

Towards automatic skin cancer detection by combining multispectral imaging and deep convolutional neural networks

P.H.M. Voorter

*Department of Biomedical Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands
p.h.m.voorter@student.tue.nl*

Abstract—Skin cancer is a prevalent disease, which can be fatal when not detected early. An automatic skin lesion classifier could aid a dermatologist in diagnosing skin cancer. Deep convolutional neural networks (DCNN) have shown strong performances at this task, even though the input images were restricted to only three (RGB) channels. Multispectral imaging captures information that is beyond the human vision and enables the quantification of skin chromophores. Therefore, it is reasonable to assume that this imaging technique adds value to the existing DCNN classifiers.

Several DCNN architectures were developed to combine the advantages of multispectral imaging and DCNN. The networks were trained and validated on a multispectral dataset of 40 malignant and 146 benign lesions. To provide a more sophisticated input to the DCNN, melanin and hemoglobin maps were estimated based on the principles of light propagation through skin. The DCNN that fuses low-level features of an RGB image and a melanin-hemoglobin-residual image, achieved an AUC of 0.87 ± 0.06 , whereas the AUC of the benchmark RGB DCNN was found to be 0.85 ± 0.07 .

Follow-up work should test the proposed DCNN architectures on a larger multispectral dataset. A higher diagnostic accuracy of an automatic skin lesion classifier might be accomplished with multispectral imaging in the future, which would enable early screening for skin cancer.

Index Terms—melanoma, deep convolutional neural network, multispectral imaging, skin cancer, automatic skin lesion classification

I. INTRODUCTION

Skin cancer is one of the most common type of cancer worldwide [1]. Unfortunately, the global incidence rate is still increasing, since the world becomes more and more exposed to UV radiation [1] [2] [3]. One in five people will experience this severe disease at some point during their life [4]. Melanoma is the most aggressive type of skin cancer, since the cancer cells will grow deeply into the skin. Due to the deep invasion, these malignant cells can easily spread to other parts of the body through the lymphatic or vascular system at a later stage [3]. Therefore, early detection of melanoma is of high importance for better treatment outcome and key to saving lives and subsequent healthcare costs [5] [6].

Currently, suspicious skin lesions are assessed by a trained dermatologist, who visually inspects the lesion with a dermatoscope to give a preliminary diagnosis [7]. Dermatoscopy is a

non-invasive way to image the skin lesion with red-green-blue (RGB) color channels and thus mimicking the perception of the human eye. A dermatologist is trained to evaluate color and geometrical features in the dermatoscopic images, e.g. asymmetry, size and irregular borders of a skin lesion [8]. When the dermatologist suspects a skin lesion to be malignant, a biopsy followed by histological examination is prescribed, which is the gold standard to formulate a diagnosis with certainty [7]. Diagnosing all melanomas as malignant is priority, since they can be fatal when not removed early. However, it is also important to classify a skin lesion as benign when there are no malignant characteristics, meaning its invasive removal can be avoided [9]. Still, approximately five invasive biopsies have to be performed to detect one malignant skin lesion [10]. Although diagnostic accuracy has been improved in recent years, discriminating benign and malignant lesions remains challenging, since they often visually resemble each other [1].

A variety of imaging modalities has been studied to support the diagnosis of skin cancer, including multispectral imaging [3] [11] [12]. Although dermatoscopy is still the most widely used device, it requires a lot of user experience to give a reliable diagnosis and these digital images are limited to only three colors channels [3] [9] [12]. Multispectral imaging overcomes these drawbacks and is reported to be more objective and quantitative [3] [12]. By capturing images at multiple wavelengths that are beyond the human vision spectra, this technique is able to provide additional information about the skin lesion, e.g. melanin and hemoglobin content [13] [14]. Since assessing all these wavelength channels and/or chromophore maps as information sources is hard for humans, artificial intelligence was often applied in previous studies for the analysis of multispectral images [9] [15] [16] [17].

In recent years, deep learning techniques were shown to have significant potential to aid the dermatologist in diagnosing skin cancer [3] [18]. Especially deep convolutional neural networks (DCNN) achieved remarkable diagnostic results and proved to be comparable with or even outperform dermatologists [19] [20] [21] [22]. However, to our knowledge a DCNN has not yet been implemented to process multispectral skin

lesion images.

The aim of this study is to investigate whether multispectral images would improve the performance of a DCNN skin lesion classifier. A skin lesion dataset was acquired with a low-cost multispectral device, which captured an RGB image as well as a multispectral image. The first objective of this study was to quantify and localize the hemoglobin and melanin content in the skin lesion by using the measured reflection spectra. As a second objective, the performance of a DCNN classifier that combines the multispectral information (i.e. chromophore maps) is compared to a benchmark DCNN, which only uses RGB images as input. For this aim, several DCNN architectures were explored and evaluated. The long-term aim of this project is to provide a low-cost imaging device that autonomously assesses a skin lesion with high diagnostic accuracy, which would enable early screening of skin cancer.

II. MATERIALS AND METHODS

A. Multispectral measurement device

The skin lesions were imaged with a self-developed multispectral camera. This measurement device contained an RGB camera and a monochrome camera. The monochrome camera was used to capture the skin reflection at specific wavelengths. The skin was sequentially illuminated with LED's that emit light at the following wavelengths: 417, 468, 525, 548, 600, 660, 736, 760, 861 and 950 nm. These wavelengths were chosen in order to capture the specific shape elements of the reflection curves of skin chromophores [23]. The equipment was integrated in a handheld device, ensuring that the skin was only illuminated by the LED's. In addition to the RGB and multispectral images, an environment image was acquired with no LED's on, which was used for calibration of the measured reflection. The images were acquired with a resolution of 75 pixels/mm and captured an area of 27.4x32.7 mm². All images contained a square reference sticker around the lesion, which was build in the device. The region of interest (ROI) within the reference sticker was extracted from the images, containing 1350x1350 pixels. The measured reflection of these ROI's was obtained by the following formula:

$$R_{measured}(\lambda) = \frac{I_{measured}(\lambda) - I_{environment}}{I_{white}(\lambda)} \quad (1)$$

where $I_{measured}$ was the measured intensity, $I_{environment}$ was the measured intensity with no LED's on and I_{white} was the measured intensity of a white object that gave a 100% diffuse reflection.

