

# Combining Deep Learning and Hand-Crafted Features for Skin Lesion Classification

Tomáš Majtner, Sule Yildirim-Yayilgan and Jon Yngve Hardeberg

Faculty of Computer Science and Media Technology,  
NTNU Norwegian University of Science and Technology, Gjøvik, Norway

e-mail: {tomas.majtner, sule.yildirim, jon.hardeberg}@ntnu.no

**Abstract**—Melanoma is one of the most lethal forms of skin cancer. It occurs on the skin surface and develops from cells known as melanocytes. The same cells are also responsible for benign lesions commonly known as moles, which are visually similar to melanoma in its early stage. If melanoma is treated correctly, it is very often curable. Currently, much research is concentrated on the automated recognition of melanomas. In this paper, we propose an automated melanoma recognition system, which is based on deep learning method combined with so called hand-crafted RSurf features and Local Binary Patterns. The experimental evaluation on a large publicly available dataset demonstrates high classification accuracy, sensitivity, and specificity of our proposed approach when it is compared with other classifiers on the same dataset.

**Keywords**—Convolutional Neural Network, SVM, RSurf Features, Local Binary Patterns, Skin Lesion Classification.

## I. INTRODUCTION

Based on the report from 2011, about 1500 melanoma cases are diagnosed each year in Norway [1]. The incidence of melanoma in this country is currently among the highest in the world and is still rising. Early diagnosis with rapid evaluation and surgical treatment cure up to 90 % of patients [2]. This is one of the examples, why the risk of melanoma is considered as a very important worldwide problem. Much research is currently focused on this issue with the aim to help doctors and patients in early melanoma recognition.

In the simplest form, the presented problem can be expressed as recognition of whether a captured dermoscopic image of skin contains a melanoma lesion or a benign lesion. Examples of these two types of lesions are shown in Fig. 1. The problem of correct skin lesion recognition is a difficult task. It usually consists of several stages, where the captured image is first preprocessed, a set of features is extracted, and as the last step, the input is classified to one of the pre-defined categories based on the extracted feature set.

Currently, two different types of feature sets are commonly employed in image recognition. The conventional type is based on the so called hand-crafted features, which are designed by researchers with the aim to capture visual characteristics of an image, typically its colour or texture. Recently, a new type of feature set inspired by biological processes and derived from the Convolutional Neural Network (CNN) [3] was introduced. In combination with deep learning, these new features tend to outperform hand-crafted features and therefore they have become very popular in computer vision.

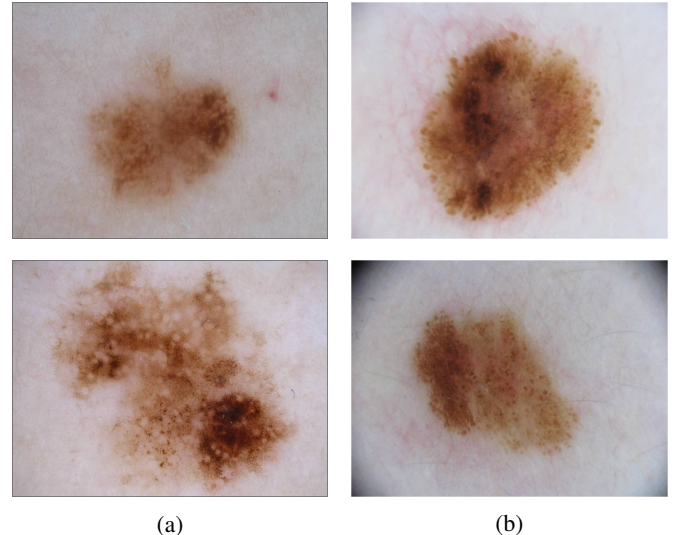


Fig. 1: An example of malignant melanoma lesions (a) and benign lesions (b). Images are from ISIC Archive (<https://isic-archive.com/>).

