Pre-trained CNN based deep features with hand-crafted features and patient data for skin lesion classification

Sule Yildirim-Yayilgan[1], Blend Arifaj[2], Masoomeh Rahimpour[3], Jon Yngve Hardeberg[1], and Lule Ahmedi[2]

[1] Faculty of Information Technology and Electrical Engineering, NTNU – Norwegian University of Science and Technology, Gjøvik, Norway
[2] Faculty of Electrical and Computer Engineering, University of Prishtina, Prishtina, Republic of Kosova
[3] Catholic University of Leuven, Leuven, Belgium
sule.yildirim@ntnu.no

Abstract. Skin cancer is a major public health problem, with millions newly diagnosed cases each year. Melanoma is the deadliest form of skin cancer, responsible for the most over 6500 deaths each year in the US, and the rates have been rising rapidly over years. Because of this, a lot of research is being done in automated image-based systems for skin lesion classification. In our paper we propose an automated melanoma and seborrheic keratosis recognition system, which is based on pre-trained deep network combined with structural features. We compare using different pre-trained deep networks, analyze the impact of using patient data in our approach, and evaluate our system performance with different datasets. Our results shown us that patient data has impact on characteristic curve metric value with around 2-6% and different algorithm in final classification layer has impact with around 1-4%.

Keywords: Deep Networks · CNN · Handcrafted features · Skin Lesion Classification and Segmentation · Image Processing.

1 Introduction

Malignant melanoma are among the most rapidly increasing types of cancer in the world with a considerable mortality rate. The American Cancer Society predicted more than 100.350 melanoma cases in 2020, of which more than 6.850 persons are expected to lose their lives [1]. However, early detection of melanoma leads to a cure rate of over 90% in low risk melanoma patients, and the use of dermoscopy images [6] plays an important role in this process. Dermoscopy images are mainly used to evaluate pigmented skin lesions in order to distinguish malignant skin lesions from benign ones such as melanocytic nevus and seborrhoeic keratosis. However, due to the visual similarity of different skin lesions, melanoma diagnosis using human vision alone can be subjective and inaccurate even among experienced dermatologists. Lesions' varying size, shape and

fuzzy boundaries, different skin colors, presence of different artifacts including hair and bubbles are some of the reasons that make the detection process more challenging.

In this paper, we propose two architectures using two different categories of features extracted from images of skin lesions, namely handcrafted and deep features. We compare the performance of these two kinds of features separately and also consider a combined use of these features to provide a single classification decision for melanoma diagnosis.

The remaining paper is organized as follows. In Section 2 a survey of state-of-the-art on melanoma classification is presented. Section 3 describes the theory behind the proposed method including deep networks and dermoscopic features. The experimental results of the proposed method on melanoma classification and a comparative study are presented in Section 4. Finally, the discussion and conclusion remarks are given in Section 5.

2 Related works

The success of melanoma detection task highly depends on the features extracted from the lesion. There are different clinical approaches that are commonly used to extract diagnostic features from dermoscopy images such as ABCD(E) rule, pattern analysis, seven-point checklist and Menzies method [36]. The ABCD(E) rule is based on asymmetry, border, colour, diameter, and evolution (or elevation) features [34]. The pattern analysis approach facilitates melanoma diagnosis by determining the presence of specific patterns visible in dermoscopic images. The features employed in pattern analysis approach refer both to the chromatic characteristics and to the shape and/or texture of the lesion. These include Atypical Pigment Network, Blue-whitish Veil, Atypical Vascular Pattern, Irregular Streaks, Irregular Pigmentation, Irregular Dots/Globules, and Regression Structures [19]. The seven-point checklist provides a simplification of the standard pattern analysis and compared to ABCD(E), allows less experienced observers to achieve higher diagnostic accuracy values. The Menzies approach facilitates to identify the colour patterns and asymmetry within the lesion. Computational methods based on the Menzies' criteria have been proposed to analyze the presence of six basic colour classes (white, red, light brown, dark brown, blue–grey, and black) for dermoscopic images [33].

The features used in automated dermoscopy image analysis mostly have been inspired by the clinical methods which have been mentioned above. Table 1 shows a distribution of feature categories known as hand-crafted features commonly employed in the literature. We categorize these features in four main classes including structural (shape), geometrical (spatial), chromatic (colour) and texture features. The interested reader is referred to [33] for detailed classification.

