

Search for Resolution Invariant Wavelet Features of Melanoma Learned by a Limited ANN Classifier

GRZEGORZ SURÓWKA

Faculty of Physics, Astronomy and Applied Computer Science,
Jagiellonian University, 30-151 Kraków, Poland
Institute of Technology, State Higher Vocational School, 33-300 Nowy Sącz, Poland
e-mail: *grzegorz.surowka@gmail.com*

Abstract. This article addresses the Computer Aided Diagnosis (CAD) of melanoma pigmented skin cancer. We present back-propagated Artificial Neural Network (ANN) classifiers discriminating dermoscopic skin lesion images into two classes: malignant melanoma and dysplastic nevus. Features used for our classification experiments are derived from wavelet decomposition coefficients of the image. Our research objective is i) to select the most efficient topology of the hidden layers and the network learning algorithm for full-size and downgraded image resolutions and, ii) to search for resolution-invariant topologies and learning methods. The analyzed classifiers should be fit to work on ARM-based hand-held devices, hence we take into account only limited learning setups.

Keywords: melanoma, CAD, wavelets, ANN

1. Introduction

Computer aided classification of the benign form (dysplastic nevus) and the malignant form (melanoma) of the pigmented skin lesions plays a crucial role in the prevention of human melanoma. Melanoma, as contrasted with the other forms of skin cancer, is extremely dangerous due to early metastases. Early diagnosis of this tumor is a life-saving factor. Lack of specialists, too late detections and an increased melanoma

morbidity rate have become a medical problem for some time and, on that account, a challenge for computer assisted diagnosis (CAD) [1–3]. The standard therapy, biopsy, is not feasible for all the patients due to treatment costs and some health reasons. Even if affordable, excision must be done at an early stage, when standard clinical diagnosis based on the ABCD criteria may fail [4]. For that reason, analysis of skin lesions images have become a useful diagnostic tool. The most popular and cheapest form of lesion screening is dermoscopy (ELM-Epiluminescence Microscopy) [5]. This technique uses optics and white light illumination of the lesion and stores the skin image usually on a computer. This is extremely useful for comparing past and current advances of the lesion. Some advanced instruments can illuminate the lesion at different angles or with a set of different wavelengths. This helps penetrate deeper layers of the skin to reveal its spatial structure [6–8]. Their coverage is however limited and the most common are the cheapest, handy dermatoscopes.

Stages of pigment cells atypia are recognized by medical doctors with help of the semi-quantitative descriptive measures: ABCD(E), the 7-Point Checklist, Menzies and other less common [9]. Clinical diagnosis of Melanoma depends on the appearance of classic dermoscopic features but unfortunately the so called early melanomas are mainly featureless [10].

The computer aided melanoma diagnosis is roughly divided into two classes of methods:

- Segmentation-based methods: they assume that melanoma can be visually examined based on shape, color and structure content. There are two main reasons why those methods may fail:
 - different illumination and optical magnification plus noise and artefacts (hair, blistering areas) can make such observations/analyses not sensitive enough [11],
 - search for featureless melanomas is infeasible.
- Texture-based methods (wavelet- or curvelet- based): they attempt to search for some frequency and scale information of the skin texture [12] to distinguish between benign and malignant skin progression [13, 14]. The related works referenced in this article (Sec. 2) prove that especially wavelet-based methods are robust and show high performance in the melanoma classification.

There exists a big market for (para)medical smartphone applications. Both medical doctors, who are interested in medical decision-supporting tools, and plain consumers, who want to quickly 'self-diagnose' and be aware of personal health issues, show interest. The self-examination/self-diagnosing optical extensions and grip panels for mobile phones are very popular. There is also a growing market for handy dermatoscope-like devices with optics and ARM-based processors for a 'mobile' CAD and melanoma diagnosis [15, 16].

Since the developed machine learning and image recognition and interpretation algorithms may demonstrate high complexity and small handheld devices have limited processing power and memory, it is of great importance to search for methods that analyze optimally both full-size dermoscopic images and preserve high efficiency also in downgraded image resolutions.

In Sec. 2 research on wavelet based features and methods is presented. One conclusion of the latest advances [17] is that different wavelet families perform not equally in terms of the melanoma classification efficiency. The reverse bi-orthogonal (RBio) and bi-orthogonal (Bior) wavelets prove to be the most efficient, robust, and resolution invariant wavelet families for the machine learning of dermoscopic images. In [17] ensemble learning of different model types and optimized for various quality measures is performed for the wavelet features. In this work we want to check efficiency and classification performance of a single, homogeneous model taking into account the most successful wavelet families concluded in [17]. We pick the RBio3.1 wavelet base and study back-propagated artificial neural network (ANN) classifiers of the melanoma (RBio3.1)-based features with limited topologies for the hidden layers to take into account execution platforms with limited processing resources.

Within these bounds our objectives are:

- to select the best NN topology and learning method in terms of absolute melanoma classification performance upon condition of efficient cost/time of ANN back-propagation learning for three different resolutions of the dermoscopic images,
- to search for a resolution-invariant model of ANN for the original clinical data and the descendant downgraded image resolutions.

Below we review previous research on both wavelet-based features of melanoma and NN-based melanoma classifiers. Then we show methodology of our machine learning experiments and present and discuss the results.

2. Related work

The first contribution to wavelet based decomposition of melanoma dermoscopy images belongs to Patwardhan et al. [18, 19]. This group successfully studied binary classification models for benign nevus and melanoma by decomposing different frequency scales of the skin texture (wavelet packets). This approach also known as selective wavelet trees in each iteration decomposes all the sub-bands of the pigmented skin texture rather than the low-frequency sub-band only (recursive algorithm). Since that time the wavelet-based features have also been successfully studied by other groups [20–28]. The binary classifiers from Patwardhan [18, 19], and early contributions from [20–22], were using only one wavelet base (Daubechies 3) to build classification models. Recent research on efficiency of different wavelet bases in classification of melanoma and benign dermoscopy images shows that bi- and reverse bi-orthogonal wavelets outperform other wavelet families. In this article we use the reverse bi-orthogonal wavelet 3.1 (RBio3.1) as the transform base.