In this paper, we propose a combination of both types of features to be used for the skin lesion classification problem. We directly compare our results with the results from the recent melanoma classification challenge, which was hosted by the International Skin Imaging Collaboration (ISIC) and organized to support research and development of algorithms for automated diagnosis of melanoma [4]. The dataset used for this competition contains hundreds of images from different sources and competition results represent state-of-the-art for this classification problem.

## II. RELATED WORK

Several rules were established for doctors to help them correctly recognise skin lesions, namely the ABCD rule of dermoscopy or the 7-point checklist [5]. In automated diagnosis systems, texture and colour features are considered as two fundamental visual characteristics, which are vital for the detection of melanoma. Employing divergence-based colour features was presented in 2015 by Møllersen et al. [6]. These features are based on the divergence between the distribution of the pixel colour values of a lesion image, and the pixel colour values of either a benign or a melanoma

model. Several studies have already been done employing texture-based features like generalized co-occurrence matrices [7], gradient histograms [8], and RSurf features [9].

The combination of texture and colour features for the classification of dermoscopy images was presented by Riaz et al. [10]. Their texture features consist of a variation of Local Binary Patterns (LBP). For the colour feature extraction, the authors used standard HSV histograms. Methods based on deep learning were also employed for this problem. Codella et al. [11] present a combination of deep learning, sparse coding, and support vector machine (SVM) learning algorithms for melanoma classification. Their method achieved a classification accuracy of 93.1 % on a dataset containing 2624 clinical cases of melanoma, atypical nevi, and benign lesions. For the classification itself, various approaches have been used for melanoma images. The most common ones include  $k$ -NN [12], neural network [13], fuzzy  $c$ -means [14], linear discriminant analysis [6], and SVM [15].

Our contribution differs from the previous works in several manners. We use the combination of two different SVM classifiers. The first one uses hand-crafted features as input, represented by RSurf features [16] and Local Binary Patterns [17]. The second one employs features derived from CNN, where we had inspiration in recent research by Kawahara et al. on deep features for skin lesions [18]. Both of our classifiers predict the class for each tested image with a certain posterior probability or predicted class score [19]. These scores are subsequently used to determine the final classification.

We performed the evaluation on the publicly available dataset provided by ISIC for better comparison of our proposed method with state-of-the-art approaches. The same dataset was used in 2016 for the melanoma classification challenge [4]. The training part of this dataset contains 900 images (727 images of benign lesions and 173 images of melanoma) and the independent testing part contains 379 images (304 images of benign lesions and 75 images of melanoma). The entire dataset consists of images from various sources with various image distortions and classification tasks, and therefore better simulates real world usage of proposed solutions. Our method is among the top performing approaches in terms of classification accuracy, sensitivity, and specificity. In the subsequent sections of this paper, we introduce each individual component of our method and in the last section we present the classification results and ideas for future development and improvements.

### III. PROPOSED METHOD

As we already mentioned, our proposed method combines two different SVM classifiers. The first one is based on RSurf features and LBP. These image descriptors expect greyscale image as the input. Therefore, we first converted the original RGB image to greyscale by considering only the channel with the highest entropy  $E$ , which is defined as

$$E(I) = - \sum (p * \log_2 p),$$

where  $I$  is the examined image and  $p$  contains the histogram counts for every intensity value in  $I$ . This approach was used in [9], [20] and we chose it because the resulting greyscale image better captures the varying spatial distribution of the intensity values than traditional conversion from RGB to greyscale. Since the dataset images were from different sources and therefore had different resolutions, we also downsampled each of them to speed-up the feature extraction process, thus the resolution in the longest axis does not exceed 800 pixels.

#### A. RSurf Features

The first method we used for image description was introduced in 2014 and is called RSurf features [16]. Despite the fact that it was designed for images from a different domain, the high potential of RSurf features for melanoma recognition was also demonstrated [9]. The idea behind this feature set is based on dividing the image into parallel sequences of intensity values from the upper-left corner to the bottom-right corner. In these intensity sequences, all the local intensity maxima and minima are localized. Therefore, intensity values between two extreme points form either non-increasing or non-decreasing intensity sequences. We refer to them as *slopes* in further text. In general, one can sequentially scan the image and extract slopes in any direction. It was already recommended in [16] to scan the image in various directions in order to increase the robustness to rotational differences.