Recently, Convolutional Neural Networks (CNNs) have shown an impressive performance in image classification tasks. Different features detected at the different convolutional layers enable the network to handle large variations in the dataset which may result in higher diagnostic accuracy and sensitivity in classi-

Table 1. Hand-crafted categorizes in automated melanoma classification

Feature	structural	geometric	chromatic	textural
Sadeghi et al [35]	√	√	√	√
Møllersen et al [31]			Divergence-based	
Ballerini et al [7]			RGB, HSV, CIE	GCM
Tomas et al [26] [25]				LBP and RSurf
Codella et al [14]	Edge Histogram		Color Histogram	MSLBP
Ma et al [24]	√		√	√
Barata et al [10]			√	√
Garnavi et al [19]	√			√
Zhou et al [44]	Fourier transformation			
Damian et al [16]	Fourier transform amplituden	asymmetry, circularity	color distribution	

fication. In comparison with fully connected networks, CNNs have much fewer connections and parameters and hence they are easier to train. They build the property of translation invariance of image statistics by replicating the same neurons. They model the low-level features of images locally by enabling the local connectivities, and make the high-level features of image coarser as they are in nature [40] using repeated pooling/sub-sampling layers.

In 2014, Simonyan et al [37] investigated the effect of depth component in CNNs by introducing VGG network which achieved 7.3% error rate in ILSVRC (ImageNet Large Scale Visual Recognition) competition. This network is characterized by its simplicity which has increased the depth of network by adding more convolutional layers. In comparison to the previous networks, VGG (with 16 or 19 layers) models more non-linearity and has less parameters by employing small convolutional filters (3*3) in each layer. Szegedy et al [40] presented two different concepts in CNNs including a new level of organization in the form of the "Inception module" and increased network depth. They proposed a deeper network (22-layer) with computational efficiency in the case of the GoogLeNet.

However, there was another challenge by increasing the depth of network; it has been depicted that adding extra layers beyond the certain layers does not help obtain promising results and may results in higher training and validation errors. In 2015, He et al [20] presented a new architecture called Deep Residual Network (ResNet) which won ILSVRC 2015 with an incredible error rate of 3.57%. ResNet consists of a number of residual blocks with each block comprising of several convolution layers, batch normalization layers and ReLU layers. The residual block enables to bypass a few convolution layers at a time [41].

Most of these architectures (VGGNet, GoogLeNet, ResNet) are available as the pre-trained models initially trained on approximately one million natural images from the ImageNet dataset [17]. Such networks are widely employed for melanoma detection. Codella et al [13] used a pre-trained CNNs named Caffe to extract the features for skin image classification and proved that deep features outperform traditional hand-crafted features. Majtner et al [27] used AlexNet as

pre-trained deep network for features extraction and proved that deep feature has potential on image classification. Devassy et al [29] replaced the complex handcrafted features with standard and deep feature, extracted from ResNet. This combination outperformed state of the art results. Zhao et al [43] showed that CNN has a great ability for localizing the objects. They proposed a technique called Class Activation Mapping (CAM) by using global average pooling (GAP) in CNNs. A class activation map for a particular category indicates the discriminative image regions used by the CNN to identify the category.

Despite all these algorithms, there is still room for automated algorithm that shows reasonable results. This motivate us to develop a new approach that combines pre-trained deep networks and structural features to detect skin lesions.

3 Methodology

We proposed a structure that combines state-of-the-art developments of deep neural networks and machine learning techniques to automated analysis of melanoma images. The main pipeline of our proposed framework for melanoma classification is depicted in Fig 1. Similar to most common approaches employed in computational systems for skin lesion diagnosis, it includes six main steps 1) Preprocessing, 2) Data Augmentation 3) Segmentation 4) Feature Extraction 5) Feature Selection and 6) Classification. In the following, more details for each step are presented.