Neural learning of wavelet-based features at various stages of melanoma development was first studied by Walvick et al. [13]. Their feature set was then optimized by removing redundancies with help of principal component analysis. A feed forward neural network trained with the back propagation algorithm was then used in the classification process to obtain better classification results. In [29] a two level wavelet transform was performed on the segmented image in order to take into account only the lesion region and not the surrounding tissue. In order to maintain the

original image size, a stationary wavelet transformation was used which suppresses down-sampling, producing a wavelet matrix of the same size as the input image. For the approximations and details components mean and variance of wavelet coefficients were calculated and normalized (0,1). Three wavelets from four different families (Daubechies, Coiflet, BiorSplines, Symlet) were tested to see the impact on the classification. The feed forward back propagation neural network consisted of 8 input nodes (for 8 features), 2-5 hidden nodes and one output node and was trained/tested by the LOO technique. The overall classification accuracy was about 83%-86% for all the wavelet bases for the best performing network with 3 hidden nodes.

Also [24] employed the variance and mean of wavelet decomposition coefficients to extract useful features of dermoscopy images. Those coefficients were used as the inputs of neural network. Results showed 90% ability in distinction between benign and malignant lesions.

Work [30] segmented suspicious lesions from the normal skin and extracted features that distinguish malignant melanomas from benign moles with 2D wavelet transform in Matlab. Some selected features formed the input to an ANN classifier. Authors claim accuracy to be 84% but no details about the procedures are available.

An interesting research was done in [31]. This paper proposes a melanoma classification system based on coefficients created by wavelet and curvelet decompositions. The curvelet transform returned the set of curvelet coefficients indexed by scale, orientation and location parameters. Fast Discrete Curvelet Transform via wedge wrapping was used to find the features. Recognition accuracy of the three layers (40-25-10) back-propagation neural network classifier with wavelets was 51.1%, which is a surprisingly low value, and with curvelets 75.6%, not high neither.

Good indication for lesion malignancy is border irregularity. Work [32] presented a contours harmonic wavelet coefficients forming a sequence of multi-scale roughness descriptors to characterise the distribution of energy across contour's line. The descriptors for regular and irregular borders were compared with ground truth from experts. Differentiating between benign and malignant lesions according to those descriptors led to maximum classification accuracy of 93.3% with sensitivity of 80%.

Similar work was reported in [14] where wavelet decomposition was used to extract energy distributions among different wavelet sub-bands. Statistical and geometric irregularity descriptors based on the wavelet coefficients were used to model structural components from the contour. The effectiveness of the descriptors was measured using the Hausdorff distance between sets of data from melanoma and mole contours. The best descriptor outputs were input to a back propagation neural network to construct a combined classifier system. Experimental results showed that the selected thirteen multi-scale features with small sample set produced the best discrimination performance with $AUC = 0.89$ and best specificity of 90%, and sensitivity of 83%.

Artificial neural networks (ANN) in general are widely used in various pattern recognition problems, also in CAD. There are different types of neural systems well suited to classify visual features from (medical) images. In this short review we list only standard back-propagated neural networks (not e.g. CNN). In the melanoma classification problem ANNs are used as one of the following:

- the main classification engine (melanoma-benign lesion) for all/selected features,
- dedicated expert system for a single segmentation task (e.g. edge detection), or

- auxiliary ML system for initial/post- analysis and verification.

First ANN classification of melanoma and benign lesions comes from [33] where about 80% correct classifications were reported. The network topology (14-X-1, X-hidden neurons) accepted two asymmetry- and twelve color-based features.

In [34] authors analyzed basic, shape and color features with different normalization conditions and concluded that on both dichotomous and trichotomous tasks, the ANNs performed (sensitivity: 91%, specificity: 94% for 1619 lesion images) similarly as logistic regression and SVMs, and better than k-nearest neighbors and decision trees.

In [35] authors analyzed 48 parameters belonging to four categories: geometry, color, texture, and color clusters inside the lesion and performed step-wise feature selection to identify an optimal subset of 10 variables (starting from the most significant: red multicomponent, decile of red, border homogeneity, mean value of red, grey-blue areas, contrast, interruptions of the border, mean skin-lesion gradient, background regions imbalance, variance of the border gradient). The clinical/dermoscopic equivalents of those variables are: multicomponent pattern/homogeneity, lesion darkness, border cleanliness, mean color of the lesion, grey-blue areas, network analysis, variation in the border cleanliness, grading of the border, color asymmetry, intensity in the border interruptions. Distinguishing melanoma from benign lesions with these optimal features gave the maximum sensitivity of 93% and specificity 92.75%.

Project DANAOS [36] aimed at analyzing robustness of the NN-based machine learning system with respect to multi-population lesion samples. Its authors concluded that the performance of their NN expert system is comparable with that of clinicians, with the average AUC of 84,4%.

Rajab et al. [37] investigated the neural network edge detection (NNED) in the iterative thresholding segmentation. They drew conclusions on the method performance over a range of different border irregularity properties and signal-to-noise ratio.

Preprocessing of dermoscopic images through the Fourier and log-polar transforms was used to build an unsupervised image segmentation and image registration system where neural networks and discriminant analysis were used to find the best classification rules for the extracted border- and color-based features [38].

Article [39] described a multi-layer perceptron classifier for melanoma recognition with accuracy of 77.7%. The number of features for classification was optimised to only five which speeds up the CPU time.

In [40] a service on the Internet was introduced to upload dermoscopy images for on-line extraction of the tumor area and calculation of 428 global features (color, symmetry, border, and texture ABCD) for the characterization of the lesion. The extracted features classified the lesion as melanoma or nevus using a neural network classifier achieving a sensitivity of 85.9% and a specificity of 86.0% on a set of 1258 dermoscopy images using cross-validation.

A co-operative neural network-based edge detection on enhanced colors and contrast of the image is reported in [41]. As supervisors (ground truths) three expert dermatologists were used.

An automatic neural skin cancer classification system was developed in [42] for dermoscopy images and optimized for different types of neural network topologies and different preprocessing modes. The authors reported best recognition accuracy

of the 3-layers back-propagation neural network classifier as 89.9% and of the auto-associative neural network as 80.8%. In the system some features were extracted through 2-D wavelet packet decomposition under performance tests of seven different wavelet bases (the best wavelet base was Bior5.5, and the most stable experimentally Db1 or Db10). Some optimization was done to the number and structure of the hidden layers. The best topology was reported for the three hidden layers with 40-25-10 neurons.

