Melanoma Thickness Prediction Based on Convolutional Neural Network with VGG-19 Model Transfer Learning

Joanna Jaworek-Korjakowska, Pawel Kleczek, Marek Gorgon
AGH University of Science and Technology, Krakow, Poland
Department of Automatic Control and Robotics
{jaworek, pkleczek, mago}@agh.edu.pl

Abstract

Over the past two decades, malignant melanoma incidence rate has dramatically risen but melanoma mortality has only recently stabilized. Due to its propensity to metastasize and lack of effective therapies for most patients with advanced disease, early detection of melanoma is a clinical imperative. Thickness is one of the most important factor in melanoma prognosis and it is used to establish the size of the surgical margin, as well as to select patients for sentinel lymph node biopsy. However, little work has concentrated on the evaluation of melanoma thickness both from the clinical as well as computer-aided diagnostic side. To address this problem, we propose an effective computer-vision based machine learning tool that can perform the preoperative evaluation of melanoma thickness. The novelty of our approach is that we directly predict the thickness of the skin lesion into one of three classes: less than 0.75 mm, 0.76-1.5 mm, and greater that 1.5 mm. In this study, we use transfer learning of the pre-trained, adapted to our application VGG-19 convolutional neural network (CNN) with an adjusted densely-connected classifier. Due to the limited data we investigate the transfer learning method where we apply knowledge from model trained on a different task. Our database contains 244 dermoscopy images. Experiments confirm the developed algorithms ability to classify skin lesion thickness with 87.2% overall accuracy what is a state-of-the-art result in melanoma thickness prediction.

1. Introduction

Malignant melanoma (Latin: *melanoma malignum*) is a deadly type of skin cancer which in the United States is the fifth most common cancer among men and women [14]. Over the past two decades, melanoma incidence has dramatically risen but melanoma mortality has only recently stabilized [21]. According to the World Health Organization about 132,000 new cases of malignant melanoma are diagnosed worldwide each year. In some parts of the world, especially in New Zealand and Australia, melanoma is becoming more common every year and has more than doubled in the past 30 years. Estimated rates of melanoma are increasing in 2019. It is prognosed that incidance rate will be 5.6% higher than in 2018 [14].

One of the main goals in prevention of malignant melanoma is awareness, early diagnosis and surgical excision. The introduction of dermoscopy has improved the accuracy of melanoma diagnosis. Digital dermoscopy is currently the most used technology, although novel methods, such as confocal microscopy, show promising result.

Malignant melanoma originates in pigment producing cells called melanocytes and is fast-growing and highly malignant tumor often spreading to nearby lymph nodes, lungs, and brain. It is well accepted that only early detection can reduce mortality, since the prognosis of patients with melanoma depends on the thickness of the tumor at the time of surgical treatment [3].

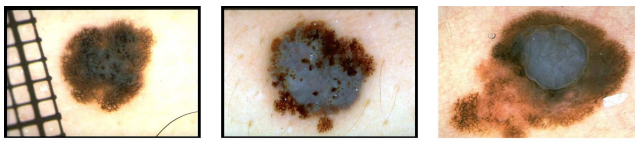

Figure 1. Dermoscopy images of various melanoma thicknesses: a) thin melanoma, b) intermediate melanoma, c) thick melanoma [3].

We wish to stress that the main goal of the proposed solution is to classify melanocytic lesions into three classes indicating the thickness of the diagnosed skin lesion:

- **thin melanoma** (T-M) including melanoma in situ (T-M-IS) with thickness less than 0.75 mm,
- **intermediate melanoma** (I-M) with thickness between 0.76-1.5 mm,

- **thick melanoma** (TK-M) with thickness greater that 1.5 mm.

Figure 1 presents examples of different stages of melanoma with varied depths.

Artificial intelligence research has been around for more than half a century but in recent years a huge progress is observed in widely understood machine learning. Advanced statistical techniques, known as deep learning models, have been exploited with impressive results. Deep learning is currently one of the most popular method from the group of artificial intelligence methods, used for image segmentation, classification and analysis. Deep Neural Networks (DNN) are state-of-the-art algorithms for challenges that haven't been solved yet. Progress in the hardware, software, algorithms and availability of huge datasets allow employment of high accuracy image recognition systems. Convolutional neural networks (CNN) and its variations have the key advantage of automatic extraction and learning of image features. Deep learning in general and convolutional neural networks in particular have been used in variety of pattern recognition problems like retinal vessel segmentation, lung area detection, or breast cancer classification. A systematic review on the use of deep neural networks, especially CNN, to analyze skin changes can be found in [7].

