Real-time skin chromophore estimation from hyperspectral images using a neural network

Lou Gevaux, Jordan Gierschendorf, Juliette Rengot, Marie Cherel, Pierre Séroul, Alex Nkengne, Julie Robic, Alain Trémeau, Mathieu Hébert

Real-time skin chromophore estimation from hyperspectral images using a neural network

Lou Gevaux [1], Jordan Gierschendorf [2], Juliette Rengot [2], Marie Cherel [2,*], Pierre Séroul [3], Alex Nkengne [4], Julie Robic [4], Alain Trémeau [1] and Mathieu Hébert [1]

[1]Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert, Curien UMR 5516, F-42023, Saint-Etienne, France ;
[2]Newtone Technologies, Lyon, France ;
[3]Bull Technologies, Atos, Grenoble, France ;
[4]Laboratoire Clarins, Pontoise, France ;
**mcherel@newtone.fr*

Acknowledgements: This work was supported by Newtone Technologies, the Région Auvergne-Rhône-Alpes within the program ARC 6, and the French National Research Agency (ANR) as part of the program "Investissements d'Avenir" (ANR-11-IDEX-0007) in the framework of the LABEX MANUTECH-SISE (ANR-10-LABX-0075) of Université de Lyon.

Disclosures: Jordan Gierschendorf, Marie Cherel and Juliette Rengot were full-time employees of Newtone Technologies at the time of the study, which uses the SpectraCam® hyperspectral camera for commercial purposes.

Abstract

Background: Hyperspectral imaging for in vivo human skin study has shown great potential by providing non-invasive measurement from which information usually invisible to the human eye can be revealed. In particular, maps of skin parameters including oxygen rate, blood volume fraction, and melanin concentration can be estimated from a hyperspectral image by using an optical model and an optimization algorithm. These applications, relying on hyperspectral images acquired with a high-resolution camera especially dedicated to skin measurement, have yielded promising results. However, the data analysis process is relatively expensive in terms of computation cost, with calculation of full-face skin property maps requiring up to 5 hours for 3-megapixels hyperspectral images. Such a computation time prevents punctual previewing and quality assessment of the maps immediately after acquisition.

Methods: To address this issue, we have implemented a neural network that models the optimization-based analysis algorithm. This neural network has been trained on a set of hyperspectral images, acquired from 204 patients and their corresponding skin parameter maps, which were calculated by optimization.

Results: The neural network is able to generate skin parameter maps that are visually very faithful to the reference maps much more quickly than the optimization-based algorithm, with computation times as short as 2 seconds for a 3-megapixel image representing a full face and 0.5 seconds for a 1-megapixel image representing a smaller area of skin. The average deviation calculated on selected areas shows the network's promising generalization ability, even on wide-field full-face images.

Conclusion: Currently, the network is adequate for preview purposes, providing relatively accurate results in a few seconds.

Keywords: hyperspectral, imaging, in vivo, non-invasive, machine learning.

1. Introduction

Skin, which acts as a barrier to the body's surrounding environment, is an organ whose study has significant social stakes. The study and diagnosis of skin pathologies is crucial in the fields of medicine and dermatology, such as for pathologies like skin cancer, associated to high rates of mortality, for which early detection increases the chances of healing. The study of skin is also of considerable interest to cosmetologists, as the appearance of someone's skin plays an important part in the perception of beauty, age and health [1]. In recent years, the use of optical methods for skin analysis has grown markedly, as they allow for non-invasive in vivo measurement with no discomfort for the patient. Among the emerging optical methods, imaging methods, which can offer spatial information measurements on large areas and with high resolution, are especially promising for the study of skin heterogeneous properties.

Hyperspectral imaging, which collects spectral information at many narrow wavebands for each pixel of an image, has shown high potential as a tool for skin analysis [2]. It allows high-resolution measurement of skin spectral reflectance, whose shape, or "spectral signature", can be related to skin structure and composition using an optical model. In particular, maps showing skin absorption properties such as oxygen rate, blood volume fraction or melanin

concentration can be generated from hyperspectral images using an optical-based model [3,4]. This analysis method yields quantitative information, which can be helpful for assessing skin properties objectively.