B. Data

A total of 246 images were acquired of suspicious skin lesions with the multispectral measurement device. After removal of duplicate images, 218 images remained for evaluation by a dermatologist, who gave a preliminary diagnosis. The lesions that required a biopsy, were surgically excised and subjected to histopathological evaluation within a maximum of 14 days from the image acquisition. The lesions were labeled with the biopsy result when available.

Otherwise, the preliminary diagnosis of the dermatologist was given as label, which was the case for 112 benign lesions. Subsequently, binary classification labels were given to the lesions, consisting of the classes 'benign' and 'malignant'. Moreover, 32 lesions were excluded from the analysis due to the following reasons: no diagnosis (29), movement during image acquisition (1) or erroneous output of chromophore map calculation (2). The remaining dataset contained 146 benign lesions (127 melanocytic nevi, 2 skin atrophies, 7 seborrheic keratoses, 1 blue nevus, 1 actinic keratosis, 3 angiomas, 1 sebaceous cyst, 3 keratoacanthomas and 1 hyperpigmentation) and 40 malignant lesions (11 melanomas, 5 squamous cell carcinomas, 24 basal cell carcinomas). All patients provided written informed consent before participating in the study.

C. Diffuse reflectance skin model

The multispectral images were used for the quantification and localization of relevant skin chromophores. The skin can be structured into different layers through which light travels. The composition and the thickness of these layers determine how much the light is absorbed and scattered [24]. Several mathematical skin models have been proposed in literature, varying from a complex multi-layer description (7 to 9 layers) [14] [25] [26] to a simplified two-layer approach [13] [27]. Since the spectral resolution of our measured reflection was limited, the simplified skin model approach of Jolivot et al. was chosen, which consisted of two layers: the epidermis and the dermis [13]. The epidermis is the upper layer of the skin. Its main absorber is melanin, which gives a brown color to the skin when its concentration is high [28]. The dermis is located beneath the epidermis and oxy- and deoxyhemoglobin are assumed to be the major absorbers in the dermis [13] [29]. The skin model provided the reflectance as a function of the melanin fraction in the epidermis (f_{mel}), the blood fraction in the dermis (f_{blood}), the percentage of saturated blood (C_{oxy}), the thickness of the epidermis ($D_{epidermis}$) and the thickness of the dermis (D_{dermis}). A more detailed explanation of this model can be found in Appendix A.

The skin model was implemented in Python. All input parameters were fitted to the measured reflectance spectra in a pixel wise manner using the function *curve_fit* from the optimize module of SciPy. In the fitting process, the parameters were restricted to the physiological ranges for human skin found in literature [13], see Table I. However, the upper bound of f_{mel} and f_{blood} was highered to 0.99 and 0.25 respectively, allowing the model to give an appropriate fit for pixels in the skin lesions as well. By providing these boundaries to *curve_fit*, this function solves the system of equations with the Trust Region Reflective method, which is reported to be quite robust for bounded problems [30]. In addition to the obtained f_{mel} -, f_{blood} -, C_{oxy} -, $D_{epidermis}$ - and D_{dermis} -maps, an amount-of-melanin-map (A_{mel} -map) and a *residual*-map were created. The A_{mel} -map was the result of multiplying the f_{mel} -map with the $D_{epidermis}$ -map. Furthermore, the *residual*-map contained information about the quality of the pixel-wise model fit. This map was created

TABLE I: Overview of the skin model parameters and their physiological ranges [13], which were used as boundaries in the skin model fitting process.

Skin parameter	Symbol	Range
Melanin fraction	f_{mel}	0.013 – 0.99
Blood fraction	f_{blood}	0.002 – 0.25
Oxygenated hemoglobin	C_{oxy}	25 – 90%
Thickness of epidermis	$D_{epidermis}$	0.01 – 0.015 mm
Thickness of dermis	D_{dermis}	0.6 – 3 mm

by calculating the mean squared error of the fitted and the measured reflection data points. It was regularly observed that the difference between the modeled and the measured reflectance was large at 950 nm. Hence, this wavelength was excluded from the fitting process.

D. Deep convolutional neural networks

The multispectral images and chromophore maps were processed by DCNN’s. The DCNN models were written in Python, where the networks were trained and validated using the fastai library [31]. The ROI images were downsampled to a size of 224x224 pixels using a median filter, since all developed DCNN architectures require this input size.

Network architectures: Several network architectures were created to process the RGB images, multispectral images and/or the chromophore maps, based on a pretrained ResNet50 [32]. It has previously been shown that transfer learning improved the performance of a skin lesion classifier when training on small datasets, since it significantly reduces the chance of overfitting the model [33] [34]. Therefore, the weights of the ResNet50 backbone were pretrained on the large-scale ImageNet database [35].

An overview of the network architectures with the required input is shown in Fig. 1. The fusion of the chromophore image features with the RGB image features can be done at a low level or at a high level in the DCNN, see ‘Early fusion’ versus ‘Late fusion’ in Fig. 1b and 1d. The performances of both networks were explored. Furthermore, a DCNN using the 10-channel multispectral images as input was evaluated (Fig. 1c). The network architectures are discussed below in more detail.

RGB ResNet: The original ResNet50 was used to provide the reference performance with the RGB images as input (Fig. 1a), where the last Linear layer ensured a binary classification into ‘benign’ or ‘malignant’.

Early fusion of RGB + skin model maps: All input images (RGB + a combination of three skin model maps) were stacked to a 6-channel input. Originally, the first layer in ResNet50 is a 2D convolutional layer. In this layer, 64 kernels with a size of 3x7x7 are applied to the 3-channel image with a stride of 2, which gives an output size of 112x112x64

for this specific layer. To allow a 6-channel image input, the size of the kernels in this 2D convolutional layer was adjusted to 6x7x7, providing the same layer output size as in the 3-channel network. The other network layers were kept similar to the original ResNet50. This network was shown in Fig. 1b.

Early fusion of multispectral images: The original ResNet50 was adapted in a similar way as described above. All multispectral images were stacked to a 10-channel input. The kernels in the first 2D convolutional layer were adjusted to a size of 10x7x7 and the rest of the network remained the same, see Fig. 1c.