In this paper we present a new and one of the first approaches to the preoperative evaluation of melanoma thickness. We examine the possibility to use pre-trained VGG-19 convolutional neural network, which is a competition-winning 19-layer model, and CNN transfer learning from non-medical to medical image domains. We firstly localize the lesion and crop the region of interest, secondly due to the limited and imbalanced data we generate synthetic samples employing the Synthetic Minority Over-sampling Technique (SMOTE). Subsequently, we use the VGG-19 pre-trained ConvNet architecture which is a common and highly effective approach to deep learning on small image datasets. The network architecture is adapted to the specification of our problem including the densely-connected classifier layer.

The novelty of this work can be summerized as:

- we present a deep learning based solution for the preoperative melanoma thickness prediction into three depth classes,
- we propose a convolutional neural network architecture with transfer learning from the VGG-19 pre-trained model and an adjusted densely-connected classifier,
- our method enables the adaptation of the VGG-19 model to the melanoma thickness prediction with a limited data amount based on data augmentation and transfer learning method.

This paper is organized in four sections as follows. Section 1 *Introduction* presents the skin cancer awareness and covers background information on clinical importance of melanoma thickness prediction, deep learning methods enhanced with transfer learning, and describes the motivation of the undertaken research and state-of-the-art. Section 2 *Melanoma thickness prediction method* shows in detail the methodology used in this research including the CNN architecture, VGG-19 pre-trained model, training and classification stages. Section 3 *Results* presents the used database as well as the conducted tests and results with comparison to related works. Section 4 *Conclusion* exposes the conclusions and suggests new lines of research.

1.1. Clinical importance and motivation

Thickness is the most important factor in melanoma prognosis; therefore, examining the most recent trends in thickness and survival may provide key insights [21]. Breslow index [6] is a method to measure the tumor depth of melanoma by means of pathological examination after biopsy of the suspected lesion. It is measured vertically from the top of the granular layer of the epidermis down to its deepest point of invasion (Fig. 2).

Dermoscopic criteria for the *in vivo* detection of the various phases of melanoma progression as well as tumor depth have been described in [1, 2]. In particular, a significant association was found between pigment network, blue-whitish veil, and vascular pattern on one side and melanoma thickness on the other. Tumor depth is one of the cornerstones of the current AJCC TNM staging of malignant melanoma [10]. A large study validated the importance of tumor depth as one of the three most important prognostic factors in melanoma [4]. According to Breslow the measurement of melanoma thickness is a consistent and well-aimed predictor for the prognosis of further treatment and disease development [6]. The assessment of melanoma thickness is not only used for the medical diagnosis but furthermore to establish the size of the surgical margin [17] and to select patients for sentinel lymph node biopsy [2, 3]. In the United States, the Surveillance, Epidemiology, and End Results (SEER) Program is the authoritative source for population-based data on melanoma incidence and mortality. In [22] it has been reported that of 182184 cases, 24329 (13%) have unknown thickness. Moreover, unknown thickness cases had a significantly increased risk of death due to melanoma than known thickness cases with an increasing trend over time.

The possibility of preoperative evaluation of melanoma thickness has many advantages including: (i) better organization of surgical and diagnostic priorities based on distinction of lesions at high and low risk of progression, (ii) excision with sufficient surgical margins at the first operation, avoiding a second more radical time and cost consum-

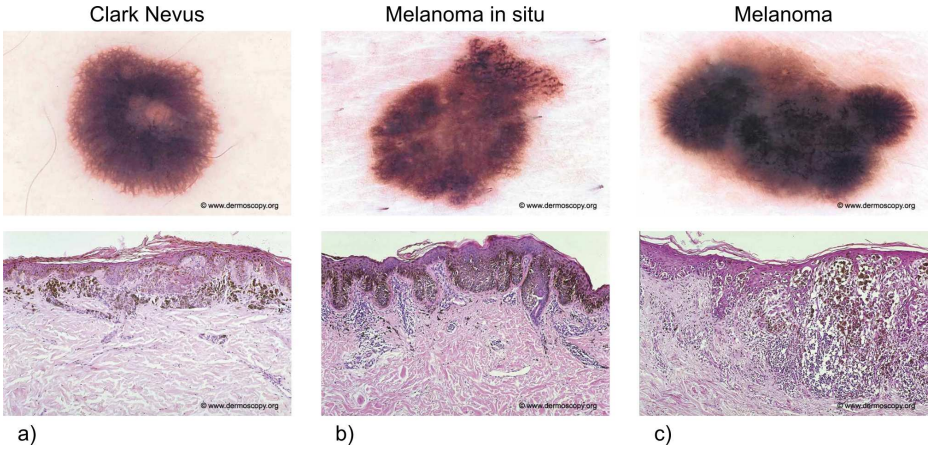

Figure 2. Dermoscopy and hematoxylin-eosin-stained histopathology images corresponding to different skin lesion types. Specific underlying histopathologic correlates as well as the increasing thickness can be observed: a) Clark nevus - benign lesion, b) melanoma in situ - stage 0, c) melanoma - malignant lesion [3].