In our previous works [3–5], we have used two hyperspectral systems to acquire skin spectral reflectance, SpectraCam® and SpectraFace® (Newtone Technologies, France). These systems have been specially designed for skin measurement, and can capture hyperspectral images at 31 wavelengths in the visible spectrum in around 2 seconds. The acquired images have then been analyzed using an optical model. We have selected a two-flux light transport model and a model of skin as a two-layer material, applying an inverse approach to retrieve skin properties from the hyperspectral measurement. Given the complexity of the inverse problem, it has been solved by optimization, an iterative process which requires significant computation time: typically, around 5 hours for a 3-megapixel image of a full face. In the context of a cosmetology study, this is a significant limitation as it prevents the user from evaluating the quality of the analysis results immediately after the hyperspectral image has been captured.

Machine learning applied to regression problems present an interesting alternative to time-consuming optimization-based algorithms, which have hitherto been used in many works in the field of tissue optics [6–10]. Nowadays, many machine learning tools are open source and supported by large communities of users, easing their use in many domains, though creating an adequate dataset to train a neural network can be challenging. Over the course of our studies, we have managed to collect a sufficient amount of data comprising hyperspectral images and their corresponding property maps, to train a neural network in order to replace the classical optimization-based algorithm used for skin analysis, allowing results to be obtained in a few seconds rather than several hours.

Many researchers in the field have used neural networks to model skin spectral reflectance as a function of skin properties (see Figure 1.a). Using a detailed skin model and a Monte Carlo method to model light transport in skin, a "synthetic" dataset can be directly simulated. This synthetic dataset comprises skin parameters and their associated spectral reflectances, and is used to train a neural network that is able to estimate skin parameters from spectral reflectance [7–10]. By contrast, in this work, the light-skin interaction model is already defined and the neural network is trained to replace the optimization algorithm (see Figure 1.b). This allows us to use real world noisy data in the training set rather than synthetic data. In this way, we aimed to obtain better results on real input data which contain similar noise as the training set.

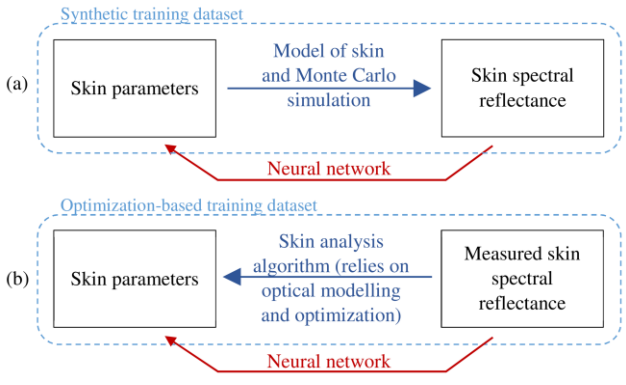

Figure 1. Flowchart of two possible approaches for skin parameters estimation using neural network. Approach (a) relies on a synthetic training dataset generally simulated using a Monte Carlo method. Approach (b), presented in this paper, relies on the analysis of real world measurements by the optimization-based algorithm detailed in Section 3.

We selected a relatively simple type of artificial neural network, the multi-layer perceptron (MLP), and applied it to our specific problem of skin analysis from hyperspectral images. Our aim was to understand to what extent the classical optimization algorithm can be replaced by a faster method with accurate results. Working on skin presents a challenge, as it is a complex material whose spectral reflectance varies significantly according to several parameters, such as location on the body, age, skin color, genetic traits, living conditions, and skin pathologies. Consequently, the data selected to train the neural network must be as diverse as possible to faithfully represent the wide range of skin spectral reflectance that can be measured. As such a dataset cannot be exhaustive, the capacity of the network to perform well on data that are not part of the training set depends on the network generalization ability. A second challenge lies in implementing a method that accounts for the irradiance drifts occurring on the curved parts of the body. If irradiance drifts are not accounted for, the neural network model cannot correctly predict chromophore concentrations from full face images. We proposed to address this issue by data augmentation, adding data artificially altered by irradiance drifts to the training dataset.

The rest of the paper is organized as follows: Section 2 details the hyperspectral acquisition setups; Section 3 describes the optical model and optimization used in the classical skin analysis method; the artificial neural network implemented to replace the classical analysis method is described in Section 4; results are presented in Section 5; and finally, conclusions are drawn in Section 6.