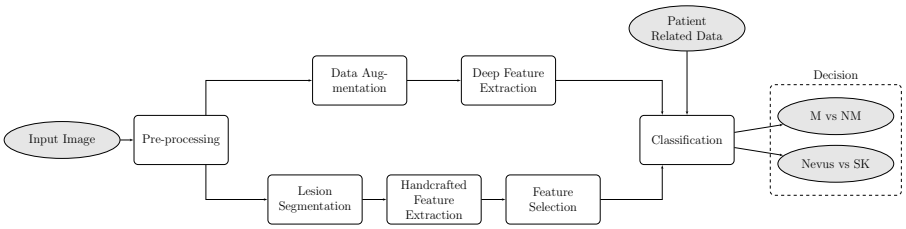

Fig. 1. Pipeline of automated melanoma image classification

3.1 Preprocessing

Dermoscopy images are often obtained with different imaging devices under varying acquisition conditions. Changes in the illumination of image acquisition devices adversely affect the color of images and reduce the performance of diagnostic systems. Hence, it is a crucial task to normalize the color and illuminance balance of images by applying color constancy methods. Different research groups have proposed color normalization strategies to deal with dermoscopy images. Gray_world, Max_RGB, Shades of Gray and General Gray World methods are among these methods that are fast, easy to implement and require the tuning

of few parameters [8]. In this study, we applied the Shades of Gray method. Interested readers is referred to [12] for a detailed description of different methods.

3.2 Segmentation

First, automatic border detection is performed to separate the lesion from the background skin. To do this job we applied deep Residual U-Net [42]. The main benefits from this model are: residual units ease training of deep networks and this model allow us to design a deep network with fewer parameters however and better performance [42]. To train this network we applied 2000 pairs of lesion-mask and for validation we applied 200 pairs lesion-mask. Those pairs lesion-mask are provided from ISIC 2017 challenge [15].

3.3 Data Augmentation

It is well known that data augmentation notably improves the performance of deep neural networks. Particularly if the amount of training data is limited, augmentation would improve the robustness of the model [23]. There are different kinds of geometrical transformations that can be applied on the images before providing them as input to the CNN to generate new versions of the images such as applying random transformations: cropping, flipping, etc. In the following, the process of data augmentation employed in this study has been described:

1. Considering four rotation angles (45, 90, 135 and 180), we generate four rotated images per original image.
2. Finding the largest inner rectangle, we crop each image to ensure that all pixels belong to the original image.
3. Finally, we perform square crops to resize each image of 224*224, since the CNN network requires square input images.

Considering the orientation changes and crops preformed, we generate eight versions of each image which, then will be provided to the trained network [18].

3.4 Feature Extraction

The performance of automated approaches for skin lesion detection greatly depends on the choice of meaningful descriptive features derived from the dermoscopic images. In this study, we have considered three types of features which are useful in automated skin lesion detection for improving the overall classification performance.

Deep Features As our first attempt for feature extraction, we employed a model based on the VGG network. According to the CNN models and based on our observations in the experiments using VGG and VGG-Deep, we employed ResNet as the next learning architecture to extract the deep features. Our CNNs were fine-tuned with the training samples from the initially pre-trained model to perform classification on the ILSVRC 2015 data [20]. In the last layer of ResNet,

there are 2048 features related to high-level concepts of the input image which are ready to be fed to the classification layer.

Handcrafted Features Feature extraction aims to represent a set of descriptors to separate each image into different classes by feeding them to the classifier. As mentioned is previous section, in order to extract the deep features, all the images need to be resized to a fixed size of 224×224 pixels. As a result, some important features related to the size and shape of object might be eliminated. To address this problem, some valuable features called handcrafted features have been extracted and added to previous deep features to strengthen our feature set. Once the image has been segmented into the object and background as stated in Section 3.2, the important handcrafted features must be extracted from the object. Three main categories of handcrafted features have been employed in this study including shape features, color features, and texture features.

Shape features: Shape features provide important information about the lesion that can be deterministic for lesion's classification [39]. To shape feature we used lesion's mask, provided as output from lesion segmentation step in Fig. 1.

Color features: Color feature descriptors delineate the colour distribution of images. In order to quantify the color features, statistics features over the channels of different color spaces were calculated where the color spaces consist of RGB, HSV, and L*a*b*. The most used descriptors are mean, standard deviation, skewness, kurtosis, and median. *Texture features*: Texture features provide important information about the spatial disposition of the grey levels and the relationship with their neighbourhoods. For texture feature we have applied the following methods: SURF [11], SIFT [22], ORB [3], and LBP [32].