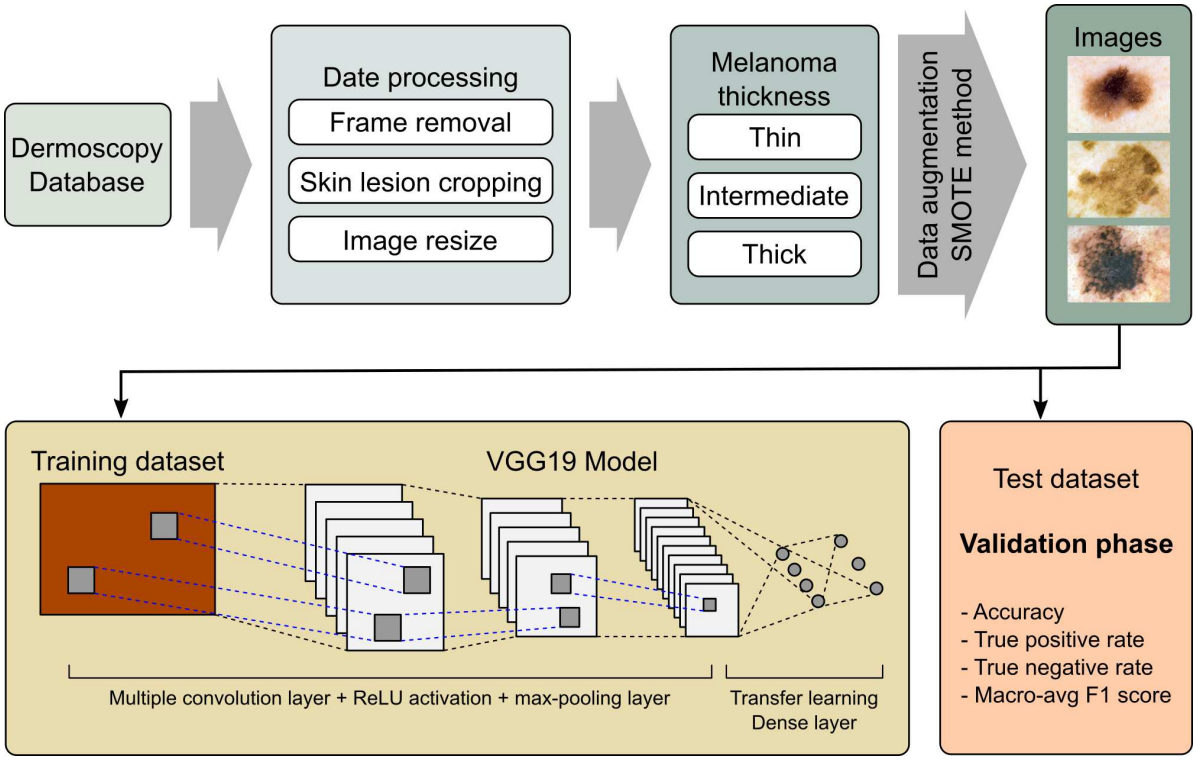

Figure 3. Melanoma thickness prediction diagram including following steps: data processing, data augmentation, training and test data division, training process of the classification layer, and validation phase.

ing operation, (iii) excision and sentinel lymph node biopsy (if needed) in a single operation, saving time and costs [18].

Moreover, clinicians suggest that the determination of Breslow index by non-invasive techniques like dermoscopy or ultrasonography, would be a great advance in every day clinical management of melanoma.

1.2. Related works

Due to the difficulty and subjectivity of human interpretation, the computerized image analysis techniques have become an important tool in the interpretation of dermoscopic images. Different groups of researchers have focused on the extraction and classification of significant fea-

tures to provide distinction between benign and malignant skin lesions; however, when dealing with melanoma thickness only few works have been published in the scientific literature. In 2010 Rubegni *et al*. presented a method for the classification of tumors into two classes: thin and thick melanomas [18]. The authors performed the classification process with a commercial software using 49 features including color, geometry, and texture. The melanoma depth has been assessed for 141 images from a private database and achieved overall accuracy 86.5%. Specifically, 97 of 108 thin melanomas (specificity 89.8%) and 25 of 33 thick melanomas (sensitivity 75.7%) were correctly classified [18].

