHit_i\right)^2 + \left(w_i - nearMiss_i\right)^2 \tag{5}$$

The weight representing particular feature decreases if it differs from that near feature of the same class more than nearby instances of the other class, and increases in the reverse case. After n iterations each weight is divided by n and represents the relevance vector. Features of their relevance values greater than the assumed threshold are selected as the most important.

In Fisher discriminant (FD) criterion [24], the importance of feature f is represented by the so called discrimination coefficient $S_{AB}(f)$. For two classes A and B it is defined as follows

$$S_{AB}(f) = \frac{\left| c_A(f) - c_B(f) \right|}{\sigma_A(f) + \sigma_B(f)} \tag{6}$$

The parameters c_A and c_B are the mean values of the feature f in the class A and B, respectively. The variables σ_A and σ_B represent the standard deviations determined for both classes. The larger the value of $S_{AB}(f)$, the better is separation ability of the feature f for these two classes. As it is seen the applied selection methods are relied on different principles and hence their results are highly independent.

Further increase of the independence of the members of ensemble has been achieved by applying three types of classifiers, which rely their decision on different principles of operation. The first one is the softmax built in the classical CNN structure. It is supplied by the set of activation signals with the randomly selected dropout ratio, serving as the applied selection. The second classifier is support vector machine and the third random forest of decision tress.

The support vector machine [23] works in high dimensional feature space obtained by the non-linear mapping of input vector $\mathbf{x}$ into a L-dimensional feature space (L>N) by using a kernel function $K(\mathbf{x},\mathbf{x}_i)$. The SVM of the Gaussian kernel was used in our application. The regularization constant C and Gaussian kernel width have been adjusted by repeating the learning experiments for the set of their predefined values and choosing the best one on the basis of the validation data set.

The Breiman random forest represents an ensemble of many decision tress for classification [25]. By constructing many different decision trees at training time it outputs the class being the mode of the classes output pointed by individual trees. The good generalization ability is obtained by applying randomness in selecting the learning data and using the limited set of decision variables chosen randomly in each node of the tree.

Both classifiers have very good reputation as the most accurate classification tools. To increase the independence of classifiers we have added small amount of noise of normal distribution to each learning sample. The structure of the used ensemble is presented in Fig. 6. This structure is trained 10 times on the randomly selected learning data (at the same subset of testing data). Testing results of each run are used in voting to elaborate final decision of ensemble.

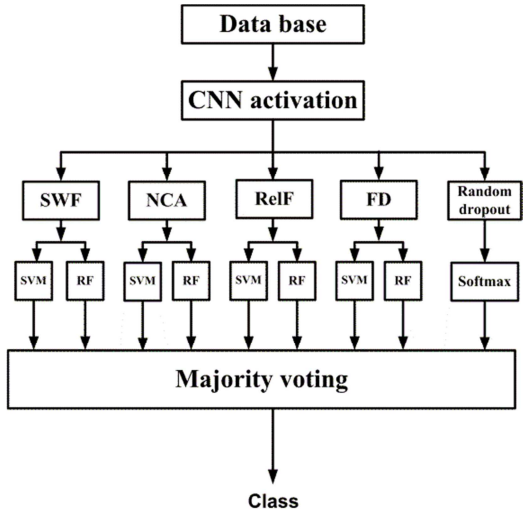

Fig. 6. The general structure of ensemble of classifiers.

The process of learning was performed by using 70% of available images. The learning phase of each unit used 90% randomly selected samples from this learning part, leaving 10% for validation. The remaining 30% of data were used only for testing purposes. The testing images were the same for all members of ensemble. Ten epochs were enough to train the whole circuit structure and the time of learning each CNN unit was around 2 minutes on PC with graphical processing unit.

V. Results of Experiments

The first experiments have been performed on ISIC 2017 data base in recognizing two classes: melanoma and non-melanoma cases. The first class contained 945 and the second 1543 samples (2488 samples together). The experiments of training and testing have been repeated many times and the results averaged and compared. Due to different contents of learning and testing data they are varying slightly. Table I depicts the statistics of individual results of classifiers obtained in 10 repetitions of experiments using ISIC 2017 data base. It is evident, that individual classifier results differ a lot. The most accurate is SVM in combination with stepwise fit selection. The least accurate is softmax.