Patient Related Data Dermatologists do not base their diagnosis solely on the analysis of the skin lesion. Patient-related context is also of relevance in making a diagnosis. There are a lot of factors that may affect the final decision such as age, gender, skin type, personal disease history and part of body [4].

3.5 Feature Selection

Feature selection is an important preprocessing step in many machine-learning tasks. The purpose is to reduce the dimensionality of the feature space by eliminating redundant, irrelevant or noisy features. From the classification perspective, there are numerous potential benefits associated with feature selection: (i) reduced feature extraction time and storage requirements, (ii) reduced classifier complexity, (iii) increased prediction accuracy, (iv) reduced training and testing times, and (v) enhanced data understanding and visualization. To do this step we have employed PCA[5], as feature selector, and select 250 features.

3.6 Classification

This block aims to perform classification among three types of skin lesions. This includes two independent binary classifications. The first classifier is used for the categorization of dermoscopic data into (a) melanoma and (b) nevus

and seborrheic keratosis. The second classifier uses the extracted features for categorizing the dermoscopic images into (a) seborrheic keratosis and (b) nevus and melanoma. We have exploited different kinds of classifiers including SVM, Linear SVM (LSVM), RUSBoost Classifier.

4 Experimental Results

Dataset and Evaluation metrics: The proposed structure is trained on the dermoscopic image sets provided from the International Skin Imaging Collaboration (ISIC) 2017 Challenge "Skin Lesion Analysis towards Melanoma Detection" – Part 3: Lesion Classification and the Seven-Point Checklist dataset [21]. The created model is evaluated on two datasets: ISIC 2017 Test Dataset [15] and on PH-2 Dataset [30] (patient data are not provided on this dataset).

Results: The first part of the model consisted of the VGG 16 layers up to the last convolutional layer, but without the last max pooling layer and the fully-connected layers. We initialized our network with the weights of a VGG 16 model pretrained on the ImageNet dataset. Also, we did the same for ResNet model. During training, the inputs are pre-processed by zero mean unit standard deviation normalization. We evaluate the proposed structure as follows:

1. Compare the VGG-16 network to the deeper ResNet-101.
2. Compare the performance of proposed appoaches
3. Compare the performance of different classifier as a final decision layer.
4. Attempt to use the patient data (age and sex) on classification.

The classification results are evaluated under metrics:

- Characteristic curve (ROC) - measure of how well a parameter can distinguish between two diagnostic groups.
- Sensitivity (SE) - the fraction of true positives that are correctly identified
- Specificity (SP) - the fraction of true negatives that are correctly identified
- Accuracy (ACC) - the number of correct predictions divided by the total number of predictions

Experiments are made on ISIC 2017 Test Dataset, that provide us 600 images (117 melanomas, 90 seborrhoeic-keratoses and 393 nevus) and to strengthen our conclusions we test our approach on PH2 dataset, that provided us 200 images (140 nevus and 60 melanomas).

To make comparison between deep networks, ResNet and VGGNet, we run experiments on three classifiers with every possible combination between deep features, handcrafted features (HF) and patient data (PD). Results are summarized in Table 2, in columns we have features that are used with ResNet respectively VGGNet. We can observe that the results with ResNet, as deep feature extractor, outperforms the results with VGGNet, as feature extractor, in every possible combination. In classifier A, ResNet as deep feature extractor, outperform the VGGNet in every combination. Best results achieved in classifier A from ResNet features is 67.45% and for VGGNet features is only 58.12%. In

classifier B, also, ResNet features outperform VGGNet features in every combination. The best result archieven from ResNet features, in classifier B, is 75.65% and from VGGNet features is 71.05%. In PH2-dataset results are more closer. The difference between ResNet feature and VGGNet feature is approx 2%. Also, in this dataset ResNet achieved better result than VGGNet. The best result achieved from ResNet features is 86.25% and from VGGNet features is 84.38%. Another thing that we can observe in Table 2, is that it is clear that patient data increase ROC value from 1% to 2% in classifier A and from 4% to 6% in classifier B. Also, with patient data we receive an highest Specificity in classifier A with 64.10% and in classifier B with 84.44%. Also, from Table 2 we can clearly observe that, In classifier A if we use only handcrafted feature the best results that we can achieve, based on ROC metric, is 58.10%. If we use deep features, result is improve to 64.00% and when we combine those features we improve result to 67.45%. For classifier B when we use only handcrafted and deep feature separately we achieve results 62.94% respectively 71.99%, but result is improve, when we combine those features to 75.65%. In PH2 dataset we can figure out that combination of deep and handcrafted feature increase result to 86.25%, in comparison when we use those feature separately, 69.94% respectively 71.81%.