In 2016 Aurora Sáez *et al*. described two classification schemas: a binary classification in which melanomas are classified into thin or thick, and a three-class scheme (thin, intermediate, and thick) [20]. Extracted features were based on the clinical findings indicating the tumor depth. The proposed methodology achieved 77.6% accuracy for the binary problem, and 68.4% accuracy for the multiclass classification, respectively. As the database contained 250 dermoscopy images from the same research dataset for skin image analysis as we are applying, a direct comparison of the statistical analysis will be presented in Section 3.

Since there is a vast amount of both computer-aided preoperative melanoma thickness prediction algorithms and literature on clinical evaluation of melanoma thickness we take advantage of the most advanced and latest methods in deep learning to perform an accurate classification of the melanoma thickness.

2. Melanoma thickness prediction method

The strength of deep networks is to learn the multiple layers of concept representation, which corresponds to miscellaneous levels of abstraction. For visual datasets, like dermoscopy images, the low levels of abstraction might describe edges and dark local structures in the image, while high layers in the network refer to object parts and even the category of the object viewed (global features). The effectiveness of ConvNets has been proven in many tasks of computer vision due to their powerful feature representation, especially for segmentation and classification tasks.

In this work we tested a convolutional neural network architecture with transfer learning from VGG-19 pre-trained model and an adjusted densely-connected classifier for melanoma thickness prediction. The algorithm flowchart is presented in Figure 3 and described in detail in the following parts. The preprocessing step is described shortly, as it can be easly found in literature, while the network architecture, learning transfer, and training phase are presented in detail. The preprocessing and image extraction have been implemented in Matlab 2017a while the neural architecture

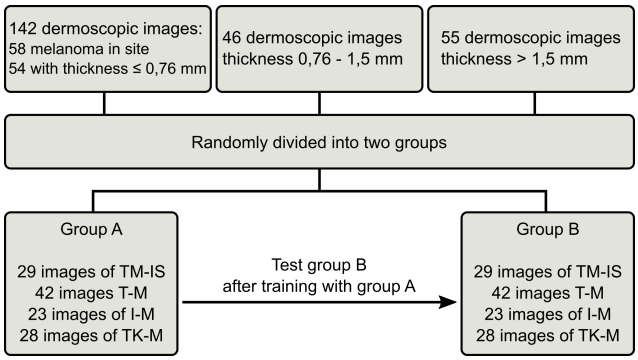

Figure 4. The diagram presents the specificity of individual classes in the database which are divided into teaching and test parts.

and classification process have been performed in Python using neural-network library Keras that provides an Applications interface for loading and using pre-trained models including VGG-19.

The dataset comprises three categories of dermoscopy images including thin, intermediate and thick melanomas. The dataset has been described in detail in Section 3.1, however due to the limited data, imbalance problem, learning and classification process, the dataset split has been presented in Figure 4. 244 dermoscopy images were splited into two subsets (group A and B); half of the total image dataset was selected for training and the rest for testing.

2.1. VGG-19 Convolutional neural network architecture

Building a convolutional neural network from scratch has it pros and cons but above all a very large database is required. A very promising approach is to use pre-trained models over large datasets like VGG-16, VGG-19, ResNet, and so on. The concept of transfer learning helps us to use the existing models for our tasks. There are several strategies of performing transfer learning by reusing the same model to extract feature representations from new images. Due to our limited dataset we have chosen the pre-trained model as feature extractor which means that only the fully connected classifier will be trained. In this work, we use

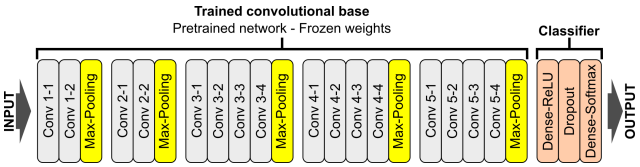

Figure 5. Schematic overview of the personalized VGG-19 network architecture with description of layers.

the pre-trained model VGG-19, proposed in 2014 by K. Simonyan and A. Zisserman from the University of Oxford

in the paper [23]. The VGG-19 CNN architecture is reported to achieve high accuracies for image processing large datasets such as ImageNet. VGG-19 model has roughly 143 million parameters, where the parameters are learned from the ImageNet dataset containing 1.2 million general object images of 1,000 different object categories for training [25]. The VGG-19 contains 19 trainable layers including convolutional and fully connected layers as well as max pooling, and dropout. In our solution we use the trained convolution base with a personalized classification part including densely-connected classifier and dropout layer for regularization. The modified version of the VGG-19 model is presented in Figure 5 and 6.

Convolutional layers apply a convolution operation over the image (feature map) and perform operation at each point, passing the result to the next layer [12]. Filters belonging to the convolutional layer are trainable feature extractors typically of 3×3 size. Each of the convolutional layers stack is followed by a rectified linear unit (ReLU) activation function, and max-pooling operation. The ReLU is currently the most popular nonlinear activation function defined as the positive part of its argument where x is the input to a neuron:

$$f(x) = max(0, x) \tag{1}$$

The ReLU function compared to sigmoid function is computationally efficient, shows better convergence performance, and solves the vanishing gradient problem.