TABLE I.
THE STATISTICAL RESULTS OF ACCURACY IN MELANOMA RECOGNITION OF 9 INDIVIDUAL CLASSIFIERS IN 10 RUNS OF EXPERIMENTS (EXPRESSED IN %).

	Mean	Std	Median	Max
Softmax	64.76	2.62	66.19	67.40
SWF+SVM	76.37	1.75	75.85	79.27
SWF+RF	73.64	1.31	73.34	76.25
NCA+SVM	69.67	1.89	69.81	72.03
NCA+RF	72.99	1.02	72.93	75.05
RelF+SVM	74.26	0.91	74.04	75.85
RelF+RF	73.58	1.13	73.94	74.64
FD+SVM	74.72	2.03	74.74	77.86
FD+RF	73.58	1.13	73.94	74.64
Average	72.62	1.53	72.75	74.78

In next experiments we have investigated the ensemble composed of the same classifiers and integrated by majority voting. Different arrangements of parameters of individual classifiers have been tried. The best setting was used in final experiments. The integration process has included the results of 9 units tested on the same contents of testing data. The average accuracy results of ensemble testing have been changing from 79% to 81% in experiments at different (random) split of data into learning and testing parts. The detailed statistical results (accuracy – ACC, sensitivity –SENS, specificity – SPEC, precision - PREC and area under ROC curve - AUC) of 10 runs are depicted in table II. They represent the average of 10 repetitions of experiments.

TABLE II.
THE STATISTICAL RESULTS OF MELANOMA RECOGNITION OF THE ENSEMBLE COMPOSED OF 9 UNITS

ACC	SENS	SPEC	PREC(class1)	PREC(class2)	AUC
80.86%	66.14%	89.61%	79.62%	81.18%	0.853

The plot of receiver operating characteristic corresponding to one case is presented in Fig. 6. The area under ROC curve is equal AUC=0.853.

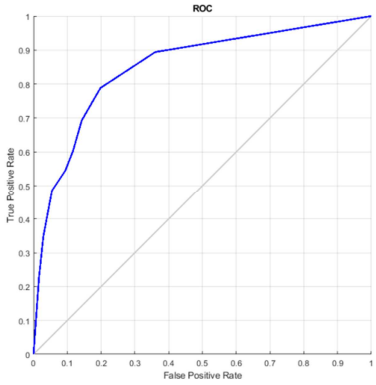

Fig. 7. ROC curve of the ensemble of classifiers.

The results of ensemble are much better than that of the individual classifiers. It is well seen on the example of statistics corresponding to results of accuracy of individual classifiers involved in ensemble (in 10 runs of experiments). The average accuracy of the same classifiers not integrated in ensemble were as follows: mean =72.62%, std=1.53%, median=72.75%, max =79.27%.

Fig. 8 depicts comparison of the accuracy of individual classifier and the ensemble in a graphical way. Evidently the ensemble has delivered the best result.

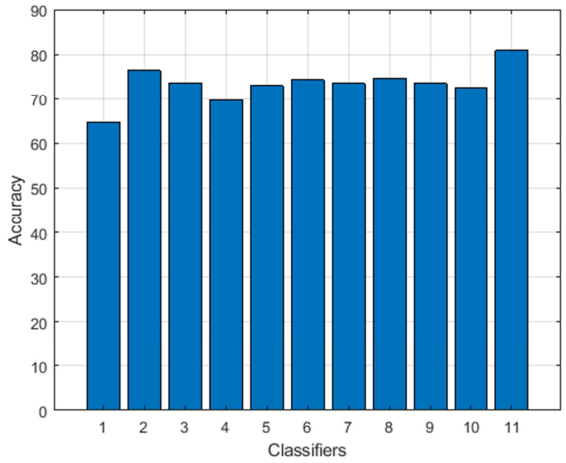

Fig. 8. The comparison of accuracy of individual classifiers and ensemble: 1- Softmax, 2 -SWF+SVM, 3 – SWF+RF, 4 –NCA+SVM, 5 – NCA+RF, 6 – RelF+SVM, 7 – RelF+RF, 8 – FD+SVM, 9 – FD+RF, 10 – Average of individual classifiers, 11- Ensemble.