To find out which algorithm perform better results in our approach, we have run experiments with three different algorithms: SVM, Linear SVM (LSVM) and RUSBoosting. In Table 3 we have summarized the best results for each algorithm. From results we observe that RUSBoost perform better results in all three classifier under ROC metric. In classifier A the ROC value is 4.46% higher than SVM and 2.26% higher than LSVM. In classifier B the ROC value is 4.25% higher than SVM and 3.50% higher than LSVM. In PH2 dataset the ROC value is 2% higher than SVM and 5 % higher than LSVM. In Table 4, we have summarized the best results from our approaches and the other results, include state of the art, to the best of our knowledge those are the only one who did the validation on PH2 dataset. From those results, we observe that our approach performs better than others under accuracy metric (90.00 %), and under ROC metric (86.50%) we are the just behind state of the art (89.50 %). Also, our approach achieve sensitivity 92.50% that outperform state of the art with 4.50%, but our specificity is for about 3% lower than state of the art.

5 Conclusions

In this study, we did not aim to develop a new deep network, but we tried to use the deep features extracted from available architectures. Considering the restrictions in ISIC 2017 dataset including the limited size and the problem of unbalanced data, we feed these features to a RUSboost classifier and we apply data augmentation. From the results discussed above, we can come to some conclusions. First conclusion is that: combination of deep with handcrafted features approach improve the results in comparison to approach with only deep or handcrafted features. Second conclusion from the results is that: feature extracted from ResNet are more representative than those extracted from VGGNet.

Table 2. Comparing ResNet vs VGGNet

		ResNet				VGGNet			
Challenge	Metric	-	PD	HF	HF+PD	-	PD	HF	HF+PD
ISIC2017	ROC	63.84%	64.00%	65.15%	**67.45%**	55.12%	58.08%	57.37%	58.12%
Classifica	SE	76.39%	93.78%	67.90%	70.80%	93.99%	69.15%	93.37%	59.83%
tion A	SP	51.28%	34.20%	62.39%	64.10%	16.24%	47.00%	21.37%	56.41%
	ACC	71.50%	82.17%	66.83%	69.50%	78.83%	64.83%	79.33%	59.17%
ISIC2017	ROC	68.10%	71.99%	69.93%	**75.65%**	65.16%	68.79%	66.37%	71.05%
Classifica	SE	91.76%	91.76%	70.98%	65.68%	92.55%	63.13%	92.74%	57.64%
tion B	SP	44.44%	52.22%	68.89%	84.44%	37.78%	74.44%	40.00%	84.44%
	ACC	84.67%	85.83%	70.67%	68.50%	84.33%	64.83%	84.83%	61.67%
	ROC	71.81%	-	**86.25%**	-	69.38%	-	84.38%	-
PH2 dat	SE	90.63%	-	92.50%	-	88.75%	-	93.75%	-
aset	SP	53.00%	-	80.00%	-	50.00%	-	75.00%	-
	ACC	83.00%	-	90.00%	-	81.00%	-	90.00%	-

Table 3. Compare the performance of different classifiers as a final decision layer

Challenge	Algorithm	ROC	SE	SP	ACC
ISIC 2017 Classification A	SVM	65.04%	95.03%	35.04%	83.33%
	LSVM	67.24%	90.89%	43.59%	81.67%
	RUSBoost	67.45%	70.80%	64.10%	69.50%
ISIC 2017 Classification B	SVM	71.40%	93.92%	48.89%	87.17%
	LSVM	72.15%	90.98%	53.33%	85.33%
	RUSBoost	75.65%	65.68%	84.44%	68.50%
PH2 Dataset	SVM	84.38%	93.75%	75.00%	90.00%
	LSVM	81.56%	90.63%	72.50%	87.00%
	RUSBoost	86.25%	92.50%	80.00%	90.00%