In [43] a back-propagation neural network (BPN) is used for segmentation. The results are compared with the ground truth images which shows that BPN has slightly worse segmentation accuracy and slower training period than Extreme Learning Machine (ELM).

Authors in [44] claims rather low accuracy of the melanoma discrimination with back propagation neural network which yields 60%–75%. This outcome is outperformed by the SVM models.

Group [45] segmented dermoscopic images using Maximum Entropy Threshold and extracted features using Gray Level Co-occurrence Matrix (GLCM) for classification into cancerous or non-cancerous cases using back-propagated neural networks (BPN). The reported accuracy is 88%.

Paper [46] proposed the flow: feature extraction, dimensionality reduction and classification to discriminate skin lesions into 'normal' and 'abnormal' skin cancer classes. In the first stage authors used discrete wavelet transforms, in the second stage PCA. In the classification stage a feed forward back-propagated artificial neural network and a k-nearest neighbor paradigmes were applied. Accuracy of those experiments was: 95% (ANN) and 97.5% (kNN).

Discrimination between the six Menzies color classes in the calibrated RGB dermoscopy images were studied in [47]. The JeffriesMatusita and transformed divergence separability distances were used to evaluate the color class separability. A non-linear cluster transformation allowed almost the total separation of each color class in the feature space. Several neural networks in competition were used as classifiers. Classification achieved 93% of sensitivity, 62% of specificity and 74% of accuracy (average). Authors claim that it might be possible to evaluate a lesion based on the presence of Menzies colors in the dermoscopic image, mimicking the human diagnosis.

In [48] statistical features and dermoscopic features (ABCD: Asymmetry, Border, Color and Diameter) for detection and diagnosis of melanoma were used. Segmentation-based thresholding plus statistical feature extraction using GLCM were used to calculate the Total Dermoscopy Score (TDS). The combined result of the TDS parameter and a neural network classification yielded accuracy of 88% which was claimed to be efficient for the skin cancer detection and diagnosis.

Article [49] presents a melanoma detection system working in two phases: the first phase detects whether the skin lesion is of pigment type, the second phase distinguishes between malignant melanoma and benign skin lesions. The reported classification results from the neural network are about 98% (phase 1) and 93% (phase 2). Although algorithms and methods are presented step-by-step some details cannot be derived. The high results are controversial.

Article [50] analyzes the recognition performance of three different classifiers: support vector machine (SVM), artificial neural network (one hidden layer, sigmoid transfer function) and k-nearest neighbor. From experiments run on a database of more

than 5000 dermoscopy images they concluded that the SVM approach outperforms the other methods reaching an average recognition rate of 82.5% comparable with those obtained by skilled clinicians. So this is an upper limit for average recognition rate by neural network.

In [51] several methods of melanoma classification were proposed: a multilayered perceptron, a Bayesian classifier and the K nearest neighbors algorithm. These methods worked independently and also in combination making a collaborative decision support system. The performance factors obtained for seven neurons in a single hidden layer were: classification rate: 86.73%, sensitivity: 78.43%, specificity: 95.74%, slightly less than in the collaborative method: classification rate: 87.76%, sensitivity: 78.43%, specificity: 97.87%.

In the review [52] authors surveyed computer-based systems according to acquisition, feature definition, extraction and skin lesion classification. Authors concluded that some widely used lesion parameters like lesion size, shape, color, and texture do not correspond to known biological phenomena and the structural patterns that are considered essential for manual lesion categorization are absent in the analysis due to their complexity. From the analyzed machine learning methods (discriminant analysis, neural networks, support vector machines) SVM performed the best. Neural Networks using two principal components as input produced 85% correct matches (sensitivity 79%, specificity 90%) for the vertical growth phase of melanoma development and 94% correct matches (sensitivity 86%, specificity 90%) for the radial growth phase.

Using meta-analytical methods Authors of [53] compared the diagnostic accuracy of the different dermoscopic algorithms with each other and with the artificial intelligence methods for the detection of melanoma. They concluded that pooled sensitivity for artificial intelligence was slightly higher than for dermoscopy (91% vs. 88%) and pooled specificity for dermoscopy was significantly better than artificial intelligence (86% vs. 79%). There was no significant difference in the diagnostic performance of various dermoscopy algorithms (ABCD, 3-point checklist, 7-point checklist, Menzies score). A useful 1994-2006 summary for characteristics of the included studies is included.

Authors of article [1] classified and researched literature on the CAD systems for melanoma identification. They tabularized lots of references according to the generic steps of CAD: image preprocessing, border detection, features/descriptors extraction and classification. Low opinion was expressed on the lack of common standards and absence of benchmark datasets for standardized algorithm evaluation.

In the review [2] the state of the art of melanoma CAD was examined: in-vivo imaging techniques, image acquisition, pre-processing, segmentation methods, feature extraction and selection, and classification of dermoscopic images. Of high value are the indications of various conditions that affect the technique's performance. The case of artificial neural networks for the melanoma classification is broadly represented in the following references.

Article [54] reviews to a lesser extent the 2011-2014 works in the skin cancer detection. Classifier performance results from other existing melanoma CAD systems can be found in [55]. The last review [3] gives a vast look at different feature types

and classification methods. It reports and classifies the 2007–2015 studies paying attention at referring them to the global and local patterns.

3. Data and Methods

3.1. Data

The database collected for this study included anonymous images of the moles from 185 patients of one private clinic in Poland (the formal agreement forbids publications of the location details to keep the patients data unidentified). The examinations were performed with plain digital camera with an extra dermoscopy extension and immersion liquid to remove light reflections. The primordial JPEG pixel resolution was 2272x1704 and the RGB color depth 24-bit. The resection and hist-pat examination of the moles allowed to assign labels to 102 malignant melanoma and 83 dysplastic nevus cases. The 83 non-melanoma images were picked up randomly from the predominant majority of about 2000 displastic lesions. In our analysis there were no 'unknown' or 'don't care' labels.