The next experiments have been done using another data base collected in Warsaw Memorial Cancer Center and Institute of Oncology, Department of Soft Tissue/Bone Sarcoma and Melanoma. This time the images were cropped by the medical experts to the proper regions presenting the lesions.

The ensemble was created once again by 9 CNN classifiers, integrated by majority voting. Each time the verdict of ensemble was compared to the mean of results of all individual members of ensemble. The accuracy, sensitivity and specificity have been compared.

The results of individual runs change, because of small size of the available samples in the data base. To obtain statistically reliable results, we have performed 10 runs of experiments with randomly selected sets of learning data by keeping the testing subset the same in each run.

Table III presents the average accuracy of melanoma versus non-melanoma recognition in 10 repetitions of experiments at application of this small data base (only testing results are presented in the table). These are the results of ensemble and the mean of the same classifiers working independently.

The average sensitivity in melanoma recognition obtained by ensemble in 10 repetitions (learning/testing sessions) of experiments was equal 97.69%, while the specificity was 89.07%. In the case of individual classifiers (not arranged in the form of ensemble) the average sensitivity was equal 94.30% and specificity 85.76%. The precision of melanoma and non-melanoma class recognition of the ensemble was 89.3% and 91.3%, respectively. These values for average of individual classifiers were 86.7% and 87.3%, respectively. The area under ROC was AUC=0.8424 for ensemble and AUC=0.8227 for the mean of individual classifiers. There is an evident superiority of ensemble over the individual operation of classifiers. All quality measures (accuracy, sensitivity and specificity) have been significantly increased with respect to the average of individual classifiers.

TABLE III.
THE AVERAGE ACCURACY OF MELANOMA VS NON-MELANOMA OBTAINED IN 10 RUNS OF THE ENSEMBLE OF CNN CLASSIFIERS.

Repeating sessions	**Applied approach**	**Accuracy [%]**
1	Mean of classifiers	90.83±2.46
	Ensemble of classifiers	94.44
2	Mean of classifiers	88.58±3.71
	Ensemble of classifiers	90.74
3	Mean of classifiers	89.72±4.28
	Ensemble of classifiers	92.61
4	Mean of classifiers	88.79±5.31
	Ensemble of classifiers	94.48
5	Mean of classifiers	90.37±6.12
	Ensemble of classifiers	96.30
6	Mean of classifiers	94.63±2.99
	Ensemble of classifiers	98.15
7	Mean of classifiers	89.35±4.45
	Ensemble of classifiers	92.61
8	Mean of classifiers	91.20±4.28
	Ensemble of classifiers	94.39
9	Mean of classifiers	91.11±3.15
	Ensemble of classifiers	92.59
10	Mean of classifiers	88.05±3.70
	Ensemble of classifiers	92.65
Mean	**Mean in all sessions**	**90.26±1.88**
	Ensemble of classifiers in all sessions	**93.89±2.14**

The quality measures obtained in deep learning approach are much better than traditional results obtained at application of specially prepared features. For example the enlarged set of features (textural, colorimetric, Kolmogorov-Smirnov, percolation and maximum subregions descriptors) combined with SVM and random forest classifiers [7] applied to the same data base of Warsaw Memorial Cancer Center has resulted in accuracy changing from 83.1% to 92.76% depending on the applied classification system. The maximum sensitivity and specificity obtained in the best case were equal 95.2% and 92.4%, respectively.

VI. CONCLUSIONS

The paper has proposed the ensemble of CNN based classifiers in recognition of images representing melanoma versus non-melanoma. We have used the activations generated by the locally connected convolutional layers of CNN to create the set of diagnostic features. These features were subject to selection by using different procedures. The selected features represent the input attributes to three different classifiers: SVM, RF and softmax, integrated into ensemble.

Thanks to different form of classifiers and selection methods high independence of classification units forming the ensemble has been obtained. Integrating the results of its members has led to significant quality improvement of the class recognition system. The results of ensemble verdicts have been compared to the average results of individual classifiers forming this ensemble. The quality measures of ensemble were much better than the mean of the individual classifiers, proving the advantage of the proposed solution.

The important conclusion from these experiments is that application of many individual classifiers based on the activation signals of CNN and arranged in the form of ensemble, provides better results in mammogram recognition. The cooperation of many units integrated into one final decision leads to the significant improvement of the whole system operation and affects simultaneously accuracy, sensitivity, precision and specificity values.