2. Skin hyperspectral imaging

Hyperspectral imaging is the acquisition of a two-dimensional image with spectral information in each pixel, resulting in three-dimensional data often called "hypercube" [11]. A hyperspectral camera measures spectral radiance over many narrow and non-overlapping bandwidths that contiguously cover a part of the radiative spectrum, providing high spectral resolution data.

The devices used in this work, shown in Figure 2, are specifically dedicated to in vivo skin measurement. They have been developed by Newtone Technologies (France) for small-area imaging (SpectraCam®) [3,5] and for full-face imaging (SpectraFace®) [4].

SpectraCam® SpectraFace®

Figure 2. SpectraCam® and SpectraFace® systems developed by Newtone Technologies (France), hyperspectral cameras dedicated to skin imaging, respectively for small area and full face acquisition.

These cameras comprise a liquid crystal tunable filter (LCTF) serving as a wavelength selection filter, placed in front of a monochrome CMOS camera, and LEDs that homogeneously illuminate the scene. The LED lighting units are designed using blue and white LEDs in order to provide sufficient irradiance on the skin over the whole visible spectrum. Linear polarizing filters are added to each light source and oriented in a cross-polarization configuration with the LCTF polarization direction. This cross-polarization configuration prevents the capture of light specularly reflected at the skin-air interface. In this way, only the diffusely reflected light, corresponding to the light that has travelled within skin and that contains information about skin structure and composition, is measured. In the rest of the paper, only the diffuse component will be addressed. The hyperspectral image is acquired via a temporal scan of the spectral information: images of 2048 × 2048 pixels are sequentially acquired for 31 wavelengths covering the visible spectrum (400 to 700 nm) with 10 nm steps. The total acquisition time is around 2 seconds.

The signal captured by the hyperspectral camera depends on a multitude of parameters beside skin surface reflectance, including incident irradiance, sensor spectral response and optics transmittance. Skin spectral reflectance is obtained independently from the acquisition parameters after applying a calibration step that relies on the acquisition of black and white diffusing tiles. This calibration does not account for the variations in illumination occurring on non-flat samples (such as the face). In those cases, the measurement is affected by what we call irradiance drifts: the calibration step yields reflectance multiplied by a constant that depends on the local shape of the surface.

The pictures in Figure 3 are examples of images at three different wavelengths from a hypercube acquired using SpectraFace®. They show how skin properties vary with wavelength, with melanin spots easily identifiable at blue wavelengths (420 nm), blood vessels highly visible at 490 nm and the skin appearing very uniform in red wavelengths (700 nm), the part of the visible spectrum for which skin is the most translucent.

3. Skin analysis using optical modeling and optimization

By combining hyperspectral imaging with an optical model, we obtain skin property maps showing oxygen rate, blood volume fraction, melanin concentration, bilirubin concentration, and epidermis thickness. In this section, we summarize the skin model, the light-skin interaction model and the optimization method used for skin analysis and refer to Refs. [3,4] for the detailed equations.

In our model, skin is described as a two-layer semi-infinite material that is homogeneous over the area of interest, i.e., the area imaged into one pixel of the sensor. The top layer, of thickness *h*, roughly corresponds to the epidermis and contains melanin and a baseline. The bottom layer, corresponding to the