Melanoma incidence rate may fluctuate over countries, but clinical statistics show an average of about 5% melanoma images as a fraction of all the dermoscopic images of the pigmented nevi. This means that the melanoma class is under-represented compared to the benign class. Learning classifiers from such cases would require special rules to properly treat the imbalanced class i.e. to draw equal attention to the minority class [56–58]. In this experiment we took the whole statistics for the minority class (melanoma, 102 cases) and then randomly selected the 83 non-melanoma images from the larger pool (about 340) covering the pigmented benign lesions (dysplastic junctional nevus, displastic compound nevus). In clinical practice only the above mentioned benign pigment cases are confused with melanoma. The ground truth was in each case the hist-pat examination.

In this work we do not analyze in detail the impact of data balancing techniques for the classification of malignant melanoma (see e.g. [58]) but we estimated if such impact affects our classifiers. Three different under-sampling procedures of the majority class in the data space were performed to balance the data vectors learned by the classifier before cross validating the classifier models. No major change in classification performance ($< 2\%$) was observed. This allows us to treat the data imbalance problem in our experiment under control. This of course does not mean that there is no impact of the data imbalance problem at all but that the bias comes preferably from the 'clinical' source and not the data statistics or procedures.

In the analysis there were three sets of images: the original set 2272x1704 (A) and the two downsampled sets (by averaging neighbor values in 2x2 elements) of 1136x852 (B) and 568x426 (C) pixels respectively.

3.2. Features

There were no apparent artefacts on the analyzed dermoscopy images (black borders, hairs, droplets of immersion fluid, etc.) or there were few negligible distortions so no preprocessing tasks to the images took place.

To support wavelet transformations the dermoscopy images of all three sets (A, B, C) were transformed to indexed images with linear, monotonic color maps of double precision numbers. Each iteration of the wavelet decomposition downscales the input image by 2 both in rows and columns and three such iterations were done.

Wavelet decomposition of signals is well established in theory after works of Mallat [59], Daubechies [60] and the others (Gabor, Morlet). It is widely applied especially to discrete signals in the form of DDWT-Discrete Dyadic Wavelet Transform. This transform is widely used to analyze the signal structure, signal de-noising and compression capabilities.

Images are two-dimensional signals and the wavelet transform to the images are done according to the Mallat algorithm [59]. One iteration of this algorithm produces 4 downscaled sub-images which can be considered as LL, LH, HL and HH filters (L-low-pass, H-high-pass filter) after one-dimensional wavelet transform on the rows and then on the columns.

DDWT can be applied recursively to the low-frequency sub-band only, but in our analysis we used the wavelet packets so each of the four filters was subject to further wavelet decompositions (not only LL).

Altogether in the three iterations $1+4+16=21$ different transformation branches were produced. In one branch the following 12 simple features were calculated: $(e_i, i=1,2,3,4)$ – energies of the sub-images, $(e_i/e_{max}, i=1,2,3,4)$ – maximum energy ratios and $(e_i/\Sigma e_k, k \neq i, i=1,2,3,4)$ - fractional energy ratios, after [18–20]. Energy was defined as a sum of absolute values of the pixels. This procedure was repeated for the three sets A, B, C of different image resolutions yielding $21 \times 12 = 252$ attributes in each single set.

For reasons presented in the Introduction for the skin texture analysis we took RBio3.1 wavelet base. Reverse bi-orthogonal wavelets (wavelet pairs) have the property of perfect reconstruction i.e. if X-image, A-reconstructed image of approximations and D-reconstructed image of details, then $X=A+D$. This property is possible due to two separate filter sets, one for decomposition and another one for image reconstruction. This wavelet is symmetric function and is not orthogonal ($X^2 \neq A^2 + D^2$).

Search for the best subset of features [61] can be used to i) reduce bias (overtraining), ii) reduce computational burden and iii) enhance classification performance. This is usually done because the simplest approach, an exhaustive or random search to evaluate the best feature set, is infeasible or even computationally prohibitive. In this work we do not take advantage of any feature selection or extraction algorithms. This follows the results presented in [62] where widely known feature selection mechanisms (CFS, PCA, GSFS) were applied to melanoma classification problem. It was concluded that although application of feature selection algorithms may reduce the complexity of the classification, the performance is highly dependent upon the classifier. Therefore, it was opted to use all the features and preserve them for some late-selections. It is also the objective of this work – to search for efficient ANN clas-

sifiers in terms of their topologies and/or error minimization approaches and not by the feature selection of the data base.

3.3. Algorithm

Artificial neural network is a black-box approach to the knowledge acquisition and can model complex relationships between inputs and outputs. As a learning method it is a 'standard' and well recognized approach [63–65]. Our objective was to collate this 'simple' learner with ensemble models discussed in [17] for the same (high-performing) wavelet families (Bio/RBio). Since both the topology (hidden layers) and the learning details (the choice for the learning function) can bias the classification performance of melanoma, we examined and compared different setups. Since neural computations can be burden on CPU/memory and can be time-consuming, we arbitrarily focused on one static feedforward back-propagated artificial neural network (ANN) without any recursive or meta (deep) learning extensions. Parallel processing (both CPU- and GPU-based) was implemented.

In our study we used a static feedforward back-propagation artificial neural network (ANN) to classify the dermoscopy images based on the calculated 252 wavelet features. As a preprocessing phase normalization of the the input was done and the labels were fixed to '1' (Malanoma) and '0' (Dysplastic Nevus).

The ANN structure was:

- 252 input nodes that represent the wavelet features,
- a number of hidden nodes grouped into one or two hidden layers ($\mathcal{T}opol$) subject to change ($\mathcal{T}opol \in \{10, 20, [10 - 10], [10 - 20], [20 - 20]\}$),
- 2 output neurons, each one activated on the vectors belonging to one class only (so in the binary classification in mutually exclusive way).

Two setups were analyzed within the scope of the activation functions:

- NN1: hyperbolic tangent sigmoid transfer function (a1) for the hidden layers and linear activation function (a2) for the output layer [65]. This is a generic ANN to model any kind of input to output mapping.
- NN2: hyperbolic tangent sigmoid transfer function (a1) for the hidden layers and also hyperbolic tangent sigmoid transfer function (a1) for the output layer. This ANN should perform more efficiently while classifying inputs according to target classes.