Lated fusion of RGB + skin model maps: In addition to the low-level feature fusion architectures discussed above, also a high-level feature fusion network was build. This network contained two parallel ResNet50 backbones. One backbone processed the RGB images and the other one processed a combination of three skin model maps. Both backbones gave an output of 4096 features, which are concatenated to a total of 8192 features. The two linear layers after the concatenation combined these features and ensured that the output of the network was again a prediction for two classes. The architecture of this late fusion network was visualized in Fig. 1d.

The early and the late fused networks both require six maps as input. These six maps consisted of the red, green and blue channel plus a combination of three skin model maps. To reduce the amount of skin model map combinations, we assumed that the f_{blood} -map and the A_{mel} -map contained the most relevant information. Therefore, these were included in all combinations, that were given as input to the DCNN’s. An overview of all used image combinations is given in Fig. 2.

Cross validation: The dataset was divided into training and validation by using 5-fold cross-validation. Usually, including a separate test set is preferred, since the fine-tuning of the network parameters could affect the performance of the validation set [36]. However, given that the number of malignant skin lesions is limited in our dataset, a test set would not be representative and is therefore not used. Furthermore, the number of malignant skin lesions was equally divided over the five folds to reduce the variety of the validation sets.

Training: During training, the parameters of the ResNet50 backbone were frozen. The trainable parameters in the head of the DCNN of the model were optimized with the weighted CrossEntropy loss function. In this loss function, more weight was put on the loss of the malignant lesions (90%) than on the loss of the benign lesions (10%), since the dataset was imbalanced (i.e. more benign than malignant lesions) and the detection of malignant lesions is considered to be more important than benign lesions. The number of epochs was set to 100. In every epoch, all training images were randomly flipped and/or rotated by multiples of 90° to prevent overfitting. The

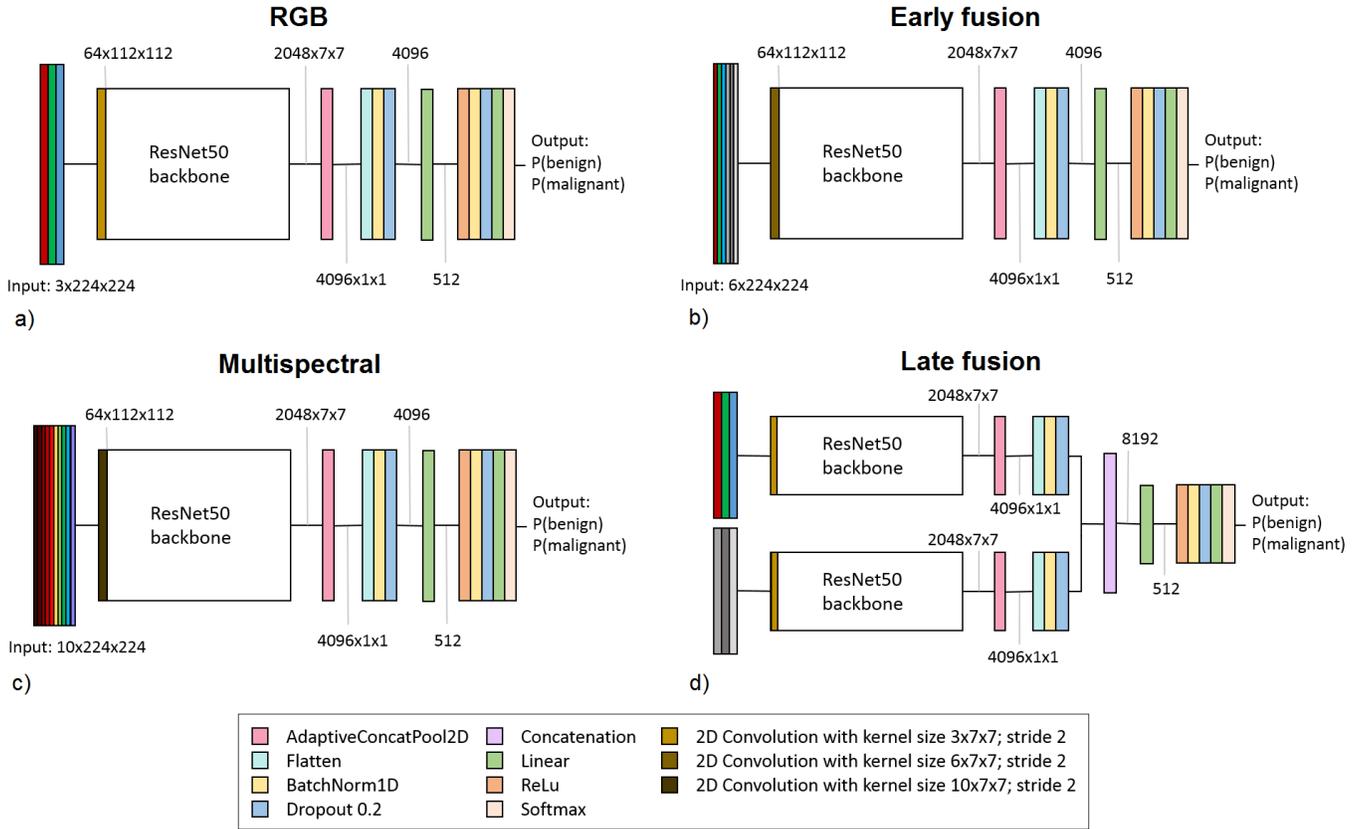


Fig. 1: a) Original ResNet50, which requires a 3-channel image as input (RGB) and gives a prediction (P) for the classes 'benign' and 'malignant' as output, b) Early fused DCNN, where the first layer of the original ResNet50 was modified to allow a 6-channel image as input, c) This multispectral DCNN architecture allows a 10-channel image as input, again by adjusting the first layer of the original ResNet50, d) Late fused DCNN, where the input of two 3-channel images is processed separately through the ResNet50 backbone and the obtained features are concatenated at a higher level.