The set of all slopes extracted from all directions are used to extract slope properties. Four characteristic functions  $\Phi_1, \dots, \Phi_4$  were originally designed for mapping each slope to a real value. The first function  $\Phi_1$  corresponds to the length of the slope, the second function  $\Phi_2$  returns the height difference between the highest and the lowest point of the slope, the third function  $\Phi_3$  defines the sum of all intensity values, which belong to the slope, and finally the fourth function  $\Phi_4$  corresponds to the number of sign changes of the second derivative of the slope. In the end, for each characteristic function, a normalized histogram of function values is built. The concatenation of histogram values from all four characteristic functions forms the feature vector. The schematic visualization of RSurf features is shown in Fig. 2.

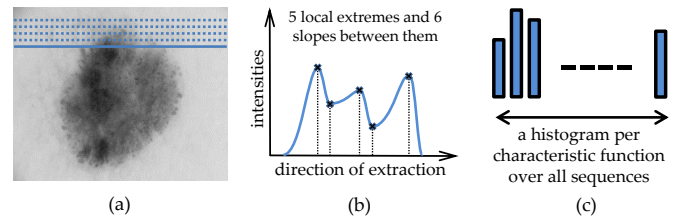


Fig. 2: The illustration of RSurf features. (a) parallel sequences, each represented by line dots, (b) slopes of a sequence, (c) 4 histograms, one for each characteristic function.

From the description one can see that RSurf features can be noise sensitive and therefore the use of Gaussian filter with small  $\sigma$  value is suggested prior to the calculation of

the features to suppress the noise in the input image. As it was demonstrated in [16], stronger smoothing decreases the classification performance of these features significantly. Therefore, prior to feature extraction, we apply Gaussian filter with  $\sigma = 0.5$ .

In the slope extraction process, the number of used directions should be determined. For various combinations of directions, we obtained almost identical results, thus this parameter has no significant influence on the results. Therefore, we decided to use the default option of  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ , where  $0^\circ$  represents the horizontal direction.

Another important parameter is the number of histogram bins used for each characteristic function, which defines the feature vector length. For each characteristic function, the same number of bins is used. The feature vector is a concatenation of all histogram bins from all four characteristic functions. We tested various feature vector lengths and the best results were achieved for 2000 elements, which corresponds to 500 bins for each characteristic function.

### B. Local Binary Patterns

Local Binary Patterns (LBP) were first introduced in 1994 by Ojala et al. [17]. The idea behind this descriptor is based on the model of texture units, where an input image is described through its texture spectrum. Ojala et al. extended this model and proposed its two-level version, which is now the core of the original LBP.

The computation of the original LBP consists of several steps. First, each pixel of the image is compared with its 8 immediate neighbours. If the intensity value of the neighbouring pixel is greater than or equal to the value of the examined pixel's intensity, write 1 in the neighbouring pixel. Otherwise, write 0. The described process gives us an eight-digit binary number for each pixel in the image. For convenience, we convert this number to decimal by taking the neighbours in clockwise order from the top-left corner. These decimal numbers are used to form a histogram and in the end, a concatenation of the normalized histogram values gives us the feature vector. Fig. 3 illustrates the process of building the feature vector for the original LBP version.

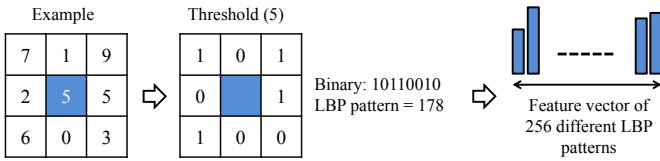


Fig. 3: The process of building the feature vector in original LBP variant.