Future research in this field will be directed toward the following tasks:

- Increasing the number of classifiers by including different architectures of CNN (among other [28]) and coupling them in an ensemble. More types of ensemble members will provide new points of view on recognition problem, hopefully increasing the final accuracy of the system.
- Additional experiments should be done to check the performance of the system on other data bases of melanoma and implement this procedure in large scale in medical practice for supporting the melanoma diagnosis.

REFERENCES

[1] Melanoma Skin Cancer, American Cancer Society website. Available at http://www.cancer.org/acs/groups/cid/documents/webcontent/003120-pdf.

[2] S. E. Yagerman and A. Marghoob A, "The ABCDs and beyond", The Skin Cancer Foundation Journal, vol. 94, pp. 31-61, 2013.

[3] S. M. Goldsmith, "A unifying approach to the clinical diagnosis of melanoma including "D" for "Dark" in the ABCDE criteria", Dermatol Pract Concept, vol. 4, pp. 75–78, 2014.

[4] M. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. Aslandogan, W. Stoecker and R. Moss, "A methodological approach to the classification of dermoscopy images", Computerized Medical Imaging and Graphics, vol. 32, pp. 362-373, 2007.

[5] 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, pp. 1-13, 2012.

[6] C. Barata, M. Ruela, M. Francisco. T. Mendonça and J. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features", IEEE Systems Journal, vol. 8, pp. 965-979, 2013.

[7] M. Kruk, B. Świderski, S. Osowski, J., Kurek, M. Słowińska and I. Walecka, "Melanoma recognition using extended set of descriptors and classifiers", Eurasip Journal on Image and Video Processing, vol. 43, pp. 1-10, 2015. DOI 10.1186/s13640-015-0099-9.

[8] H. Ganster, A. Pinz, E. Wildling, M. Binder and H. Kittler, "Automated melanoma recognition", IEEE Trans. on Medical Imaging, vol. 20, pp. 233–239, 2001.

[9] P. Sabouri, H. H. Gholam, T. Larsson and J. Collins., "A cascade classifier for diagnosis of melanoma in clinical images", Engineering in Medicine and Biology Society (EMBC) 36th Annual Intern. Conf. of the IEEE, Chicago, 2014.

[10] 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.

[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] Matlab user manual. Natick (USA): MathWorks: 2019.

[13] I. Goodfellow, Y. Bengio and A. Courville.' "Deep learning". New York: MIT Press; 2016.

[14] A. Krizhevsky, I. Sutskever and G. Hinton, "Imagenet classification with deep convolutional neural networks, NIPS, 2012.

[15] https://isic-archive.com/.

[16] https://challenge.kitware.com/#challenge/n/ISIC_2017%3A_Skin_Lesion_Analysis_Towards_Melanoma_Detection.

[17] https://github.com/BVLC/caffe/tree/master/models/bvlc_googlenet.

[18] 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.

[19] A. A. Milton, "Automated skin lesion classification using ensemble of deep neural networks in ISIC 2018", skin lesion analysis toward melanoma detection challenge, 2018, http://arXiv.org/abs/1405.0312

[20] 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", 2017, http://arXiv.org.abs/1610.04662

[21] 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 4/5, 2017.

[22] T. Shane, "Applied Computer Science"(2nd ed.), Springer, Berlin, 2016

[23] B. Schölkopf and A. Smola, "Learning with kernels", MIT Press, Cambridge MA, 2002.

[24] P. N. Tan, M. Steinbach and V. Kumar V., "Introduction to data mining", Pearson Education Inc., Boston, 2006.

[25] L. Breiman, "Random forests", Machine Learning, vol. 45, No 11, pp. 5–32, 2001.

[26] R. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF". *Machine Learning*, vol. 53, pp. 23-69, 2003.

[27] W. Yang, K. Wang and W. Zuo, "Neighborhood component feature selection for high dimensional data", Journal of Computers, vol. 7, No 1, pp. 161-168, 2012.

[28] C. Szegedy, S. Ioffe, V. Vanhoucke, "Inveption-v4, Inception-ResNet and the impact of residual connections on learning". arXiv:1602.07261v2, 2016.