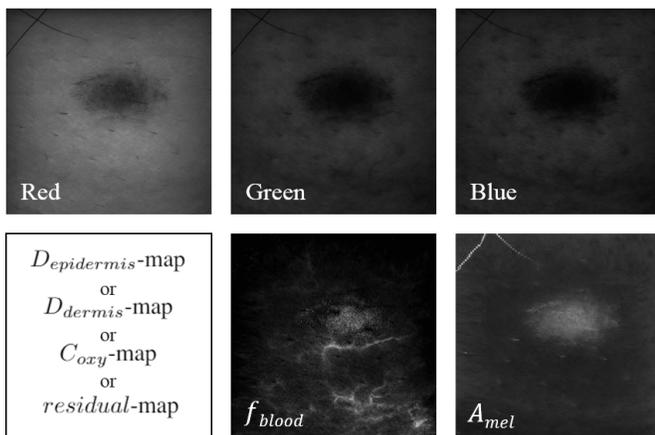


Fig. 2: Overview of the image combinations that were given as input to the early and the late fused DCNN's: 5 maps were fixed (R, G, B, f_{blood} and $A_{melanin}$), and the other map was varied: $D_{epidermis}$, D_{dermis} , C_{oxy} or residual.

network was trained with the function `fit_one_cycle` from the `fastai` library. Using the one cycle policy was recommended, since its performance is better in terms of computational speed and accuracy [37].

Validation: Receiver operating characteristic (ROC) curves were created to evaluate the performances of the different networks. A ROC curve gives an overview of the true positive rate (TPR) and the false positive rate (FPR) when choosing different thresholds for converting the probability output of the network to a binary output. The TPR was defined as the number of correctly classified malignant lesions divided by the total number of malignant lesions. The FPR was defined as the number of benign lesions that were wrongly classified as malignant divided by the total number of benign lesions. The area under the ROC curve (AUC) was determined for all five folds. Subsequently, these five AUC values were averaged and the standard deviation (std) was determined. Last, an average ROC curve was created by calculating the TPR and FPR over all five folds.

III. RESULTS

The diffuse reflectance skin model was successfully applied to 186 multispectral images. However, two skin lesion images gave erroneous results for the chromophore maps, hence excluded from further analysis.

An example of the measured reflection versus the model fit is shown in Fig. 3 for a skin pixel and for a lesion pixel. The reflectance in the lesion is lower for all wavelengths compared to the surrounding skin, which resulted in a high value for the parameter f_{mel} . This figure also demonstrates that the skin model was able to fit the different reflection signatures shapes of lesion versus skin.

A few examples of the f_{blood} -maps and the A_{mel} -maps with the corresponding dermatoscopic images were displayed in Fig. 4. The first two rows in this figure show malignant lesions (top row: melanoma, middle row: basal cell carcinoma), whereas the bottom row visualizes a benign lesion (melanocytic nevus). It was observed that the superficial blood vessel structures were extremely enhanced in the f_{blood} -map compared to the dermatoscopic image. Furthermore, the skin lesion contains a higher amount of melanin compared to the healthy skin.

The chromophore maps and multispectral images were given as input to the DCNN's as described in the *Material and Methods* section. The performances of the DCNN classifiers were summarized in Table II. Moreover, Fig. 5 shows the average ROC curves for the RGB DCNN, the multispectral DCNN and the best performing early and late fused DCNN's. The early fused DCNN, given an input of R-G-B-*residual-f_{blood}-A_{mel}* maps, achieved the highest AUC value, which was equal to 0.87 ± 0.06 . Furthermore, the AUC for the multispectral DCNN was found to be 0.82 ± 0.06 . All AUC values were comparable with the reference AUC value of 0.85 ± 0.07 , i.e. the result of the benchmark RGB DCNN classifier. There was no trend observed for the performances of the early fused networks (average AUC = 0.81) versus the late fused networks (average AUC = 0.82). Last, it is worth noting that the variety of AUC values was high among the different folds (std: 0.06-0.11).

TABLE II: Results of the DCNN classifiers

Input	DCNN	AUC mean \pm std
RGB	RGB	0.85 ± 0.07
Multispectral	Multispectral	0.82 ± 0.11
R-G-B- $D_{epidermis}$ - f_{blood} - A_{mel}	Early fusion	0.80 ± 0.09
R-G-B- $D_{epidermis}$ - f_{blood} - A_{mel}	Late fusion	0.86 ± 0.09
R-G-B- D_{dermis} - f_{blood} - A_{mel}	Early fusion	0.79 ± 0.09
R-G-B- D_{dermis} - f_{blood} - A_{mel}	Late fusion	0.82 ± 0.09
R-G-B- C_{oxy} - f_{blood} - A_{mel}	Early fusion	0.79 ± 0.09
R-G-B- C_{oxy} - f_{blood} - A_{mel}	Late fusion	0.81 ± 0.09
R-G-B- <i>residual-f_{blood}-A_{mel}</i>	Early fusion	0.87 ± 0.06
R-G-B- <i>residual-f_{blood}-A_{mel}</i>	Late fusion	0.79 ± 0.11

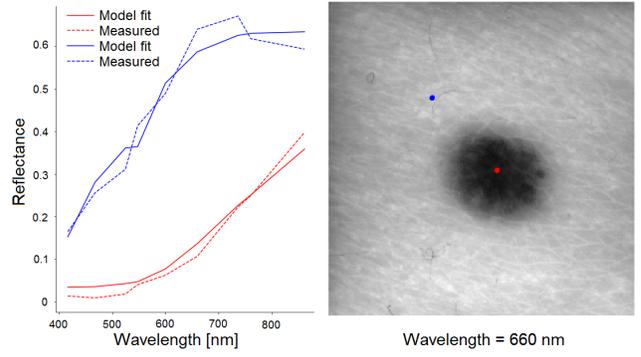


Fig. 3: Example of the measured reflectance and the result of the model fit for a pixel in skin ($f_{mel}=0.28$, $f_{blood}=1.2 \cdot 10^{-2}$, $C_{oxy}=0.90$, $D_{epidermis}=1.1 \cdot 10^{-2}$ mm, $D_{dermis}=1.5$ mm) versus skin lesion ($f_{mel}=0.99$, $f_{blood}=2.4 \cdot 10^{-2}$, $C_{oxy}=0.25$, $D_{epidermis}=1.8 \cdot 10^{-2}$ mm, $D_{dermis}=3.0$ mm).