After the ReLU activation function a downsampling max-pooling layer is applied. We basically take a filter of 2×2 size and a stride of the same length. The output is the maximum number in every subregion that the filter convolves around [16].

A series of convolutional layers (conv1, conv2, conv3, conv4, and conv5) were followed by the classification part including densely-connected classifier and dropout layer. A dense layer is a fully connected layer, where each neuron receives an input from all the neurons present in the previous layer. A densely connected layer provides learning features from all the features of the previous layer, whereas a convolutional layer relies on consistent features with a small repetitive field. For the densely-connected layer the activation function has to be specified [13].

The Dropout layer forces the network to be redundant by droping out a random set of activations in that layer by setting them to zero. The neurons are randomly removed during the training phase (p - dropout rate) which helps alleviate the overfitting problem. This layer is only used during training, and not during test time [24].

2.2. Database preprocessing and augmentation

Dermoscopic images are inhomogeneous and complex and furthermore include extraneous artifacts, such as skin lines, air bubbles, and hairs, which appear in virtually every image. The preprocessing stage consists of three steps. The first step is the removal of black frame that is introduced during the digitization process [8, 15]. With the binary masks which for most cases we obtained through the publicly available ISIC Archive we are cropping the skin lesion with a tight margin (5 px). After applying the aforementioned transformations, we crop the original image to match the bounding box and resize the it to 400×400 px -the same input size as required by the classification network. In the second step we handle the class imbalance by using Synthetic Minority Oversampling TEchnique (SMOTE) to generate synthetic samples [9]. In SMOTE method, new examples are created by averaging a few representatives of the same minority classes which are nearest neighbors in the feature space. We have performed the upsampling by adding generated images to the underrepresented classes to make the quantity of dermoscopy images in each class equal.

2.3. Training and classification

The proposed and described neural network takes 400×400 px preprocessed RGB image as an input and generates the output that indicates the melanoma thickness - thin, intermediate or thick. As we use the pre-trained VGG-19 model we only need to train the classification part containing dense and dropout layers (Fig. 6). We run the VGG-19 model with our training dataset and take output vectors from the last layer as the training input for the classification part.

As we are dealing with multi-class classification problem the categorical cross-entropy loss function, also called Softmax Loss, has been applied [19]. It consists of Softmax activation and a Cross-Entropy loss that calculates the error rate between the predicted value and the original value. The Categorical Cross-Entropy is given by:

$$CE = -\sum_{i}^{C} t_i \log\left(f(s)_i\right) \tag{2}$$

where t_i is the ground truth and $f(s)_i$ is the standard Softmax function.
For a given class s_i, the Softmax function is computed as [5]:

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \tag{3}$$

where s_j are the scores inferred by the net for each class in C.

Each sample can belong to one of C classes. The CNN has C output neurons that are gathered in a vector s (Scores). The ground truth vector t is a one-hot vector with a positive class and $C - 1$ negative classes (zeros), so