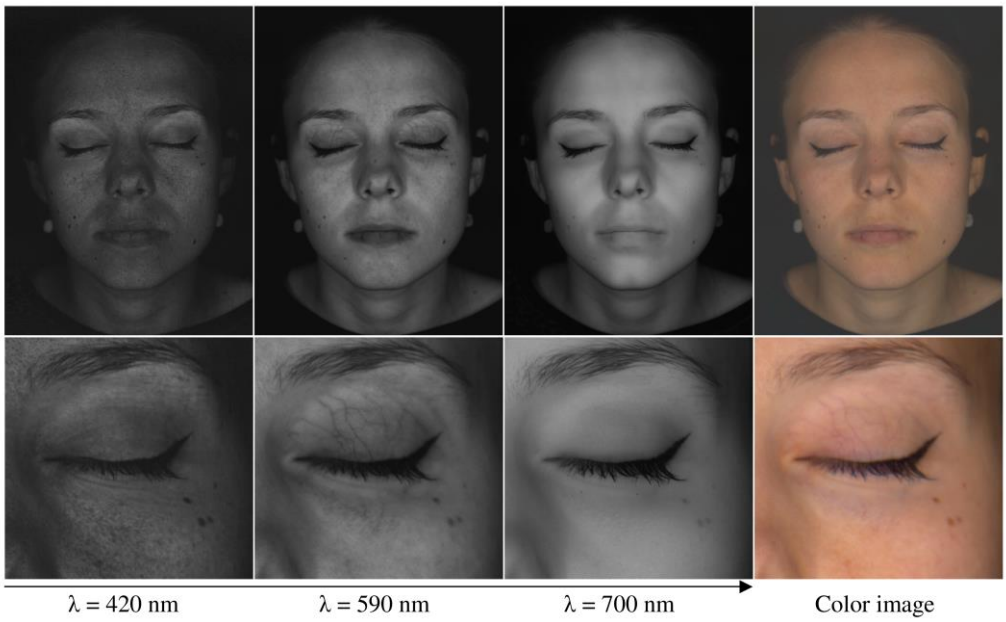

Figure 3. Hyperspectral image at several wavelengths and corresponding color image. Full face image (top), detail of right eye (bottom).

dermis, is assumed to be infinitely thick and contains oxyhemoglobin (HbO_2), deoxyhemoglobin (Hb), bilirubin, and a baseline. Each skin chromophore is characterized by its volume fraction, i.e., concentration: the quantities $\{c_{mel}, c_{HbO2}, c_{Hb}, c_{bi}\}$ respectively denote the concentrations in melanin, HbO_2, Hb and bilirubin, and can vary between 0 and 100%.

Light propagation in each layer is described using the Kubelka-Munk theory [12], by which light interactions are characterized in terms of absorption and scattering, defined respectively by the optical coefficients K and S. The overall optical model can be described in three steps illustrated in Figure 4.

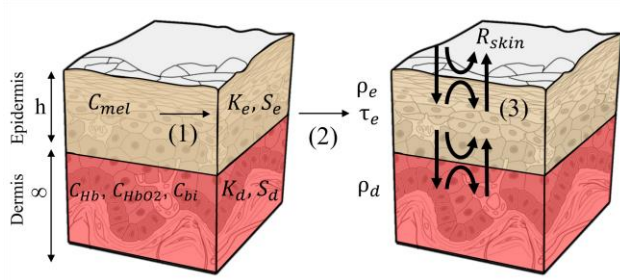

Figure 4. Skin model and light-skin interaction models used in the analysis method: (1) Beer-Lambert-Bouguer laws, (2) Kubelka Munk theory, (3) Kubelka formula and Saunderson correction.

Firstly, epidermis and dermis absorption properties $K_e(\lambda)$ and $K_d(\lambda)$ are related to skin chromophore concentrations $\{c_{mel}, c_{HbO2}, c_{Hb}, c_{bi}\}$ and the spectral absorption properties of each chromophore taken from literature [13–15] using the Beer-Lambert-Bouguer law.

Secondly, applying the Kubelka Munk theory [12], the reflectance ρ_e and transmittance τ_e of the epidermis, and the reflectance ρ_d of the dermis are obtained as functions of the layers' absorption and scattering coefficients (K_e and S_e, respectively, for the epidermis, and K_d, and S_d, respectively, for the dermis). The spectral scattering coefficients of epidermis $S_e(\lambda)$ and dermis $S_d(\lambda)$ are taken from literature [13].

Finally, the Kubelka formula [16] is used to take into account the two-layer configuration of skin, and the Saunderson correction [17] is applied to account for the light transfers (reflections and transmissions) at the interface between air and skin. These models yield skin spectral reflectance R_{skin} as a function of ρ_e, ρ_d, and τ_e.

The combination of these three steps gives the direct relationship between skin spectral reflectance R_{skin}, skin chromophore concentrations $\{c_{mel}, c_{HbO2}, c_{Hb}, c_{bi}\}$ and epidermis thickness h.

The inverse problem, in which skin chromophore concentration and epidermis thickness are estimated from the measured spectral reflectance, is solved by optimization. For each pixel, the parameters $\{c_{mel}, c_{HbO2}, c_{Hb}, c_{bi}, h\}$ that minimize the distance $d(R_{skin}, R_m)$ between the measured spectrum R_m and the predicted spectrum R_{skin} are determined using an iterative method:

$$\{c_{mel}, c_{Hb}, c_{HbO2}, c_{bi}, h\} = \underset{c_{mel}, c_{Hb}, c_{HbO2}, c_{bi}, h}{\arg\min} \left(d\left(R_{skin}, R_m\right)\right). \quad (1)$$

The optimization results are strongly dependent on how the distance between two spectra is defined. For this application, the Spectral Angle Mapper (SAM) [18], defined in Eq.(2), was chosen for its properties of rendering only two spectra shape differences while being independent from the spectra amplitude difference, which makes it insensitive to irradiance variations.