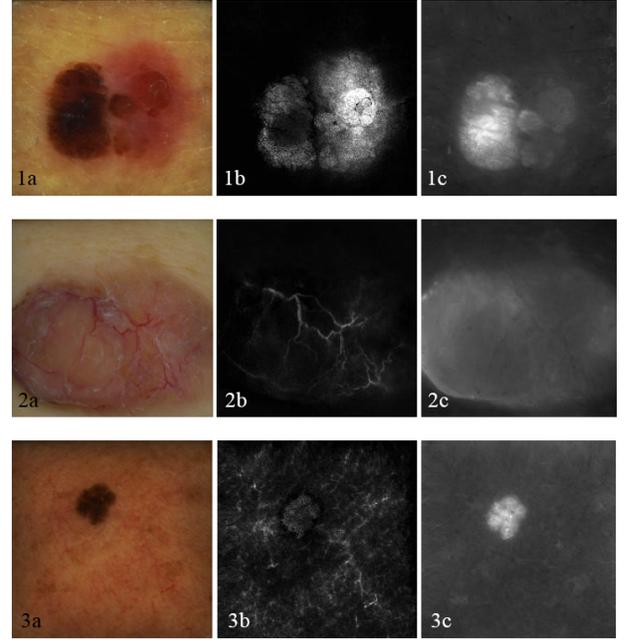


Fig. 4: Examples of dermatoscopic images (1a, 2a, 3a) with the corresponding f_{blood} -maps (1b, 2b, 3b) and A_{mel} -maps (1c, 2c, 3c) for 1) melanoma (malignant); 2) basal cell carcinoma (malignant) and 3) melanocytic nevus (benign).

IV. DISCUSSION

This study introduced a method to combine multispectral imaging and DCNN's for diagnosing skin cancer. As a first objective, information about the skin composition was retrieved from the pixel-wise reflectance signatures. For this, the skin model of Jolivot et al. [13] was implemented, which estimates the hemoglobin and melanin content based on the principles of light propagation. In addition to these skin chromophore maps, the skin model provided spatial information about the

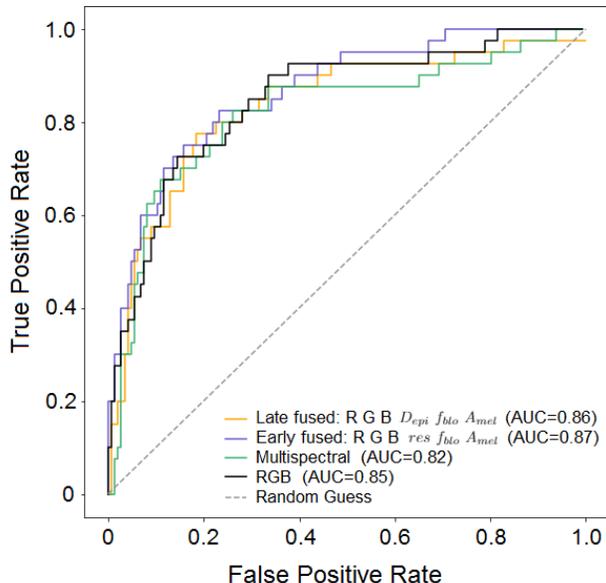


Fig. 5: ROC curves for the best performing early and late fused DCNN’s, the multispectral DCNN and the benchmark RGB DCNN.

thickness of the dermis and epidermis as well as the percentage of the blood that was bounded to oxygen. The results show that this skin model performs well at enhancing the blood (f_{blood}) and melanin (A_{mel}) structures compared to other studies, even though the spectral resolution of the measured reflection was limited. For example, the study of Hosking et al. [9] also derived blood and melanin maps from hyperspectral images to create melanoma biomarkers. However, these two maps seem to have a strong correlation, which was not the case in the current study.

The skin model maps were only validated by visual inspection, and although the model seems to extract the desired information from the multispectral images, it still has some limitations. First, the skin model gave a poor fit for wavelengths in the near-infrared region. In this region, the absorption coefficient of water becomes more relevant, which was not included in this model [38]. Therefore, adding a water component to the used skin model of Jolivot et al. might improve the model reflectance fit for wavelengths above 800 nm. Another limitation is that the skin model systematically overestimates the fraction of melanin (f_{mel}). For example, literature reports a melanin percentage of 43% in darkly pigmented Africans [14] [39], while our estimated f_{mel} in the normal skin surrounding the lesion was regularly higher than 43%. This overestimation was also seen in the melanin maps showed in the article of Jolivot et al. [13]. Here, a melanin percentage of approximately 7% was estimated for skin type II (Fitzpatrick scale), although a value of 1-3% was expected for this skin type [39]. As a consequence of this overestimation, the f_{mel} in a dark skin lesion clips to 0.99. Nevertheless, it

was observed that the epidermis map was thickened at the location of the moles where f_{mel} was clipped. Therefore, the amount-of-melanin map was introduced, which gave a better representation of the melanin content.

Furthermore, there was also a limitation of the measurement equipment that affected the model fit. The measured reflection at the boundaries of the images was slightly lower compared to the center of the image. The first reason for this was that the set up of the LED’s in the multispectral device gave a non-homogeneous field of illumination. The boundaries of the image were slightly less illuminated and therefore also the measured reflection tends to be lower. Another explanation was that more light photons could ‘escape’ at the boundaries of the image due to scattering in the skin, and thus less photons were captured by the detector. As a result of this boundary effect, the skin model parameters were fitted poorly at the boundaries of the image. These boundary effects were especially observed in the D_{dermis} -map.

For the second objective, several DCNN architectures were created that were able to 1) combine the information of the obtained skin model maps (i.e. chromophore maps) with the RGB image features or 2) process the information of the multispectral images directly to obtain a skin lesion classification result. The highest AUC (0.87 ± 0.06) was achieved by a DCNN that fuses the information of the RGB images and the $residual-f_{blood}-A_{mel}$ maps at a low level in the network. The reason that the residual map might contain relevant information is that cancerous cells can produce higher concentrations of chromophores that are not described in the skin model, e.g. keratin [40]. When the skin model cannot fit the measured reflection accurately due to the high concentration of keratin, this is captured in the residual map.