As a performance function ($\mathcal{P}erf$) for NN1 and NN2 we used the 'standard' mean square error (mse) and the cross-entropy (ent). The latter case was an attempt to check how 'information-gain'-based learning objective affect both the classification and computational performance. The binary cross-entropy (ent) is calculated as $-p * \log(p) - (1 - p) * \log(1 - p)$ (p -a priori probability of one class) and it heavily penalizes outputs that are extremely inaccurate, with very little penalty for fairly correct classifications.

For the sake of cross-validation (CV) we randomly divided our data into training (70/100), validation (15/100) and testing (15/100) set. Every epoch all the training samples were presented simultaneously to the network to train it. The validation data was used to evaluate the prediction errors hence to optimize and update the weights in the backpropagation phase. The testing set was used to calculate all the performance coefficients. The algorithm was as follows:

1. Separate data into three sets: training, validation, testing
2. Build a network (NNx, \mathcal{P} erf, \mathcal{T} opol, \mathcal{L} earn)
3. Determine the parameters (stopping conditions: maximum number of epochs, maximum training time; learning conditions: learning rates, etc. Loop (4-6) (average over initial conditions, $\#(\text{CV partitions})=6$)
4. Initialize the weights and biases randomly (but with same seed to compare different setups) Loop (5-6) over epochs until *stop1*
5. Train the network with the train data
6. Compute the network performance on validation data and back propagate the error to update the weights
7. Use the network (compute the network performance on the test data)

As *stop1* standard conditions were applied: the maximum number of epochs reached, maximum time exceeded, performance gradient fallen below e^{-6} .

Several training algorithms and refinements for ANNs have been proposed in the literature to enhance the convergence speed and reduce the generalization error of the network [64–66]. In this work we do not discuss mathematical properties of those algorithms, rather focus on the classification interest when they are run with 'standard' base parameters. The analyzed backpropagation training algorithms (\mathcal{L} earn) were: (shown with initial parameters where applicable)

- L1 Levenberg-Marquardt ($\mu = 0.001$),
- L2 Bayesian Regularization ($\mu = 0.005$),
- L3 Broyden-Fletcher-Goldfarb-Shanno,
- L4 Conjugate Gradient with Powell-Beale restarts,
- L5 Fletcher-Powell Conjugate Gradient,
- L6 Polak-Ribière Conjugate Gradient,
- L7 Gradient Descent ($lrate = 0.01$),
- L8 Gradient Descent with Adaptive Learning ($lrate = 0.01$),
- L9 Gradient Descent with Momentum ($lrate = 0.01, momentum = 0.9$),
- LA Variable Learning Rate Gradient Descent ($lrate = 0.01, momentum = 0.9$),

LB One Step Secant,

LC Resilient backpropagation ($lrate = 0.01, \Delta = 0.07$),

LD Scaled Conjugate Gradient.

Although in our ANN learning we meet the CV paradigm and validate during training, to promote better generalization for the $\mathcal{P}erf$ function 'mse' (mean squared error) we applied also the performance regularization ratio (0.01) which takes into account not only minimizing the error but also the weights and biases (for L2 set to 0).

Due to performance aspects the Matlab library was used for calls to the ANN training algorithms. The code was run on the CUDA-based NVidia GTX 1070 GPU.

In our brief review of the related literature we went through the advances of both the ANN and mobile devices for the melanoma CAD. Unfortunately, there are no methodical studies how the ANN structure (hidden layers) affect the melanoma classification performance. Arbitrary values for both the number of the hidden layers and the number of neurons in the hidden layers are published. Usually 2- and 3-hidden layers are presented. Taking into account the computational burden reported and own attempts on both CPU- and GPU-based parallel computing platforms and, last but not least, performance analyzes and benchmarks for the ARM-based mobile devices, we limited ourselves to up to 2 hidden layers with pretty small (up to 20) neurons on each layer.

4. Results and Discussion

We analyzed two networks: NN1 and NN2 (see Section 3.3.) each thought according to two performance measures: (mse) and (ent). The best CV performance for the discrimination of Melanoma from Dysplastic nevi is reached with NN2(mse) which is full sigmoid-like network. Performance of NN1(mse/ent) and NN2(ent) is slightly worse (by about 8-3%) and more sparse in terms of ($\mathcal{L}earn, \mathcal{T}opol$) coverage. Below we present the results for the best performing NN2(mse).

Our objective was to find the best performing ANN for the classification of melanoma dermoscopy images under the assumption that the algorithm for using the (trained) network and, as a next step, even the network training process take place on an ARM-based mobile device. For that reason only limited topologies were taken into account (starting with 10 hidden neuron on one hidden layer up to 20x20 hidden neurons on two layers). Although we take into account only about 10, 20, 100, 200, and 400 weights this is not a small selection compared to the literature. Close by performance our experiments show coarsely how complex in terms of epochs and time the back-propagation algorithms can perform. Statistics for the 5x13=65 different setups shows the following grouping of results in terms of pairs ($number_of_epochs, setups_finished$): (10, 17), (15, 31), (20, 38), (30, 44), (50, 45), (100, 51). As a necessary condition for further analysis we took the threshold of 20 epochs (median).

In Table 1 we show numerical results of AUC for resolutions A, B and C for the five setups of the hidden layers and for the thirteen different back-propagation algorithms assumed that the number of epochs is below 20. Absence of L7-LA proves that methods based on (variations of) gradient decent converges very slowly (maximum number of epochs even above 1000) and are out of range of mobile hand-held devices and are not feasible as CAD applications.