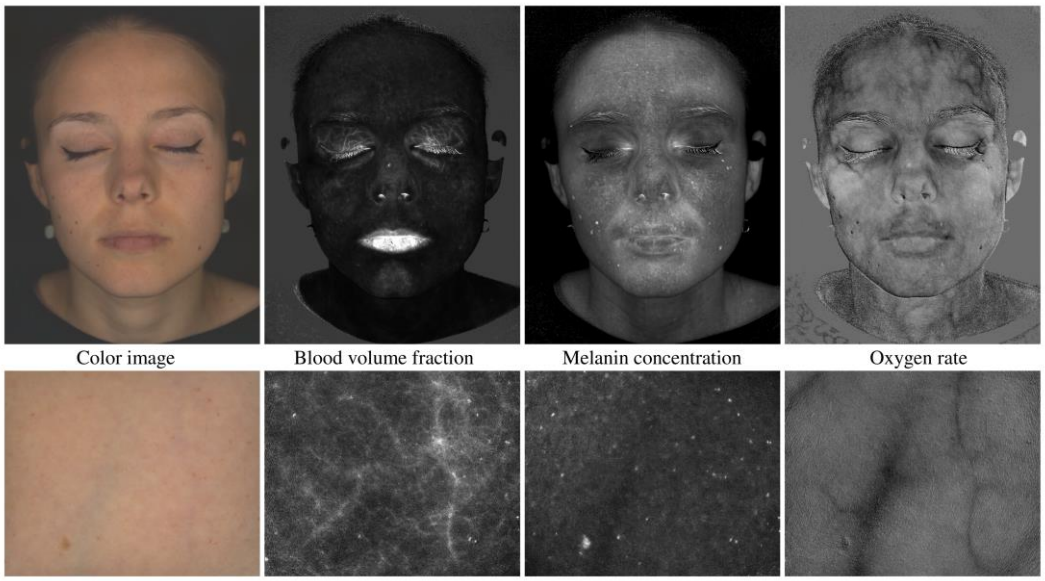

Figure 5. Estimated parameter maps using the analysis method on a full-face image (top) and a small area (4 × 5 cm) image (bottom). The grayscales of the displayed images have been optimized for better display and may vary between images.

$$d_{SAM}(R_{skin}, R_m) = \arccos\left(\frac{\sum R_{skin}(\lambda)R_m(\lambda)}{\sqrt{\sum R_{skin}^2(\lambda)\sum R_m^2(\lambda)}}\right). \quad (2)$$

Owing to these properties, the SAM metric yields good results on flat areas as well as on a full face, despite the presence of irradiance drifts.

To make interpretation of results simpler, the concentrations in Hb and HbO2, respectively c_{Hb} and c_{HbO2}, have been replaced with the blood volume fraction c_{blood} and the oxygen rate α. The blood volume fraction is the sum of the concentrations in Hb and HbO2. The oxygen rate is the ratio of the concentration in HbO2 to the blood volume fraction. Examples of skin parameter maps obtained on a full face and on a small area are presented in Figure 5. In the rest of the article, this skin analysis algorithm will be referred to as the optimization-based algorithm.

4. Skin analysis using a neural network method

The skin acquisition and analysis method presented in the previous section yields satisfactory results for a number of applications in cosmetology [3,5]. However, this method relies on iterative optimization for each pixel of the hyperspectral image, which results in a high computational cost that limits its viability for applications requiring immediate results. To tackle this limitation and reduce computation time, we trained an artificial neural network to replace the optimization algorithm. In this section, we describe the training dataset, the neural network architecture, and the training process used in this study.

4.1 Training dataset

The dataset used to train and test the neural network comprises hyperspectral images acquired during a clinical study conducted by Clarins (France) on 204 Asian patients of skin phototypes I to III. Using a SpectraCam®, hyperspectral images were acquired from the inner forearm, the cheek and the dark circles under the eyes for each patient. These images represent the spectral reflectance of healthy skin at 31 wavelengths between 400 and 700 nm for 1148 × 948 pixels.