The early fused DCNN architecture combines information of RGB images and skin model maps at a stage where spatial features are still relevant, whereas a late fused DCNN architecture combines high-level features that are extracted from the RGB and chromophore maps separately [41]. Although the information of the input maps is processed in different ways, no significant difference was seen in the performance of the early and the late fused networks.

The results do not show significant improvement in the performance of the DCNN classifier when using multispectral imaging, compared to the benchmark ($AUC = 0.85 \pm 0.07$). The main limitation of this study is that a small and imbalanced dataset was used for the validation and training of the DCNN classifiers. The large diversity in the lesions’ appearances might have made it difficult for the classifier to classify the lesions correctly, since some types of skin lesions had never been seen in training before. This was also reflected by the inconsistent performances among different folds. Therefore, future work should train and validate the DCNN’s with a larger dataset, enabling a better comparison between the DCNN’s due to the reduced uncertainty of the performances. Another limitation of the dataset was that not all benign lesions had been proved by histological examination. Since a biopsy is an

invasive procedure, it would be unethical to perform this on all lesions. Therefore, the possibility that the some benign lesions were actually malignant cannot be ruled out.

As previously concluded in the review of Brinker et al. [33], it is difficult to compare the performances of DCNN skin lesion classifiers across studies. The reason for this is that most studies use different, nonpublic datasets. Attempts to tackle this problem have already been made for RGB DCNN's, since several large-scale RGB datasets were made publicly available (e.g. ISBI 2017 Challenge dataset and ISIC Archive). Such large and widely available datasets should also be collected with multispectral imaging, but this still remains challenging. Another issue is that the definition of the classification problem differs across studies. Most researchers considered the classification problem of melanoma versus benign nevi [20] [21] [42]. However, there are also classifiers that aimed to distinguish 12 skin diseases [43] or even consider more than 150 different classes [22] [44]. In contrast to these studies, our DCNN classifies skin lesions into broader classes. Although the most benign lesions were nevi, the malignant class was more diverse, containing melanomas as well as non-melanomas (basal cell and squamous cell carcinomas).

Recent developments show great potential for the integration of multispectral imaging on smartphones [45] [46]. It is not unlikely that in the near future a mobile phone will contain various multispectral imaging applications, which can function as healthcare tools. This emphasizes the need for a fully automatic skin lesion classifier based on multispectral images, which could support the detection of skin cancer at an early stage.

V. CONCLUSION

This study proposed a new method for skin lesion classification, by combining the advantages of multispectral imaging and DCNN's. It proved the feasibility of a low-cost multispectral measurement device to provide hemoglobin and melanin maps, which have great potential to improve a DCNN skin lesion classifier. Furthermore, the DCNN architectures that were described in the current study, were able to process RGB image input as well as the additional information obtained from multispectral imaging. The performances of these state-of-the-art networks were comparable with the benchmark. However, the uncertainty of the performance was still high, since a small dataset was used. Future work in this field of research should focus on the collection of a large-scale, publicly available multispectral database, enabling a fair comparison between several DCNN methods.

REFERENCES

- [1] Z. Apalla, A. Lallas, E. Sotiriou, E. Lazaridou, and D. Ioannides, "Epidemiological trends in skin cancer," *Dermatology practical & conceptual*, vol. 7, no. 2, p. 1, 2017.
- [2] World Health Organization, "Ultraviolet (UV) radiation and skin cancer," 2017. [Online]. Available: [https://www.who.int/news-room/q-a-detail/ultraviolet-\(uv\)-radiation-and-skin-cancer](https://www.who.int/news-room/q-a-detail/ultraviolet-(uv)-radiation-and-skin-cancer)
- [3] A. Masood and A. Ali Al-Jumaily, "Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms," *International journal of biomedical imaging*, vol. 2013, 2013.
- [4] R. S. Stern, "Prevalence of a history of skin cancer in 2007: results of an incidence-based model," *Archives of dermatology*, vol. 146, no. 3, pp. 279–282, 2010.
- [5] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, no. 1, pp. 7–30, 2020.
- [6] G. P. Guy Jr, S. R. Machlin, D. U. Ekwueme, and K. R. Yabroff, "Prevalence and costs of skin cancer treatment in the US, 2002-2006 and 2007-2011," *American journal of preventive medicine*, vol. 48, no. 2, pp. 183–187, 2015.
- [7] National Collaborating Centre for Cancer (UK and others), "Melanoma: assessment and management," 2015.
- [8] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermoscopy and a new 7-point checklist based on pattern analysis," *Archives of dermatology*, vol. 134, no. 12, pp. 1563–1570, 1998.
- [9] A. M. Hosking, B. J. Coakley, D. Chang, F. Talebi Liasi, S. Lish, S. W. Lee, A. M. Zong, I. Moore, J. Browning, S. L. Jacques et al., "Hyperspectral imaging in automated digital dermoscopy screening for melanoma," *Lasers in surgery and medicine*, vol. 51, no. 3, pp. 214–222, 2019.
- [10] E. W. Breitbart, A. Waldmann, S. Nolte, M. Capellaro, R. Greinert, B. Volkmer, and A. Katalinic, "Systematic skin cancer screening in Northern Germany," *Journal of the American Academy of Dermatology*, vol. 66, no. 2, pp. 201–211, 2012.
- [11] C. Wassef and B. K. Rao, "Uses of non-invasive imaging in the diagnosis of skin cancer: an overview of the currently available modalities," *International journal of dermatology*, vol. 52, no. 12, pp. 1481–1489, 2013.
- [12] E. L. Psaty and A. C. Halpern, "Current and emerging technologies in melanoma diagnosis: the state of the art," *Clinics in dermatology*, vol. 27, no. 1, pp. 35–45, 2009.
- [13] R. Jolivot, Y. Benezeth, and F. Marzani, "Skin parameter map retrieval from a dedicated multispectral imaging system applied to dermatology/cosmetology," *International journal of biomedical imaging*, vol. 2013, 2013.
- [14] S. Vyas, A. Banerjee, and P. Burlina, "Estimating physiological skin parameters from hyperspectral signatures," *Journal of biomedical optics*, vol. 18, no. 5, 2013.
- [15] M. Aboras, H. Amasha, and I. Ibraheem, "Early detection of melanoma using multispectral imaging and artificial intelligence techniques," *American Journal of Biomedical and Life Sciences*, vol. 3, no. 2-3, pp. 29–33, 2015.
- [16] M. Carrara, A. Bono, C. Bartoli, A. Colombo, M. Lualdi, D. Moglia, N. Santoro, E. Tolomio, S. Tomatis, G. Tragni et al., "Multispectral imaging and artificial neural network: mimicking the management decision of the clinician facing pigmented skin lesions," *Physics in Medicine & Biology*, vol. 52, no. 9, p. 2599, 2007.
- [17] R. Marchesini, A. Bono, S. Tomatis, C. Bartoli, A. Colombo, M. Lualdi, and M. Carrara, "In vivo evaluation of melanoma thickness by multispectral imaging and an artificial neural network. a retrospective study on 250 cases of cutaneous melanoma," *Tumori Journal*, vol. 93, no. 2, pp. 170–177, 2007.
- [18] A. G. Pacheco and R. A. Krohling, "Recent advances in deep learning applied to skin cancer detection," *ArXiv*, vol. abs/1912.03280, 2019.
- [19] T. J. Brinker, A. Hekler, A. H. Enk, J. Klode, A. Hauschild, C. Berking, B. Schilling, S. Haferkamp, D. Schadendorf, T. Holland-Letz et al., "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," *European Journal of Cancer*, vol. 113, pp. 47–54, 2019.
- [20] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [21] H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kallou, A. B. H. Hassen, L. Thomas, A. Enk et al., "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, 2018.
- [22] S. S. Han, I. Park, S. E. Chang, W. Lim, M. S. Kim, G. H. Park, J. B. Chae, C. H. Huh, and J.-I. Na, "Augmented intelligence dermatology: Deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders," *Journal of Investigative Dermatology*, 2020.