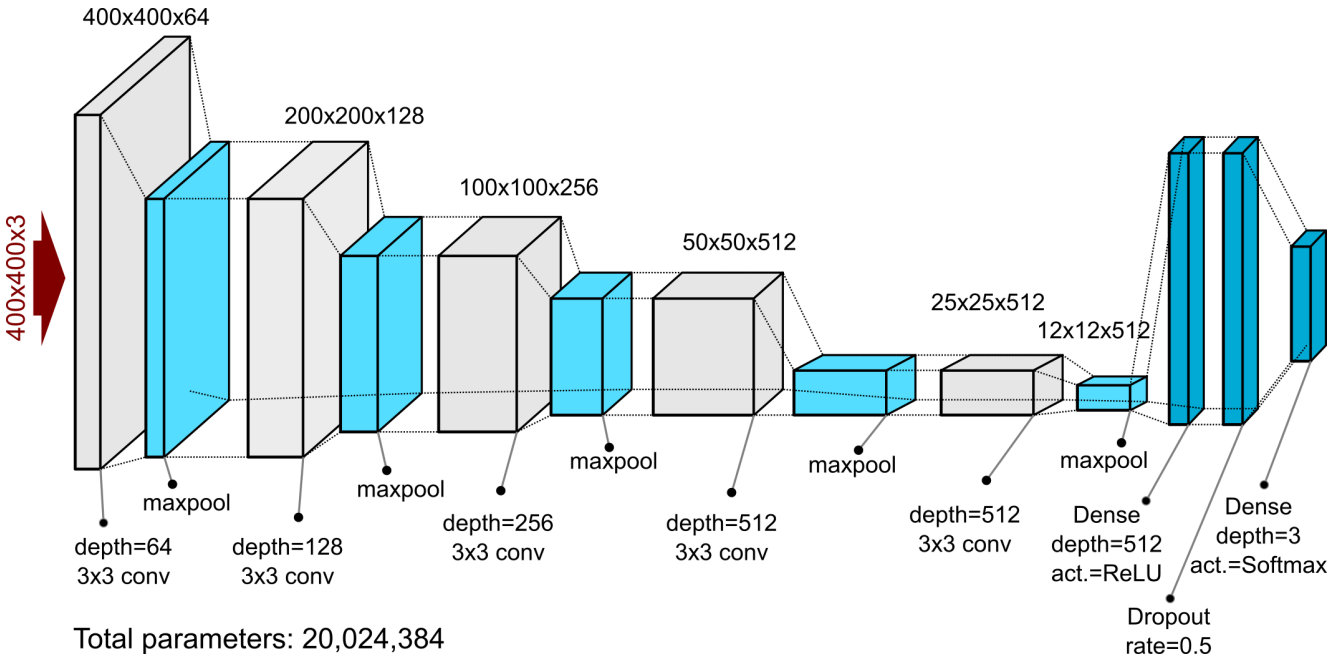

Figure 6. Illustration of the personalized network architecture of VGG-19 model.

we can write:

$$CE = -log\left(\frac{e^{s_p}}{\sum_j^C e^{s_j}}\right) \tag{4}$$

where S_p is the score for the positive class.

We have chosen the Adam optimization algorithm that is an extension of the stochastic gradient descent (SGD) that has recently seen broader adoption for deep learning applications in computer vision and natural language processing. The aim of optimization algorithms is to find optimum weights, minimize error and maximize accuracy. The optimizer is responsible for updating the weights of the neurons via backpropagation. It calculates the derivative of the loss function with respect to each weight and subtracts it from the weight. Adam optimization technique has many advantages including: computationally efficient, invariant to diagonal rescale of the gradient, appropriate for problems with very noisy gradients. Furthermore, the step size of Adam in each iteration is approximately bounded the step size hyperparameter and is invariant to the magnitude of the gradient.

3. Results

3.1. Database

The proposed melanoma thickness classification method has been tested on colour dermoscopic images from the widely used Interactive Atlas of Dermoscopy [3]. Images for this atlas have been provided by two university hospitals (University of Naples, Italy, and University of Graz, Austria) and stored on a CD-ROM in the JPEG format. The documentation of each dermoscopic image was performed using a Dermaphot apparatus (Heine, Optotechnik, Herrsching, Germany) and a photo camera (Nikon F3) mounted on a stereomicroscope (Wild M650, Heerbrugg AG, Switzerland) in order to produce digitized ELM images of skin lesions. All the images have been assessed manually by a dermoscopic expert with an extensive clinical experience. Furthermore, all of the descriptions of skin cases were based on histopathological examination of the biopsy material. The melanoma thickness has been determined during the histopathological examination as well as with the Breslow's staging method. The database included 244 cases belonging to one of three classes: thin melanoma (including melanoma in situ) - 142 images, intermediate melanoma - 46 images, and thick melanoma - 56 images. Dermoscopy colour images have different resolutions, ranging from 0.033 to 0.5mm/pixels. Figures 2 and 4 present samples from the applied database.

3.2. Statistical analysis

We analyze the discrepancies in classification, which are differences between the classification carried out by the classifier and the actual classification (ground truth) to measure the performance of the proposed solution. Although the initial problem is a multiclass classification problem (with three classes), it can be decomposed into four binary classification problems: whether an instance belongs to the

given class or not. For a binary classification problem, those indicators are given by [11]:

$$Accuracy = \frac{1}{Te} \sum_{x \in Te} I[\hat{c}(x) = c(x)] \tag{5}$$

$$TPR = \frac{\sum_{x \in Te} I[\hat{c}(x) = c(x) = \oplus]}{\sum_{x \in Te} I[c(x) = \oplus]} \tag{6}$$

$$TNR = \frac{\sum_{x \in Te} I[\hat{c}(x) = c(x) = \ominus]}{\sum_{x \in Te} I[c(x) = \ominus]} \tag{7}$$

where $\hat{c}(x)$ refers to the actual class, $c(x)$ as stated in the ground truth, Te is the test set, and the function $I[\cdot]$ denotes the indicator function. Positive and negative classes are denoted by $\oplus$ and $\ominus$, respectively.
To evaluate multi-class classification problem a commonly used measure is the macro-averaged F1 (F-measure) score which is defined as:

$$F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \tag{8}$$

where $\beta = 1$. We specify precision P_{macro} and recall R_{macro} as follows:

$$P_{macro} = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FP_i}, \tag{9}$$

$$R_{macro} = \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FN_i} \tag{10}$$

Classes	TPR [%]	TNR [%]	ACC [%]
Thin	84.5	90.9	86.9
Intermediate	78.3	87.23	85.5
Thick	78.6	92.86	89.3
Average	80.5	90.3	87.2

Table 1. The performance of the melanoma thickness classification model.