Table 1. Numerical results for AUC for (from top to bottom) resolutions A, B, C as a function of \mathcal{T} opol and \mathcal{L} earn. Filter: #Epoch<20

\mathcal{T} opol	AUC [10^{-2}]												
	$L1$	$L2$	$L3$	$L4$	$L5$	$L6$	$L7$	$L8$	$L9$	LA	LB	LC	LD
10	92	94	92	94	94	95	-	-	-	-	95	-	93
20	94	6	94	95	-	93	-	-	-	-	-	-	97
10-10	97	6	93	96	96	96	-	-	-	-	97	-	-
10-20	95	93	95	93	92	94	-	-	-	-	95	92	92
20-20	96	61	92	94	91	95	-	-	-	-	-	89	96
10	99	86	97	98	96	98	-	-	-	-	98	98	-
20	100	72	98	98	98	98	-	-	-	-	97	99	97
10-10	98	86	98	99	97	98	-	-	-	-	-	92	-
10-20	97	86	99	98	96	99	-	-	-	-	99	97	-
20-20	95	14	99	98	98	97	-	-	-	-	99	-	-
10	95	-	95	94	94	93	-	-	-	-	94	92	92
20	98	-	93	92	92	93	-	-	-	-	93	93	94
10-10	92	92	91	91	93	93	-	-	-	-	92	89	-
10-20	94	93	91	90	91	92	-	-	-	-	92	94	94
20-20	94	7	92	92	93	93	-	-	-	-	93	93	-

AUC is a good overall measure for how a classifier performs. When the classifier is used however one should pick up the threshold thereby the tradeoff between the sensitivity and specificity. This is called the ROC operating point and its 'manual' optimal selection does depend on the subject of interest. In this study we determined the optimal operating point from probabilistic considerations [67] by the following geometrical procedure. A straight line calculated with $slope = \frac{N}{P} \frac{(p(P|N) - p(N|N))}{(p(N|P) - p(P|P))}$ crossing the point $ROC(0, 1)$ we shifted down and to the right, until it intersected the ROC curve. P and N are the total instance counts in the positive (melanoma) and negative (dysplatic) class and expressions for p denote different probabilities for misclassifications, e.g. $p(N|P)$ denotes the probability of misclassifying a positive class as a negative class etc. In Figure 1 we present those optimal points from trainings with different \mathcal{L} earn and \mathcal{T} opol setups.

First we remark that the above numerical procedure fail for few points (middle field and right upper corner). All the other points are optimal operating points. The strings of 'Lx' stand for the best-performing back-propagation algorithms and the colors represent the structure of the hidden layers. There are no single winners and we should mind that all the points come from the best-converging networks. It

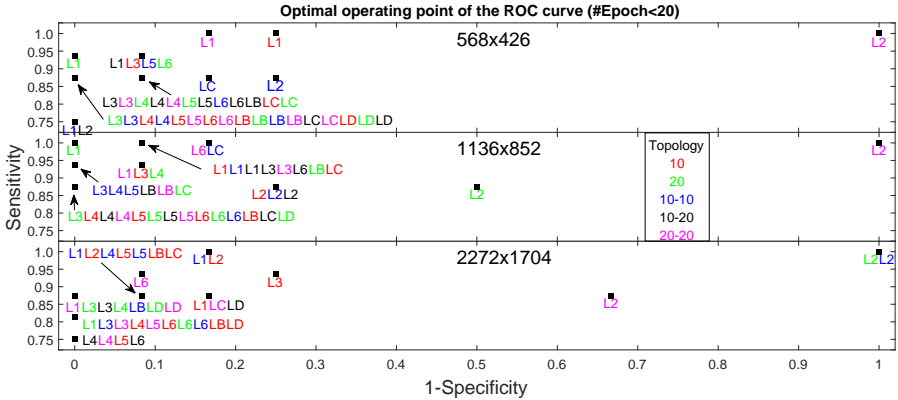


Figure 1. Optimal points (not components of one ROC) from trainings with different $\mathcal{L}earn$ and $\mathcal{T}opol$ setups. Resolutions A, B, C are presented.

seems that different combinations of learning algorithms and topologies reach upper-left ROC region. Taking into account possible system limitations of the developed applications (e.g. image resolution limits, the number of the weights, memory limits of the training cycle) one can optimally select a pair $(\mathcal{L}earn, \mathcal{T}opol)$.

L1 (Levenberg-Marquardt) shows extreme robustness for almost all topologies for resolutions A, B and C.

L6 (Polak-Ribière Conjugate Gradient) seems to be the second best for A, B, C – it is less represented but reaches high AUC values in the upper-left corner.

Figure 1 shows that the magnitude of the the operating points is preserved and stable for B and C and, unexpectedly, outperforms the initial A conditions. An extra attempt was done to degrade the resolution C by a factor of two, but it failed (performance decrease by 30%). A stable overall performance for A, B and C proves resolution invariance of the wavelet features for the original clinical data and the two descendant downgraded image resolutions. This behavior confirms results from [17] where the ensemble learning paradigm was implemented. In our setup a single 'standard' learner (ANN) performs in the same way. It seems that this performance is even better than in [17] where the downgraded resolutions showed still high but slightly worse results. In our experiments with ANN, which is a homogeneous (not complex) model, better sensitivity and specificity is reached for dermoscopy images with smaller resolutions. The maximum reached at B shows the best production resolution of 1136x852 pixels.

Whether this behavior is due to the wavelet features and not the (selected) learning methods it should be further studied both in some future experiments with other learning paradigms and by theoretical considerations about the wavelet families and their properties.

5. References

- [1] Korotkov K., Garcia R., *Computerized analysis of pigmented skin lesions: A review*. Artificial Intelligence in Medicine, 2012, 56(2).
- [2] Masood A., Ali Al-Jumaily A., *Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms*. International Journal of Biomedical Imaging, 2013, 2013(7), pp. 323268.
- [3] 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, 2016.
- [4] Skvara H., Teban L., Fiebigler M., Binder M., Kittler H., *Limitations of dermoscopy in the recognition of melanoma*. Arch. Dermatol., 2005, 141, pp. 155–160.
- [5] Stolz W., Semmelmayr U., Johow K., Burgdorf W.H., *Principles of dermoscopy of pigmented skin lesions*. Seminars in Cutaneous Medicine and Surgery, 2003, 22(1), pp. 9–20.
- [6] Wang S.Q., Hashemi P., *Noninvasive imaging technologies in the diagnosis of melanoma*. Seminars in Cutaneous Medicine and Surgery, 2010, 29(3), pp. 174–184.
- [7] Talbot H., Bischof L., *An overview of the polartechnics solarscan melanoma diagnosis algorithms*, 2003, pp. 33–38.
- [8] Boone M., Suppa M., Dhaenens F., Miyamoto M., Marneffe A., Jemec G., Del Marmol V., Nebosis R., *In vivo assessment of optical properties of melanocytic skin lesions and differentiation of melanoma from non-malignant lesions by high-definition optical coherence tomography*. Arch. Dermatol. Res., 2016, 308(1), pp. 7–20.
- [9] Johr R.H., *Dermatoscopy: Alternative melanocytic algorithms - the abcd rule of dermatoscopy, menzies scoring method, and 7-point checklist*. Clinics in Dermatology, 2002, 20, pp. 240–247.
- [10] Kittler H., Pehamberger H., Wolff K., Binder M., *Follow-up of melanocytic skin lesions with digital epiluminescence microscopy: Patterns of modifications observed in early melanoma, atypical nevi, and common nevi*. J. Am. Acad. Dermatol., 2000, 43(3), pp. 467–476.
- [11] Goodson A.G., Grossman D., *Strategies for early melanoma detection: Approaches to the patient with nevi*. J. Am. Acad. Dermatol., 2009, 60(5), pp. 719–735.
- [12] Chang T., Kuo C.C., *Texture analysis and classification with tree-structured wavelet transform*. IEEE Transactions on Image Processing, 1993, 2(4), pp. 429–44.