- [23] N. Kollias, G. Zonios, and G. N. Stamatias, "Fluorescence spectroscopy of skin," *Vibrational Spectroscopy*, vol. 28, no. 1, pp. 17–23, 2002.
- [24] C. Mignon, D. J. Tobin, M. Zeitouny, and N. Uzunbajakava, "Shedding light on the variability of optical skin properties: finding a path towards more accurate prediction of light propagation in human cutaneous compartments," *Biomedical optics express*, vol. 9, no. 2, pp. 852–872, 2018.
- [25] C. Magnain, M. Elias, and J. M. Frigerio, "Skin color modeling using the radiative transfer equation solved by the auxiliary function method: inverse problem," *JOSA A*, vol. 25, no. 7, pp. 1737–1743, 2008.
- [26] I. V. Meglinski and S. J. Matcher, "Quantitative assessment of skin layers absorption and skin reflectance spectra simulation in the visible and near-infrared spectral regions," *Physiological measurement*, vol. 23, no. 4, p. 741, 2002.
- [27] L. O. Svaasand, L. Norvang, E. Fiskerstrand, E. Stopps, M. Berns, and J. Nelson, "Tissue parameters determining the visual appearance of normal skin and port-wine stains," *Lasers in Medical Science*, vol. 10, no. 1, pp. 55–65, 1995.
- [28] R. R. Anderson, J. A. Parrish *et al.*, "The optics of human skin," *Journal of investigative dermatology*, vol. 77, no. 1, pp. 13–19, 1981.
- [29] M. Van Gemert, S. L. Jacques, H. Sterenborg, and W. Star, "Skin optics," *IEEE Transactions on biomedical engineering*, vol. 36, no. 12, pp. 1146–1154, 1989.
- [30] M. A. Branch, T. F. Coleman, and Y. Li, "A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems," *SIAM Journal on Scientific Computing*, vol. 21, no. 1, pp. 1–23, 1999.
- [31] J. Howard and S. Gugger, "Fastai: A layered API for deep learning," *Information*, vol. 11, no. 2, p. 108, 2020.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [33] T. J. Brinker, A. Hekler, J. S. Utikal, N. Grabe, D. Schadendorf, J. Klode, C. Berking, T. Steeb, A. H. Enk, and C. von Kalle, "Skin cancer classification using convolutional neural networks: systematic review," *Journal of medical internet research*, vol. 20, no. 10, 2018.
- [34] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [35] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.
- [36] F. Chollet, *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. MITP-Verlags GmbH & Co. KG, 2018.
- [37] L. N. Smith, "A disciplined approach to neural network hyperparameters: Part 1 - learning rate, batch size, momentum, and weight decay," *ArXiv*, vol. abs/1803.09820, 2018.
- [38] W. S. Pegau, D. Gray, and J. R. V. Zaneveld, "Absorption and attenuation of visible and near-infrared light in water: dependence on temperature and salinity," *Applied optics*, vol. 36, no. 24, pp. 6035–6046, 1997.
- [39] Jacques, Steven L, "Optical absorption of melanin," 1998. [Online]. Available: <https://omlc.org/spectra/melanin/>
- [40] V. Karantza, "Keratins in health and cancer: more than mere epithelial cell markers," *Oncogene*, vol. 30, no. 2, pp. 127–138, 2011.
- [41] V. Osin, A. Cichocki, and E. Burnaev, "Fast multispectral deep fusion networks," *Bulletin of the Polish Academy of Sciences. Technical Sciences*, vol. 66, no. 6, 2018.
- [42] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," *ArXiv*, vol. abs/1703.04197, 2017.
- [43] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *Journal of Investigative Dermatology*, vol. 138, no. 7, pp. 1529–1538, 2018.
- [44] X. Sun, J. Yang, M. Sun, and K. Wang, "A benchmark for automatic visual classification of clinical skin disease images," in *European Conference on Computer Vision*. Springer, 2016, pp. 206–222.
- [45] S. Kim, D. Cho, J. Kim, M. Kim, S. Youn, J. E. Jang, M. Je, D. H. Lee, B. Lee, D. L. Farkas *et al.*, "Smartphone-based multispectral imaging: system development and potential for mobile skin diagnosis," *Biomedical optics express*, vol. 7, no. 12, pp. 5294–5307, 2016.
- [46] Q. He and R. Wang, "Hyperspectral imaging enabled by an unmodified smartphone for analyzing skin morphological features and monitoring hemodynamics," *Biomedical Optics Express*, vol. 11, no. 2, pp. 895–910, 2020.
- [47] P. Kubelka and F. Munk, "An article on optics of paint layers," *Z. Tech. Phys*, vol. 12, no. 593–601, 1931.
- [48] S. Prahl, "Molar extinction coefficients for oxy and deoxyhemoglobin," 1999. [Online]. Available: <https://omlc.org/spectra/hemoglobin/>
- [49] Oregon Medical Laser Center, "Skin optics," 1998. [Online]. Available: <https://omlc.org/news/jan98/skinoptics.html>