The original algorithm was further enhanced and in 2002, the basic LBP operator was introduced [21]. The idea for this operator is based on the assumption that each texture has two complementary aspects, a pattern and its strength. The main innovation is putting no limitation to the size of

the neighbourhood (parameter  $R$ ) or the number of sampling points (parameter  $P$ ). Examples of various local circular neighbourhoods defined by these new parameters are shown in Fig. 4.

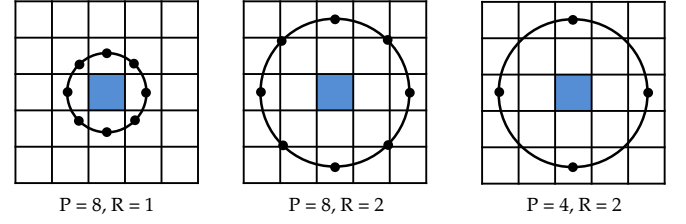


Fig. 4: The circular neighbourhoods with various parameters  $P$  and  $R$ . The pixel values are bilinearly interpolated, when the sampling point is not in the centre of the pixel.

In our experiments, we used the version with 8 sampling points ( $P = 8$ ) and three different radii ( $R = 1, 3, 5$ ), thus our feature vector length is  $3 \times 2^8 = 768$ . This version is typically denoted as  $LBP_{R=1,3,5}$ . LBP has many different variants, which were successfully applied in a number of domains [22], including the recent results for melanoma recognition [10], [23].

### C. First SVM Classifier

The first classification is based on SVM classifier with Gaussian kernel and standardized predictors. It combines RSurf features and  $LBP_{R=1,3,5}$  to estimate the class for a given input image as is shown in Fig. 5. The input feature vector consists of 2768 components. Each image class label is predicted with a certain score representing the posterior probability that an observation belongs in a particular class, given the data [19]. It can be interpreted as the strength of classification decision and we use it later in our final solution.

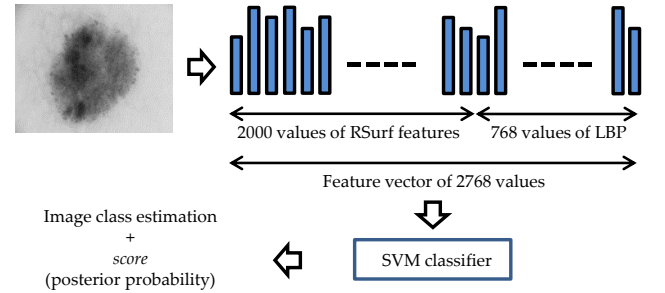


Fig. 5: The illustration of the image class estimation by the first SVM classifier.

We tested this first classifier based on the same protocol as was used in the already referred classification contest [4]. We used 900 train images to extract features and train the SVM to create the model, and subsequently we use independent 379 test images for evaluation. The independent test set is a guarantee that results are not biased, which is

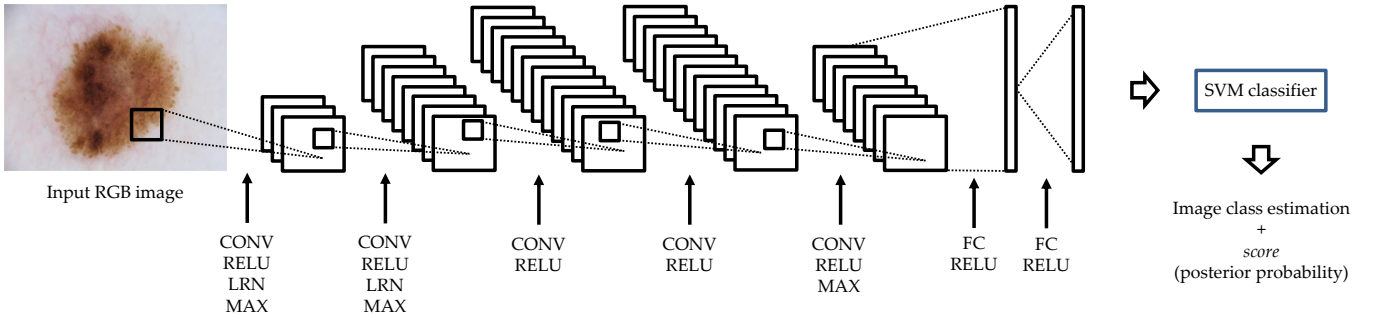


Fig. 6: The illustration of AlexNet combined with SVM, which we used in our experiments. The visualisation was inspired by work of Karnowski [24]. Individual shortcuts of the layers are described in the text.