From the results presented in Table 1, we can conclude as follows: the average accuracy is 87.5%, the average TPR is 90.3%, and the average TNR 87.2%, respectively. Figure 7 shows the training and validation accuracy with reference to the number of epoch. Furthermore, the average F1-score is 83.4% for classifying the melanoma thickness into three different categories.

3.3. Comparison to other related studies

Table 2 lists the state-of-the-art of previous studies described in Section 1.2. However, only in [20] Authors present a multi-class approach classifying melanomas into three groups like in our work. Furthermore, the database

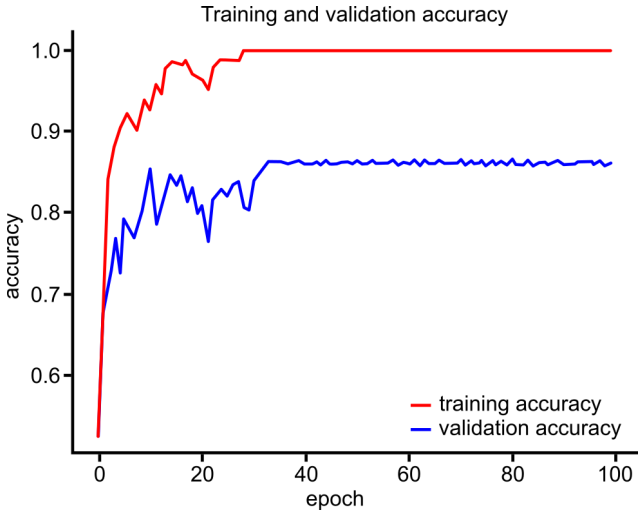

Figure 7. Training and validation accuracy.

Problem	Binary [ACC %]	Multi-class [ACC %]
Rubegni *et al.* [18]	86.5	—
Sáez *et al.* [20]	77.6	68.4
Our method	**—**	**87.2**

Table 2. Comparison with other melanoma thickness prediction methods.

used in the compared work are derived from the same source which allows us to confront the results of our classification. A study described by Rubegni *et al.* [18] presents a binary classification distinguishing between thin and thick melanomas. A direct comparison to this study is not possible due to the different classification task as well as the dataset used in that study is private.

Comparing the average accuracy we obtained higher results than presented in the literature.

4. Conclusion

Although non-invasive and automated diagnostic techniques have been introduced for the segmentation, classification, and early detection of melanoma, there is still a lack of solutions to predict the melanoma thickness. Our results demonstrate the feasibility of classifying the melanoma depth into two main classes while using deep learning approaches based on non-medical learning. The achieved results are much better than results conducted so far. Our method allowed classifying melanoma thickness with ACC= 87.2%, whereas for TPR is within the range 78-84% and TNR 87-93%, respectively. Deep learning methods have not been tested for melanoma thickness prediction for our knowledge, especially not with non-medical archive learning. This is a first-of-its-kind experiment that shows that deep learning with ImageNet training may be sufficient for general medical image depth prediction tasks.

4.1. Future works

Starting from the described framework, further research efforts will be firstly addressed to compare and integrate other promising pre-trained models not only with transfer learning method but also with fine-tuning of feature vectors. In order to improve the CNN neural network through training the parameters from scratch a larger database is required. Future research will concentrate on the possibility of melanoma thickness prediction as a regression challenge.

Acknowledgment: This scientific work was financially supported by AGH University of Science and Technology Status Funds on Decision No. 16.16.120.773.

References

[1] Giuseppe Argenziano, Gabriella Fabbrocini, P. M. Carli, Vincenzo de Giorgi, and Mario R. Delfino. Epiluminescence microscopy: criteria of cutaneous melanoma progression. *Journal of the American Academy of Dermatology*, 37 1:68–74, 1997. 2

[2] Giuseppe Argenziano, Gabriella Fabbrocini, P. M. Carli, Vincenzo de Giorgi, and Mario R. Delfino. Clinical and dermatoscopic criteria for the preoperative evaluation of cutaneous melanoma thickness. *Journal of the American Academy of Dermatology*, 40 1:61–8, 1999. 2