From these images, maps of skin parameters have been estimated using the skin analysis method described in Section 3. The measured skin spectral reflectance on one pixel corresponds to the input data of the neural network, and the output values that we are seeking correspond to the calculated skin parameters for the pixel, i.e. oxygen rate, blood volume fraction, melanin concentration and epidermis thickness. In Section 3, we mentioned that the skin analysis model yields maps of bilirubin concentrations as well. For the neural network model, however, bilirubin has been excluded. Bilirubin is a breakdown product of melanin and is present only at very low concentrations in healthy skin, typically around 0.1%. Higher concentrations generally occur in the incidence of pathologies such as jaundice, or on healing bruises. As our dataset comprises only healthy skin (with low bilirubin concentration), the data about bilirubin concentration are extremely noisy due to a low signal-to-noise ratio. As such, including these values in the training dataset would reduce the overall accuracy of the neural network.

The training dataset is a collection of spectral reflectances and the associated skin parameters corresponding to selected pixels on each available image. To train the neural network to accurately reproduce the skin analysis algorithm, the dataset must contain varied spectral reflectances that represent as widely as possible the range of possible skin spectral reflectance, and must not contain any pixels corresponding to features that are not skin, such as the eyelashes. Hence, areas of interest corresponding to the most diverse and vascularized areas have been manually created on each image. A total of 812 000 pixels were then randomly selected from these areas to form the dataset. For these pixels, it has been verified that they correspond to spectral reflectances between 0 and 1, and to skin parameters within the possible ranges of [0, 100%] for oxygen rate, blood and melanin concentrations. The dataset sampling process is illustrated by Figure 6.

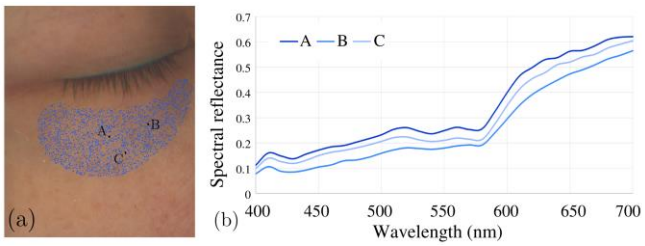

Figure 6. Illustration of the dataset sampling process. (a) Hyperspectral image (displayed here as a RGB color image) with pixels randomly selected within an area of interest and (b) examples of spectral reflectances associated with the pixels labelled as A, B and C in the image on the left.

The measured skin spectral reflectances were then modified to linearly extend the dynamic of the input spectrum within the interval [0, 1], to improve neural network performance. With r_{max} and r_{min} the maximum and minimum reflectance values over the entire dataset and for all wavelengths, the new input data are defined as:

$$R(\lambda_i) = \frac{r(\lambda_i) - r_{\min}}{r_{\max} - r_{\min}} \quad (3)$$

Finally, we applied a data augmentation method to increase the robustness of the neural network to irradiance drifts. The spectral reflectance of the original training dataset was multiplied by random constants uniformly distributed between 0.3 and 1.3, thus simulating the effect of varying lighting and viewing conditions that occur on some parts of the face when measured in full using the SpectraFace® camera. This new "irradiance drifts" dataset was added to the original data, doubling the size of the dataset.

The 1 624 000 selected training patterns (i.e., skin spectral reflectance measured on one pixel and the corresponding skin parameters) were randomly distributed between the learning

dataset (containing 1 032 000 patterns), the validation dataset (containing 272 000 patterns) and the test dataset (containing 320 000 patterns). The training dataset was used to optimize the trainable parameters of the network, and the validation dataset was used at each iteration of the training process to verify that the MLP is not over-trained, and therefore unable to perform well on any data outside the training set. The test training set was used to assess the accuracy of the neural network prediction.

4.2 Neural network architecture

The type of neural network used in this work is a multi-layer perceptron (MLP), a prevalent class of feed-forward network [19] that can be applied to a wide variety of tasks including prediction, function approximation, and pattern classification. This kind of network is a set of several perceptrons, or artificial neurons, organized into layers.