a common problem of cross-validation evaluation [25]. The average precision for the first classifier is 0.405, we achieved the classification accuracy of 0.794 with sensitivity of 0.507, and specificity of 0.866. These results are comparable with the top performing classifiers from the classification challenge, in case these traditional metrics, suggested also in [26], are considered.

#### D. Second SVM Classifier (AlexNet)

The second SVM classifier with Gaussian kernel and standardized predictors uses features derived from CNN. Specifically, we used the AlexNet [3] proposed by Krizhevsky et al. in 2012. Authors employed the CNN for the first time for image classification and won the ImageNet Large Scale Visual Recognition Challenge 2012 with significant lead over the other participants. Despite the fact that AlexNet was created for real world images, it was also applied in combination with SVM for melanoma recognition [11], [18].

The original version of AlexNet contains five convolutional layers (CONV) and three fully-connected layers (FC). Each of these layers is followed by a point-wise non-linear activation layer called rectified linear units (RELU). There is also a local response normalization layer (LRN), which follows the first two convolutional layers. AlexNet has three max-pooling layers (MAX), two after the LRN layers and the third one following the fifth convolutional layer. The core structure of AlexNet locates between the second and the third max-pooling layers, which contains three convolutional layers, each with  $3 \times 3$  convolution kernels [27]. More details about the overall structure and design of AlexNet can be found in the publication by Krizhevsky et al. [3].

In our experiments, we used MatConvNet [28], which is a MATLAB toolbox implementing Convolutional Neural Networks for computer vision applications. The original RGB input images were used and AlexNet was employed only for calculation of the second fully-connected layer, as it is illustrated in Fig. 6. This layer has 4096 elements, which corresponds to the same number of neurons. These elements are, in our experiment, considered as the feature vector and they are subsequently used as inputs to SVM. This is a standard approach employed, for example, in [29].

We evaluated the second SVM classifier with the same protocol as the previous one and results are presented in Table 1. Based on this comparison, deep learning features confirm their tendency to outperform hand-crafted features also for this particular dataset. However, it is important to emphasize that in the first SVM classifier based on hand-crafted features, only one colour channel was considered, since the employed texture-based feature extractors expect greyscale image as input. Therefore, the number of input information was not equal in both classifiers. In the case of using only greyscale input for the second classifier based on CNN features, the results will be worse than the results presented for the first classifier.

	AP	Accuracy	Sensitivity	Specificity
<i>First SVM</i>	0.405	0.794	0.507	0.866
<i>Second SVM</i>	0.470	0.805	0.533	0.872

Table 1: The comparison between first SVM classifier based on hand-crafted features and second SVM classifier based on AlexNet. AP represents the average precision.

## IV. FINAL EVALUATION AND RESULTS

As was mentioned before, both of the tested SVM classifiers provide also scores defined as the mean accuracy of the given test data and labels. We use these values to determine, which classifier is more confident about the assigned label. For each tested image, we choose the label with the highest absolute score value. The final classifier is therefore a combination of both principles, the hand-crafted features and the features derived from the deep learning mechanism, as it is illustrated in Fig. 7.

Our final classifier has the average precision of 0.473, which is only slight improvement over the previously mentioned results but the classification accuracy is 0.826 with the sensitivity of 0.533 and specificity of 0.898. This performance is among the top state-of-the-art classifiers, as is shown in Table 2, which summarises the results of the melanoma classification challenge. For completeness, we also provide the precision-recall graph in Fig. 8 and the ROC curve in Fig. 9. The area under the ROC curve is 0.780.