APPENDIX

A. Skin reflection model

A reflection spectra of illuminated skin is dependent on the composition of the skin, since the biochemical components determine the absorption and scattering in each skin layer. The skin reflection model of Jolivot et al. [13] was chosen for chromophore map calculation. Jolivot's model was based on the Kubelka-Munk theory of diffuse reflectance in a multi-layer medium [47]. In this model, the skin is represented as a 2-layer medium, i.e. the epidermis and the dermis, which have both their specific layer absorption coefficient and a layer scattering coefficient.

The absorption coefficient of the epidermis layer is defined by:

$$\begin{aligned} \mu_{a,epidermis}(\lambda) = & f_{mel} \mu_{a,melanin}(\lambda) \\ & + (1 - f_{mel}) \mu_{a,baseline}(\lambda) \quad [cm^{-1}] \end{aligned} \quad (A.1)$$

where $\mu_{a,melanin}$ is the absorption coefficient of melanin and $\mu_{a,baseline}$ is the combined absorption coefficient of the other chromophores in skin, e.g. collagen, carotene and keratin [13]. These absorption coefficients are wavelength dependent and are tabulated in Table A1 for our measured wavelengths [13]. Note that the baseline absorption is almost negligible, meaning that the amount of the epidermis absorption is mostly dependent on the melanin fraction.

The absorption coefficient of the dermis can be expressed by:

$$\begin{aligned} \mu_{a,dermis}(\lambda) = & f_{blood} C_{oxy} \mu_{a,oxy}(\lambda) \\ & + f_{blood} (1 - C_{oxy}) \mu_{a,deoxy}(\lambda) \\ & + (1 - f_{blood}) \mu_{a,baseline}(\lambda) \quad [cm^{-1}] \end{aligned} \quad (A.2)$$

where $\mu_{a,deoxy}$ and $\mu_{a,oxy}$ are the absorption coefficient of deoxygenated and oxygenated hemoglobin [13]. The values for $\mu_{a,deoxy}$ and $\mu_{a,oxy}$, retrieved from Prahl [48], slightly differ from each other, shown in Table A1.

Scattering in the dermis is mostly caused by collagen fibers [14] [49]. The scattering coefficient (μ_s) was not defined in Jolivot et al. [13], therefore this coefficient was retrieved from the Oregon Medical Laser Center website [49]. Although the scattering in the epidermis is caused by keratin fibers, the μ_s of the epidermis is comparable with the μ_s of the dermis [49]. Note that the scattering in the epidermis does not have a critical role, since the thickness of the epidermis is small [49]. The scattering coefficient for both layers is given by the following formula:

$$\mu_{s,layer}(\lambda) = 2.0 * 10^{12} * \lambda^{-4.0} + 2.0 * 10^5 * \lambda^{-1.5} \quad (A.3)$$

The wavelength-dependent reflectance (R) and transmittance (T) were calculated for each layer with the Kubelka-Munk equations:

TABLE A1: The absorption coefficients [cm^{-1}] of skin baseline ($\mu_{a,baseline}$), melanin ($\mu_{a,melanin}$), oxygenated hemoglobin (μ_{a,HbO_2}) and deoxygenated hemoglobin ($\mu_{a,Hb}$) for our measured wavelengths.

	$\mu_{a,baseline}$	$\mu_{a,melanin}$	μ_{a,HbO_2}	$\mu_{a,Hb}$
417 nm	2.11	$1.24 * 10^3$	$2.78 * 10^3$	$2.01 * 10^3$
468 nm	1.11	$8.46 * 10^2$	$1.87 * 10^2$	$9.12 * 10^1$
525 nm	$6.09 * 10^{-1}$	$5.77 * 10^2$	$1.65 * 10^2$	$1.88 * 10^2$
548 nm	$5.02 * 10^{-1}$	$5.01 * 10^2$	$2.50 * 10^2$	$2.81 * 10^2$
600 nm	$3.62 * 10^{-1}$	$3.70 * 10^2$	$1.71 * 10^1$	$7.86 * 10^1$
660 nm	$2.92 * 10^{-1}$	$2.69 * 10^2$	1.71	$1.73 * 10^1$
736 nm	$2.59 * 10^{-1}$	$1.87 * 10^2$	2.24	5.90
760 nm	$2.54 * 10^{-1}$	$1.68 * 10^2$	3.14	8.29
861 nm	$2.46 * 10^{-1}$	$1.11 * 10^2$	5.87	3.72
950 nm	$2.45 * 10^{-1}$	$8.01 * 10^1$	6.45	2.23

$$R_{layer}(\lambda) = \frac{(1 - \beta^2)(e^{K_{layer}D_{layer}} - e^{-K_{layer}D_{layer}})}{(1 + \beta)^2 e^{K_{layer}D_{layer}} - (1 - \beta)^2 e^{-K_{layer}D_{layer}}} \quad (A.4)$$

$$T_{layer}(\lambda) = \frac{4\beta}{(1 + \beta)^2 e^{K_{layer}D_{layer}} - (1 - \beta)^2 e^{-K_{layer}D_{layer}}} \quad (A.5)$$

where K_{layer} is the backward flux and β_{layer} is the forward flux in each layer, expressed by:

$$K_{layer}(\lambda) = \sqrt{2 \mu_{a,layer}(2 \mu_{a,layer} + 4 \mu_{s,layer})} \quad (A.6)$$

$$\beta_{layer}(\lambda) = \sqrt{\frac{2 \mu_{a,layer}}{2 \mu_{a,layer} + 4 \mu_{s,layer}}} \quad (A.7)$$

Hereafter, the total reflectance of a two-layer skin medium can be approximated by:

$$R_{total}(\lambda) = R_{epidermis} + \frac{T_{epidermis}^2 R_{dermis}}{1 - R_{epidermis} R_{dermis}} \quad (A.8)$$