In this work, the MLP, whose architecture is shown in Figure 7, contains 5 fully connected hidden layers that are each constituted of 31 perceptrons. The entry layer connects the 31 perceptrons of the first hidden layer with the spectral reflectance R at each wavelength λ_i. The output layer comprises 4 signals corresponding to the oxygen rate, the blood volume fraction, the melanin concentration and the epidermis thickness.

Each perceptron is characterized by the rectified linear unit (ReLU) activation function, which is recognized to yield good results for modeling regression problems. Each hidden layer is preceded by a batch normalization function according to the method developed by Ioffe and Szegedy [20]. Batch normalization increases the performance of the training process by constraining the distribution of each layer's input. In total, the MLP comprises 5460 trainable parameters, which are randomly initialized according to a normal distribution of mean value 0 and standard deviation 0.01.

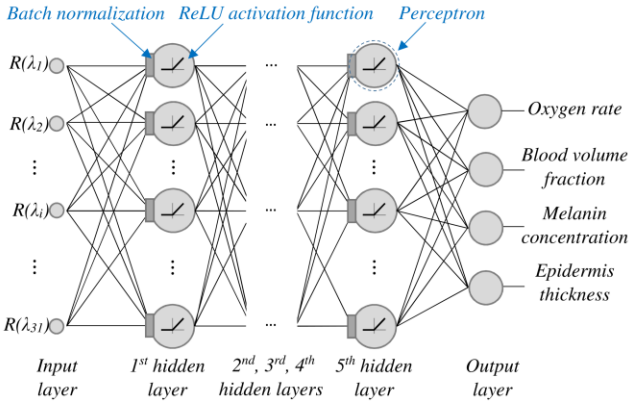

Figure 7. Architecture of the MLP used to model the function that relates skin spectral reflectance to skin properties.

The choice of the number of layers and perceptrons depends on the complexity of the function that has to be modeled [21]. Too few trainable parameters yield a poor fit of the function, and too many yield a model that overfits the training data and has poor generalization performance, i.e., performs well on the training set, but poorly on novel data. In this work, several configurations were considered, with 1, 3 and 5 hidden layers. The last configuration, with 5 hidden layers, yielded the best performance while remaining within an acceptable ratio of trainable parameters to training data (around 1% of the dataset size). The optimum values for the MLP parameters are searched using the iterative training process described in Section 4.3.

Finally, we observed that using a training dataset whose size has been doubled by data augmentation improved generalization ability, but to the detriment of fitting performance. To mitigate the fall off in fitting performance, post-processing was applied: each output signal undergoes a linear correction whose parameters are determined by training on the original dataset without the irradiance drift dataset.

4.3 Training process

The MLP has been trained using a back-propagation learning algorithm [22], an optimization algorithm that uses a gradient-descent method. The loss function that is minimized during the MLP parameters optimization is defined as follows. Each pixel i of the training set corresponds to a training pattern or input-target pair:

$$\left(R_i(\lambda);\left\{\alpha_i, c_{blood,i}, c_{mel,i}, h_i\right\}\right), \qquad (4)$$

where $R_i(\lambda)$ denotes the spectral reflectance at 31 wavelengths, α_i the estimated oxygen rate, $c_{blood,i}$ the estimation blood volume fraction, $c_{mel,i}$ the melanin concentration and h_i the epidermis thickness. The target values $\{\alpha_i, c_{blood,i}, c_{mel,i}, h_i\}$ are also referred as ground truth. We denote as $\{A_i, C_{blood,i}, C_{mel,i}, H_i\}$ the neural network output values.

The loss function is defined as a weighted sum of absolute differences between the target values and the output values:

$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{k}\begin{pmatrix} a\left|\alpha_i - A_i\right| + b\left|c_{blood,i} - C_{blood,i}\right| \\ +m\left|c_{mel,i} - C_{mel,i}\right| + h\left|h_i - H_i\right| \end{pmatrix}, \qquad (5)$$

where $\{a, b, m, h\}$ are the weights associated to each parameter, k is the sum of these weights, and N is the number of pixels. A mean square error loss function was also implemented, but it performed less well than the mean absolute error loss function given in Eq. (5).