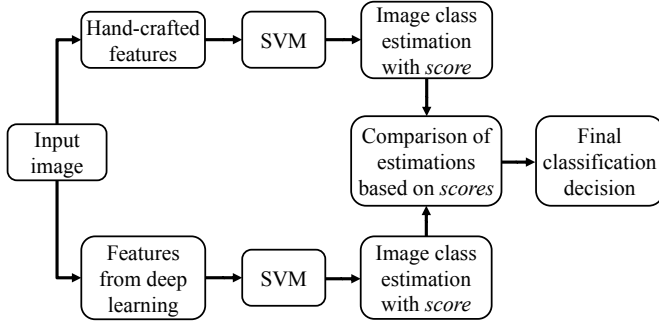


Fig. 7: The illustration of the structure of our final classifier.

Since there are more metrics that could be considered for classification evaluation, it is hard to clearly conclude, which approach is the highest performing one. For different purposes, different approaches maximizing a particular metric could be more appropriate. However, the presented results demonstrate our approach as one of the top performing for melanoma classification. There is still a room for improvement but these first results indicates that with combining hand-crafted features with deep learning features we easily build high performing solution for melanoma classification, and therefore this approach is very promising for the future.

Authors	Accuracy	Sensitivity	Specificity	AUC
Lequan Yu	<b>0.855</b>	0.507	0.941	0.804
Jordan Yap	0.844	0.240	<b>0.993</b>	0.775
Balde Thierno	0.834	0.320	0.961	<b>0.826</b>
<b>Our method</b>	0.826	0.533	0.898	0.780
Haebeom Lee	0.821	0.200	0.974	0.793
Nandita Damaraju	0.815	0.467	0.901	0.813
Sahbi Chaieb	0.813	0.573	0.872	0.802
Andre Esteva	0.786	0.667	0.816	0.796
Sebastien Paris	0.731	0.773	0.720	0.793
Mohit Agarwal	0.681	0.787	0.655	0.800
Euijoon Ahn	0.599	<b>0.853</b>	0.536	0.780

Table 2: The comparison of our final classifier with the top ten participants of the melanoma classification challenge [4] ordered by the classification accuracy. AUC represents the area under the ROC curve. The challenge results were downloaded from website <https://challenge.kitware.com/#phase/5667455bcad3a56fac786791> on 28/06/2016.

## V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a fully automated melanoma classification approach, which is based on the combination of RSurf features and Local Binary Patterns together with the features derived from a Convolutional Neural Network. Each component of the presented approach was tested on a large publicly available dataset with independent training and testing images. For evaluation purposes, we derived various classifier characteristics but our aim in designing the classifier was maximization of the classification accuracy, sensitivity, and specificity, which were recognized as one of the most important classifier metrics [26]. Based on the classification

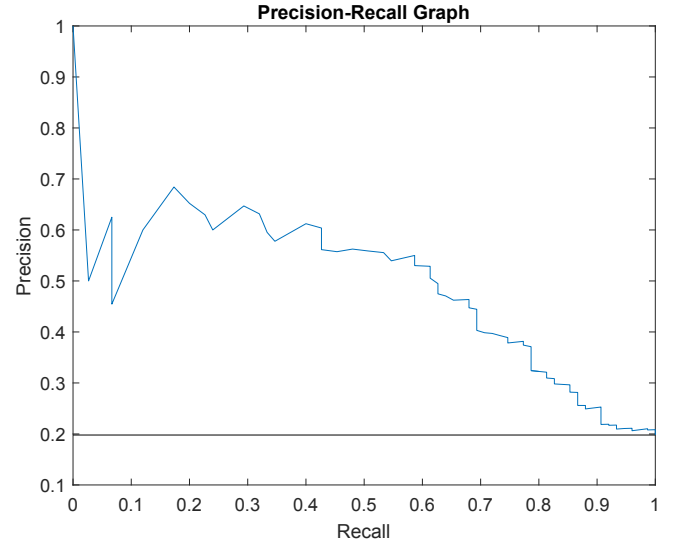


Fig. 8: The precision-recall curve for the final classifier.