- [13] Walvick R.P., Patel K., Patwardhan S.V., Dhawan A.P., *Classification of melanoma using wavelet-transform-based optimal feature set*. In: *Medical Imaging 2004*, International Society for Optics and Photonics, 2004, pp. 944–951.
- [14] Ma L., Staunton R.C., *Analysis of the contour structural irregularity of skin lesions using wavelet decomposition*. *Pattern Recognition*, 2013, 46(1), pp. 98–106.
- [15] Massone C., Hofmann-Wellenhof R., Ahlgrimm-Siess V., Gabler G., Ebner C., Soyer H.P., *Melanoma screening with cellular phones*. *PLoS ONE*, 2007, 2(5), pp. e483.
- [16] MacKinnon N., Vasefi F., Booth N., Farkas D.L., *Melanoma detection using smartphone and multimode hyperspectral imaging*. *SPIE BiOS*, 2016, 9711, pp. 971117–1.
- [17] Surówka G., Ogorzałek M., *On optimal wavelet bases for classification of melanoma images through ensemble learning*. *Artificial Intelligence and Soft Computing*, Lecture Notes in Computer Science, 2016.
- [18] Patwardhan S.V., Dhawan A.P., Relue P.A., *Classification of melanoma using tree structured wavelet transforms*. *Computer Methods and Programs in Biomedicine*, 2003, 72, pp. 223–239.
- [19] Patwardhan S.V., Dai S., Dhawan A.P., *Multi-spectral image analysis and classification of melanoma using fuzzy membership based partitions*. *Computerized Medical Imaging and Graphics*, 2005, 29, pp. 287–296.
- [20] Surówka G., Merkwirth C., Żabińska-Płazak E., Graca A., *Wavelet based classification of skin lesion images*. *Bio Alg. Med Syst.*, 2006, 2(4).
- [21] Surówka G., Grzesiak-Kopec K., *Different learning paradigms for the classification of melanoid skin lesions using wavelets*. *Proc. EMBC07 Lyon*, 2007.
- [22] Surówka G., *Supervised learning of melanocytic skin lesion images*. *Proc. HSI Kraków*, 2008.
- [23] Indira D., Jyotsna Suprya P., *Detection & analysis of skin cancer using wavelet techniques*. *International Journal of Computer Science and Information Technologies*, 2011, 2(5), pp. 1927–1932.
- [24] Fassihi N., Shanbehzadeh J., Sarrafzadeh H., Ghasemi E., *Melanoma diagnosis by the use of wavelet analysis based on morphological operators*. *Proc. Int. Multiconf. Eng. Comp. Sci. I Hong-Kong*, 2011.
- [25] Castillejos H., Ponomaryov V., Nino-de Rivera L., Golikov V., *Wavelet transform fuzzy algorithms for dermoscopic image segmentation*. *Computational and Mathematical Methods in Medicine*, 2012, 578721.
- [26] Ramteke N.S., Jain S.V., *Analysis of skin cancer using fuzzy and wavelet technique - review & proposed new algorithm*. *International Journal of Engineering Trends and Technology*, 2013, 4(6).

- [27] Sugin S., Jegadeesh A., *Segmentation of skin images using fixed grid wavelet networks*. International Journal of Engineering Research & Technology, 2014, 3(4).
- [28] Rajarathinam A., Arivazhagan A., *Timely efficient automated system by segmentation using wavelet transform*. International Journal of Science, Engineering and Technology Research, 2015, 4(8).
- [29] Sikorski J., *Identification of malignant melanoma by wavelet analysis*. Proceedings of Student/Faculty Research Day, CSIS, Pace University, 2004.
- [30] Aswin R., Jaleel J.A., Salim S., *Implementation of ann classifier using matlab for skin cancer detection*. ICMiC13, 2013, pp. 87–94.
- [31] Mahmoud M.K.A., Al-Jumaily A., Takruri M., *The automatic identification od melanoma by wavelet and curvelet analysis: Study based on neural network classification*. 11th International Conference on Hybrid Intelligent Systems, 2011, pp. 680–685.
- [32] Clawson K.M., Morrow P., Scotney B., McKenna J., Dolan O., *Analysis of pigmented skin lesion border irregularity using the harmonic wavelet transform*. Machine Vision and Image Processing Conf., 2009.
- [33] Ercal F., Chawla A., Stoecker W.V., Lee H.C., Moss R.H., *Neural network diagnosis of malignant melanoma from color images*. IEEE Trans. Biomed. Eng., 1994, 41(9).
- [34] Dreiseitl S., Ohno-Machado L., Kittler H., Vinterbo S., Billhardt H., Binder M., *A comparison of machine learning methods for the diagnosis of pigmented skin lesions*. Journal of Biomedical Informatics, 2001, 34, pp. 28–36.
- [35] Rubegni P., Burroni M., Perotti R., Fimiani M., Andreassi L., Cevenini G., Dell'Eva G., Barbini P., *Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: A retrospective study*. J. Invest. Dermatol., 2002, 119, pp. 471–474.
- [36] Hoffmann K., Gambichler T., Rick A., Kreutz M., Anschuetz M., Grünendick T., Orlikov A., Gehlen S., Perotti R., Andreassi L., et al., *Diagnostic and neural analysis of skin cancer (danaos). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy*. British Journal of Dermatology, 2003, 149, pp. 801–809.
- [37] Rajab M., Woolfson M., Morgan S., *Application of region-based segmentation and neural network edge detection to skin lesions*. Computerized Medical Imaging and Graphics, 2004, 28, pp. 61–68.
- [38] Maglogiannis I., Pavlopoulos S., Koutsouris D., *An integrated computer supported acquisition, handling, and characterization system for pigmented skin lesions in dermatological images*. IEEE Trans. Inf. Techn. Biomed., 2005, 9(1).