[3] G. Argenziano, P. Soyer, Vincenzo De Giorgio, Domenico Piccolo, Paolo Carli, Mario Delfino, Angela Ferrari, Rainer Hofmann-Wellenhof, Daniela Massi, Giampero Mazzocchetti, Massimiliano Scalvenzi, and Ingrid H. Wolf. *Interactive Atlas of Dermoscopy*. Edra Medical Publishing and New Media, 2000. Book and CD/Web Resource. 1, 2, 3, 6

[4] Charles Balch, Seng Jaw Soong, Jeffrey E Gershenwald, John F Thompson, Dr. Douglas Reintgen, Natale Cascinelli, Marshall M. Urist, Kelly Mcmasters, Merrick I. Ross, John Munn Kirkwood, Michael B. Atkins, John A. Thompson, Daniel G. Coit, David Martin Byrd, Renee L. Desmond, Yu Zhang, Ping Yang Liu, Gary H. Lyman, and Alessandro Morabito. Prognostic factors analysis of 17,600 melanoma patients: validation of the american joint committee on cancer melanoma staging system. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 19 16:3622–34, 2001. 2

[5] Christopher M. Bishop and Nasser M. Nasrabadi. Pattern recognition and machine learning. *J. Electronic Imaging*, 16:049901, 2007. 5

[6] Alexander Breslow. Tumor thickness, level of invasion and node dissection in stage i cutaneous melanoma. *Annals of surgery*, 182 5:572–5, 1975. 2

[7] Titus Josef Brinker, Achim Hekler, Jochen Sven Utikal, Niels Grabe, Dirk Schadendorf, Joachim Klode, Carola Berking, Theresa Steeb, Alexander H Enk, and Christof von Kalle. Skin cancer classification using convolutional neural networks: Systematic review. In *Journal of medical Internet research*, 2018. 2

[8] M. Emre Celebi, Hitoshi Iyatomi, Gerald Schaefer, and William V. Stoecker. Lesion border detection in dermoscopy images. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 33 2:148–53, 2009. 5

[9] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority oversampling technique. *J. Artif. Intell. Res.*, 16:321–357, 2002. 5

[10] Stephen B Edge and PhD Carolyn C. Compton. The american joint committee on cancer: the 7th edition of the ajcc cancer staging manual and the future of tnm. *Annals of Surgical Oncology*, 17:1471–1474, 2010. 2

[11] Peter Flach. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge University Press, New York, NY, USA, 2012. 7

[12] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2414–2423, 2016. 5

[13] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. 5

[14] Marc Hurlbert. Estimated rates of melanoma increasing in 2019, 2019. https://www.curemelanoma.org. 1

[15] Joanna Jaworek-Korjakowska and Pawe Keczek. Automatic classification of specific melanocytic lesions using artificial intelligence. In *BioMed research international*, 2016. 5

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60:84–90, 2012. 5

[17] G. T. Neades, Dub Orr, LeslieE. Hughes, and Kieran Horgan. Safe margins in the excision of primary cutaneous melanoma. *The British journal of surgery*, 80 6:731–3, 1993. 2

[18] Pietro Rubegni, Gabriele Cevenini, Paolo Sbano, Marco Burroni, Iris Zalaudek, Massimiliano Risulo, Giordana Dell'eva, Niccolò Nami, Antonia Martino, and Michele Fimiani. Evaluation of cutaneous melanoma thickness by digital dermoscopy analysis: a retrospective study. *Melanoma research*, 20 3:212–7, 2010. 3, 4, 7

[19] Reuven Y. Rubinstein and Dirk P. Kroese. The cross-entropy method. In *Information Science and Statistics*, 2004. 5

[20] Aurora Sáez, Javier Sánchez-Monedero, Pedro Antonio Gutiérrez, and César Hervás-Martínez. Machine learning methods for binary and multiclass classification of melanoma thickness from dermoscopic images. *IEEE Transactions on Medical Imaging*, 35:1036–1045, 2016. 4, 7

[21] Waqas R. Shaikh, Stephen W. Dusza, Martin A. Weinstock, SusanA. Oliveria, Alan Geller, and Allan Halpern. Melanoma thickness and survival trends in the united states, 1989 to 2009. *Journal of the National Cancer Institute*, 108 1, 2016. 1, 2

[22] Waqas R. Shaikh, Martin A. Weinstock, Allan Halpern, SusanA. Oliveria, Alan Geller, and Stephen W. Dusza. The characterization and potential impact of melanoma cases with unknown thickness in the united states' surveillance, epidemiology, and end results program, 1989-2008. *Cancer epidemiology*, 37 1:64–70, 2013. 2

[23] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. 5

[24] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. 5

[25] Xiangyu Zhang, Jianhua Zou, Kaiming He, and Jian Sun. Accelerating very deep convolutional networks for classification and detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38:1943–1955, 2016. 5