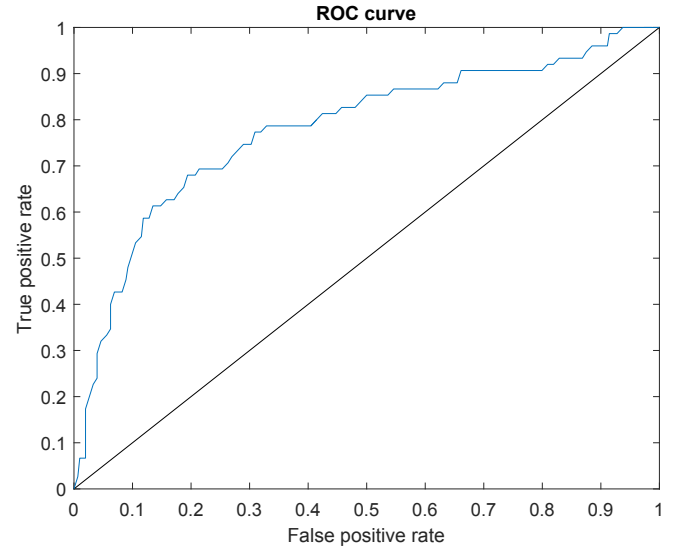


Fig. 9: The ROC curve for the final classifier. The area under the ROC curve is 0.780.

accuracy, our approach is the fourth best performing classifier, when compared with state-of-the-art classifiers participating in the melanoma recognition challenge [4]. These results are very promising and suggest that approaches based on hand-crafted features in combination with features derived from CNN have high potential in melanoma recognition.

In future work, we will concentrate on creating a new model for MatConvNet, which will be aimed specifically on skin lesion images. Currently, we used the pre-trained AlexNet model [3], which is very robust but has some limitations when used in different research domains. Since the dataset used in this article is very large and contains images from different sources, various image distortions and artefacts are also present in images. Therefore, we will aim at creating a

robust pre-processing step to eliminate all the possible artefacts and enhance the feature extraction process.

#### ACKNOWLEDGMENT

This research has been supported by the Research Council of Norway through project no. 247689 “IQ-MED: Image Quality enhancement in MEDical diagnosis, monitoring and treatment”.