Table 4. Comparing our approach best results with others in PH2 dataset

Paper	ROC	SE	SP	ACC
Abbas [2]	**89.50%**	88.00%	**91.00%**	89.00%
Proposed	86.25%	92.50%	80.00%	**90.00%**
Barata [9]	86.00%	85.00%	87.00%	87.00%
Marques [28]	85.00%	**94.00%**	77.00%	79.00%
Situ [38]	85.00%	86.00%	85.00%	85.00%

The results showed us that features from ResNet when they are combined with handcrafted feature and patient data, if they are provided, achieve the best results in compare to other combinations or any combination of features from VGGNet with handcrafted feature and patient data. Also, this combination of features from ResNet with those handcrafted outperform state of the art, on PH-2 dataset, in accuracy and sensitivity. Next conclusion is relevance of patient data. From the results is shown that when patient data are used results are improved for 1 to 6 %. Also, RusBoost as a final classifier performs better results in comparison to SVM or Linear SVM. The difference between these classifiers varies from 1 to 4%.

6 Acknowledgement

This research was supported in part by the grants from the IQ-MED (Image Quality enhancement in MEDical diagnosis, monitoring and treatment) project, funded by the Research Council of Norway; and ERASMUS+ funding.

References

1. Key statistics for melanoma skin cancer. https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html, accessed: 2020-05-01
2. Abbas, Q., Emre Celebi, M., Garcia, I.F., Ahmad, W.: Melanoma recognition framework based on expert definition of abcd for dermoscopic images. Skin Research and Technology **19**(1), e93–e102 (2013)
3. Abdulmajeed, M., Seyfi, L.: Object recognition system based on oriented fast and rotated brief (12 2018)
4. Alcón, J.F., Ciuhu, C., Ten Kate, W., Heinrich, A., Uzunbajakava, N., Krekels, G., Siem, D., De Haan, G.: Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis. IEEE journal of selected topics in signal processing **3**(1), 14–25 (2009)
5. Anna Karen Garate-Escamila, Amir Hajjam El Hassani, E.A.: Classification models for heart disease prediction using feature selection and pca. In: Informatics in Medicine Unlocked. Elsevier (2020)
6. Argenziano, G., Soyer, H., De Giorgi, V., Piccolo, D., Carli, P., Delfino, M., et al.: Dermoscopy: a tutorial. EDRA, Medical Publishing & New Media **16** (2002)
7. Ballerini, L., Fisher, R.B., Aldridge, B., Rees, J.: A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions. In: Color Medical Image Analysis, pp. 63–86. Springer (2013)
8. Barata, C., Celebi, M.E., Marques, J.S.: Improving dermoscopy image classification using color constancy. IEEE journal of biomedical and health informatics **19**(3), 1146–1152 (2015)
9. Barata, C., Marques, J., Rozeira, J.: Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model. In: "Advances in Visual Computing". pp. 40–49. Springer Berlin Heidelberg (2013)
10. Barata, C., Ruela, M., Francisco, M., Mendonça, T., Marques, J.S.: Two systems for the detection of melanomas in dermoscopy images using texture and color features. IEEE Systems Journal **8**(3), 965–979 (2014)

11. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer Vision and Image Understanding **110**, 346–359 (06 2008)
12. Cherepkova, O., Hardeberg, J.Y.: Enhancing dermoscopy images to improve melanoma detection. In: 2018 Colour and Visual Computing Symposium (CVCS). pp. 1–6 (2018)
13. Codella, N., Cai, J., Abedini, M., et al.: Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images. In: International Workshop on Machine Learning in Medical Imaging. pp. 118–126. Springer (2015)
14. Codella, N., Nguyen, Q.B., Pankanti, et al.: Deep learning ensembles for melanoma recognition in dermoscopy images. arXiv preprint arXiv:1610.04662 (2016)
15. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1710.05006 (2017)
16. Damian, F., Moldovanu, S., Dey, N., Ashour, A.S., Moraru, L.: Feature selection of non-dermoscopic skin lesion images for nevus and melanoma classification (04 2020). https://doi.org/10.3390/computation8020041
17. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database pp. 248–255 (2009)
18. Díaz, I.G.: Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions. arXiv preprint arXiv:1703.01976 (2017)
19. Garnavi, R., Aldeen, M., Bailey, J.: Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis. IEEE Transactions on Information Technology in Biomedicine **16**(6), 1239–1252 (2012)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
21. Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G.: Seven-point checklist and skin lesion classification using multitask multimodal neural nets. IEEE Journal of Biomedical and Health Informatics **23**(2), 538–546 (mar 2019)
22. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. vol. 2, pp. II–506 (05 2004)
23. Kumar, A., Kim, J., Lyndon, D., Fulham, M., Feng, D.: An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE journal of biomedical and health informatics **21**(1), 31–40 (2017)
24. Ma, L., Staunton, R.C.: Analysis of the contour structural irregularity of skin lesions using wavelet decomposition. Pattern recognition **46**(1), 98–106 (2013)
25. Majtner, T., Yildirim-Yayilgan, S., Hardeberg, J.Y.: Combining deep learning and hand-crafted features for skin lesion classification. In: Image Processing Theory Tools and Applications, 2016 6th International Conference on. pp. 1–6. IEEE
26. Majtner, T., Yildirim-Yayilgan, S., Hardeberg, J.Y.: Efficient melanoma detection using texture-based rsurf features. In: International Conference Image Analysis and Recognition. pp. 30–37. Springer (2016)
27. Majtner, T., Yildirim Yayilgan, S., Hardeberg, J.Y.: Optimised deep learning features for improved melanoma detection. Multimedia Tools and Applications **78**, 11883–11903 (05 2019)
28. Marques, J.S., Barata, C., Mendonça, T.: On the role of texture and color in the classification of dermoscopy images. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 4402–4405 (2012)

29. Melit Devassy, B., Yildirim Yayilgan, S., Hardeberg, J.Y.: The impact of replacing complex hand-crafted features with standard features for melanoma classification using both hand-crafted and deep features. Advances in Intelligent Systems and Computing (2018)
30. Mendonça, T., Ferreira, P.M., Marques, J.S., Marcal, A.R.S., Rozeira, J.: Ph2 - a dermoscopic image database for research and benchmarking. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 5437–5440 (2013)
31. Møllersen, K., Hardeberg, J.Y., Godtliebsen, F.: Divergence-based colour features for melanoma detection. In: Colour and Visual Computing Symposium (CVCS), 2015. pp. 1–6. IEEE (2015)
32. Ojala, T., et al.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: Proceedings of 12th International Conference on Pattern Recognition. vol. 1, pp. 582–585 vol.1 (1994)
33. Oliveira, R.B., Mercedes Filho, E., Ma, Z., Papa, J.P., et al.: Computational methods for the image segmentation of pigmented skin lesions: a review. Computer methods and programs in biomedicine pp. 127–141 (2016)
34. Oliveira, R.B., Papa, J.P., Pereira, A.S., Tavares, J.M.R.: Computational methods for pigmented skin lesion classification in images: review and future trends. Neural Computing and Applications **29**(3), 613–636 (2018)
35. Sadeghi, M., et al.: Detection and analysis of irregular streaks in dermoscopic images of skin lesions. IEEE Transactions on Medical Imaging (5), 849–861 (2013)
36. Sáez, A., Acha, B., Serrano, C.: Pattern analysis in dermoscopic images. In: Computer Vision Techniques for the Diagnosis of Skin Cancer, pp. 23–48 (2014)
37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
38. Situ, N., Wadhawan, T., Hu, R., Lancaster, K., Yuan, X., Zouridakis, G.: Evaluating sampling strategies of dermoscopic interest points. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 109–112 (2011)
39. Somwanshi, D., Chaturvedi, A., Mudgal, P.: Abcd features extraction-based melanoma detection and classification. In: International Conference on Artificial Intelligence: Advances and Applications 2019. Springer Singapore (2020)
40. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9 (2015)
41. Veit, A., et al.: Residual networks behave like ensembles of relatively shallow networks. In: Advances in Neural Information Processing Systems (2016)
42. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters **15**(5), 749–753 (2018)
43. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization pp. 2921–2929 (2016)
44. Zhou, Y., Smith, M., Smith, L., Warr, R.: A new method describing border irregularity of pigmented lesions. Skin Research and Technology **16**(1), 66–76 (2010)