- [39] Zagrouba E., Barhoumi W., *An accelerated system for melanoma diagnosis based on subset feature selection*. Journal of Computing and Information Technology, 2005, 13(1), pp. 69–82.
- [40] Iyatomi H., Oka H., Celebi M.E., Hashimoto M., Hagiwara M., Tanaka M., Ogawa K., *An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm*. Computerized Medical Imaging and Graphics, 2008, 32(7), pp. 566–579.
- [41] Schaefer G., Rajab M.I. Celebi M.E., Iyatomi,H., *Skin lesion segmentation using cooperative neural network edge detection and colour normalization*. Inf. Techn. and Applic. Biomed., 2009.
- [42] Lau H.T., Al-Jumaily A., *Automatically early detection of skin cancer: Study based on neural network classification*. IEEE International Conference of Soft Computing and Pattern Recognition, 2009, pp. 375–380.
- [43] Vennila G.S., Suresh L.P., Shunmuganathan K., *Dermoscopic image segmentation and classification using machine learning algorithms*. American Journal of Applied Sciences, 2012, 8(11).
- [44] Mhaske H., Phalke D., *Melanoma skin cancer detection and classification based on supervised and unsupervised learning*. International conference on Circuits Controls and Communications, 2013, pp. 1–5.
- [45] Jaleel J.A., Salim S., Aswin R., *Computer aided detection of skin cancer*. International Conference on Circuits, Power and Computing Technologies, 2013.
- [46] Elgamal M., *Automatic skin cancer images classification*. International Journal of Advanced Computer Science and Applications, 2013, 4(3).
- [47] Silva C.S., Marcal A.R., *Colour-based dermoscopy classification of cutaneous lesions: An alternative approach*. DOI: 10.1080/21681163.2013.803683, 2013.
- [48] Achakanalli S., Sadashivappa G., *Skin cancer detection and diagnosis using image processing and implementation using neural networks and abcd parameters*, 2014.
- [49] Alasadi A.H., ALSafy B.M., *Early detection and classification of melanoma skin cancer*. Int. J. Information Technology and Computer Science, 2015, 12, pp. 67–74.
- [50] Torre E.L., Caputo B., Tommasi T., *Learning methods for melanoma recognition*. International Journal of Imaging Systems and Technology, 2010, 20(4), pp. 316–322.
- [51] Ruiz D., Berenguer V., Soriano A., SáNchez B., *A decision support system for the diagnosis of melanoma: A comparative approach*. Expert Systems with Applications, 2011, 38, pp. 15217–15223.
- [52] Maglogiannis I., Kosmopoulos D.I., *Computational vision systems for the detection of malignant melanoma*. Oncology Reports, 2006, 15(Spec no. 4), pp. 1027–32.

- [53] Rajpara S., Botello A., Townend J., Ormerod A., *Systematic review of dermoscopy and digital dermoscopy/artificial intelligence for the diagnosis of melanoma*. British Journal of Dermatology, 2009, 161(3), pp. 591–604.
- [54] Sathiya S.B., Kumar S., Prabin A., *A survey on recent computer-aided diagnosis of melanoma*. International Conference on Control Instrumentation Communication and Computational Technologies, 2014, pp. 1387–1392.
- [55] Abedini M., Chen Q., Codella N.C., Garnavi R., Sun X., *Accurate and scalable system for automatic detection of malignant melanoma*. In book: Dermoscopy Image Analysis, 2015, pp. 293–343.
- [56] Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P., *Smote: Synthetic minority over-sampling technique*. Journal of Artificial Intelligence Research, 2002, 16, pp. 321–357.
- [57] Stefanowski J., Wilk S., *Selective pre-processing of imbalanced data for improving classification performance*. Data Warehousing and Knowledge Discovery, 2008, pp. 283–292.
- [58] Rastgoo M., Lemaitre G., Massich J., Morel O., Marzani F., Garcia R., Meriaudeau, F., *Tackling the problem of data imbalancing for melanoma classification*. BIOSTEC - 3rd International Conference on BIOIMAGING, 2016.
- [59] Mallat S.G., *A theory for multiresolution signal decomposition: The wavelet representation*. IEEE Transactions on pattern analysis and machine intelligence, 1989, 11(7).
- [60] Daubechies I., *Ten lectures on wavelets*. CBMS SIAM, 1994, 61.
- [61] Tang J., Alelyani S., Liu H., *Feature selection for classification: A review*. CRC Press, 2014, 37.
- [62] Maglogiannis I., Doukas C.N., *Overview of advanced computer vision systems for skin lesions characterization*. IEEE Trans. Inf. Techn. Biomed., 2009, 13(5), pp. 721–733.
- [63] Michie D., Spiegelhalter D.J., Taylor C.C., *Machine Learning, Neural and Statistical Classification*. Prentice Hall, 1994.
- [64] Haykin S., *Neural Networks: A Comprehensive Foundation*. 2 edn. Prentice Hall, 2004 ISBN 0-13-273350-1.
- [65] Demuth H.B., Beale M.H., De Jess O., Hagan M.T., *Neural Network Design*. 2 edn., 2004 ISBN-10: 0-9717321-1-6, ISBN-13: 978-0-9717321-1-7.
- [66] Battiti R., *First- and second-order methods for learning: Between steepest descent and newton's method*. Neural Computation, 1992, 4(2).
- [67] Hajian-Tilaki K., *Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation*. Caspian J. Intern. Med., 2013, 4(2), pp. 627–635.