#### REFERENCES

- [1] T. E. Røbsahm, T. B. Johannesen, and I. M. et al. Bachmann, “Nasjonale retningslinjer for diagnostikk, behandling og oppfølging av maligne melanomer,” *Oslo: Helsedirektoratet*, 2011.
- [2] J. Geisler, I. M. Bachmann, M. Nyakas, P. Helsing, H. E. Fjøsne, L. O. Maehle, S. Aamdal, N. A. Eide, H. L. Svendsen, and O. et al. Straume, “Malignant melanoma—diagnosis, treatment and follow-up in Norway,” *Tidsskrift for den Norske lægeforening: tidsskrift for praktisk medicin, ny række*, vol. 133, no. 20, pp. 2154–2159, 2013.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [4] D. Gishman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, “Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC),” *CoRR*, vol. abs/1605.01397, 2016.
- [5] G. Argenziano, H. P. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, L. Cerroni, G. De Rosa, G. Ferrara, et al., “Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet,” *Journal of the American Academy of Dermatology*, vol. 48, no. 5, pp. 679–693, 2003.
- [6] K. Møllersen, J. Y. Hardeberg, and F. Godtliebsen, “Divergence-based colour features for melanoma detection,” in *Colour and Visual Computing Symposium*. IEEE, 2015, pp. 1–6.
- [7] L. Ballerini, R. B. Fisher, B. Aldridge, and J. Rees, “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] C. Barata, M. Ruela, M. Francisco, T. Mendonça, and J. S. Marques, “Two systems for the detection of melanomas in dermoscopy images using texture and color features,” *IEEE Systems Journal*, vol. 8, no. 3, pp. 965–979, 2014.
- [9] T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, “Efficient Melanoma Detection Using Texture-Based RSurf Features,” in *International Conference Image Analysis and Recognition*. Springer, 2016, pp. 30–37.
- [10] F. Riaz, A. Hassan, M. Y. Javed, and M. T. Coimbra, “Detecting melanoma in dermoscopy images using scale adaptive local binary patterns,” in *Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 6758–6761.
- [11] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, “Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images,” in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2015, pp. 118–126.
- [12] A. Glowacz and Z. Glowacz, “Recognition of images of finger skin with application of histogram, image filtration and k-NN classifier,” *Biocybernetics and Biomedical Engineering*, 2016.
- [13] J. Premaladha and K. S. Ravichandran, “Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms,” *Journal of Medical Systems*, vol. 40, no. 4, pp. 1–12, 2016.
- [14] A. T. Beuren, R. J. G. Pinheiro, and J. Facon, “Color approach of melanoma lesion segmentation,” in *Systems, Signals and Image Processing*. IEEE, 2012, pp. 284–287.
- [15] S. Gilmore, R. Hofmann-Wellenhof, and H. P. Soyer, “A support vector machine for decision support in melanoma recognition,” *Experimental Dermatology*, vol. 19, no. 9, pp. 830–835, 2010.
- [16] T. Majtner, R. Stoklasa, and D. Svoboda, “RSurf—The Efficient Texture-Based Descriptor for Fluorescence Microscopy Images of HEP-2 Cells,” in *22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 1194–1199.
- [17] T. Ojala, M. Pietikäinen, and D. Harwood, “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,” in *International Conference on Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision and Image Processing*. IEEE, 1994, vol. 1, pp. 582–585.
- [18] J. Kawahara, A. BenTaieb, and G. Hamarneh, “Deep features to classify skin lesions,” in *13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 1397–1400.
- [19] J. C. Platt, “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods,” in *Advances in Large Margin Classifiers*. 1999, pp. 61–74, MIT Press.
- [20] E. Guerra-Segura, C. M. Travieso-González, J. B. Alonso-Hernández, A. G. Ravelo-García, and G. Carretero, “Symmetry Extraction in High Sensitivity Melanoma Diagnosis,” *Symmetry*, vol. 7, no. 2, pp. 1061–1079, 2015.
- [21] T. Ojala, M. Pietikäinen, and T. Maenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, July 2002.
- [22] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision Using Local Binary Patterns*. Computational Imaging and Vision. Springer Verlag, London, 2011.
- [23] S. Naeem, F. Riaz, A. Hassan, and R. Nisar, “Description of Visual Content in Dermoscopy Images Using Joint Histogram of Multiresolution Local Binary Patterns and Local Contrast,” in *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 2015, pp. 433–440.
- [24] Jeremy Karnowski, “AlexNet + SVM,” 2015, Available Online: <https://jeremykarnowski.files.wordpress.com/2015/07/alexnet2.png>; accessed 24/06/2016.
- [25] P. Smialowski, D. Frishman, and S. Kramer, “Pitfalls of supervised feature selection,” *Bioinformatics*, vol. 26, no. 3, pp. 440–443, 2010.
- [26] N. K. Mishra and M. E. Celebi, “An overview of melanoma detection in dermoscopy images using image processing and machine learning,” *arXiv preprint arXiv:1601.07843*, 2016.
- [27] Y. Liu, B. Yin, J. Yu, and Z. Wang, “Cross-Level: A Practical Strategy for Convolutional Neural Networks Based Image Classification,” in *CCF Chinese Conference on Computer Vision*. Springer, 2015, pp. 398–406.
- [28] A. Vedaldi and K. Lenc, “MatConvNet: Convolutional Neural Networks for Matlab,” in *23rd ACM International Conference on Multimedia*. ACM, 2015, pp. 689–692.
- [29] Y. Zhou and N.-M. Cheung, “Vehicle Classification using Transferable Deep Neural Network Features,” *arXiv preprint arXiv:1601.01145*, 2016.