The definition of weights associated to each parameter helps to balance each parameter importance in the loss function for them to be estimated with similar accuracy. Indeed, the parameters considered do not vary within the same intervals: in the training dataset, oxygen rate typically varies between 18% and 96%, blood volume fraction between 2% and 12%, melanin concentration between 21% and 51%, and epidermis thickness between 21 µm and 41 µm. Without weights in the loss function, the prediction of blood volume fraction, for example, would be much less accurate than the other parameters. An alternative to introducing these weights would be to rely on relative difference rather than absolute

difference in the loss function. After several tests, the weights chosen in our model are $a = 1$ for the oxygen rate, $b = 4$ for the blood volume fraction, $m = 2$ for the melanin concentration and $h = 2$ for the epidermis thickness. By applying these settings, we obtain $k = 9$.

During the optimization process, the loss function decreases for both training and validation datasets when the model is converging. However, when the network starts to be over-trained, the loss function on the validation dataset increases while it keeps decreasing on the training dataset. This change of trend provides a clear indication of when training should be stopped and was used for automatic early-stopping.

The model was implemented in TensorFlow 1.10 framework using Python 3.5. The training process was performed on a single GPU NVidia 1080 Ti GTX 10Gb RAM. The model trained for around 5 days, requiring roughly 5 500 steps.

The average deviation between the MLP prediction and the optimization-based algorithm results (referred to as ground truth) obtained at the end of the training process on the validation and test sets for each output are given in Table 1.

The average deviation between the MLP prediction and the optimization-based algorithm results that serve as ground truth is rather equally distributed between each output signal on the test and validation datasets. This shows that the loss function weights have been adequately chosen. Surprisingly, the relative deviation is not particularly low compared to the variations that can be observed in clinical studies (for example, Nkengne et al. observed a 5.5% significant decrease (p-value = 0.026) in melanin concentration on the area under the eyes after applying an anti-dark circle cosmetic during 56 days on patients of skin phototypes II and III [5]). This is perhaps because doubling the training dataset with data altered by irradiance drifts made the regression problem harder to solve for the neural network. We will however see in Section 5.3 that this processing of the training dataset is necessary when applying the MLP on full face image.

These levels of relative deviation on the validation and test datasets give us a hint at the minimal deviation that can be expected when applying the MLP on images.

5. Experimental results

The MLP was implemented to perform skin analysis from hyperspectral images faster than the conventional optimization-based algorithm. Our results show that computation time is much faster using this method: while computation was about 1 hour for a 1148 × 948 pixels image with the optimization-based algorithm, the computation time using the MLP is only 0.5 seconds using a GPU and 2 seconds using a CPU on the same computer (Intel®Core™ i7-6700 CPU 3.40 Ghz). This is respectively around 7 000 times and 1 800 times faster than the optimization-based algorithm. The implemented MLP is hence validated in term of computation speed.

Besides computation speed, we have to verify that the MLP yields accurate results that can be used for skin analysis and explore how it can be used in various applicative contexts. Through the training process, the MLP parameters have been optimized so that the loss function on the test dataset is as low as possible. However, this does not guarantee reliable results on real-world images, which potentially involve skin spectral measurements that were not represented in the training dataset. We therefore prefer to discuss the efficiency of the method by comparing its output with the expected results on images. In this section, various images of healthy skin are used to study the accuracy of the method.

5.1 Predicted skin parameter maps on images similar to the training dataset

Skin parameter maps estimated using the neural network (referred to as MLP) on images of the dark circles under the eyes and the inner forearm are compared with the optimization-based algorithm results (referred to as ground truth) in Figures 8 and 9. The images selected here as examples are part of the cosmetology study from which the training dataset has been built and are therefore very similar to the data used for the MLP training. The deviations, i.e., absolute differences, between the MLP prediction and the ground truth obtained on areas of interest selected under the eye and on the forearm (see Figure 10) are detailed in Tables 2 and 3, respectively.

Table 1. MLP results on the validation and test sets after training: average deviation value (i.e., absolute difference between the MLP prediction and the ground truth), average ground truth value and relative average deviation (i.e., average deviation expressed as a percentage of the average ground truth).

	Average deviation value		Average ground truth value		Relative average deviation	
Output	Validation	Test	Validation	Test	Validation	Test
A	0.0149	0.0159	0.552	0.566	2.7%	2.8%
C_{blood}	0.0020	0.0019	0.083	0.074	2.4%	2.6%
C_{mel}	0.0141	0.0154	0.373	0.373	3.8%	4.1%
H (μm)	0.87	0.90	31.9	31.3	2.9%	3.2%

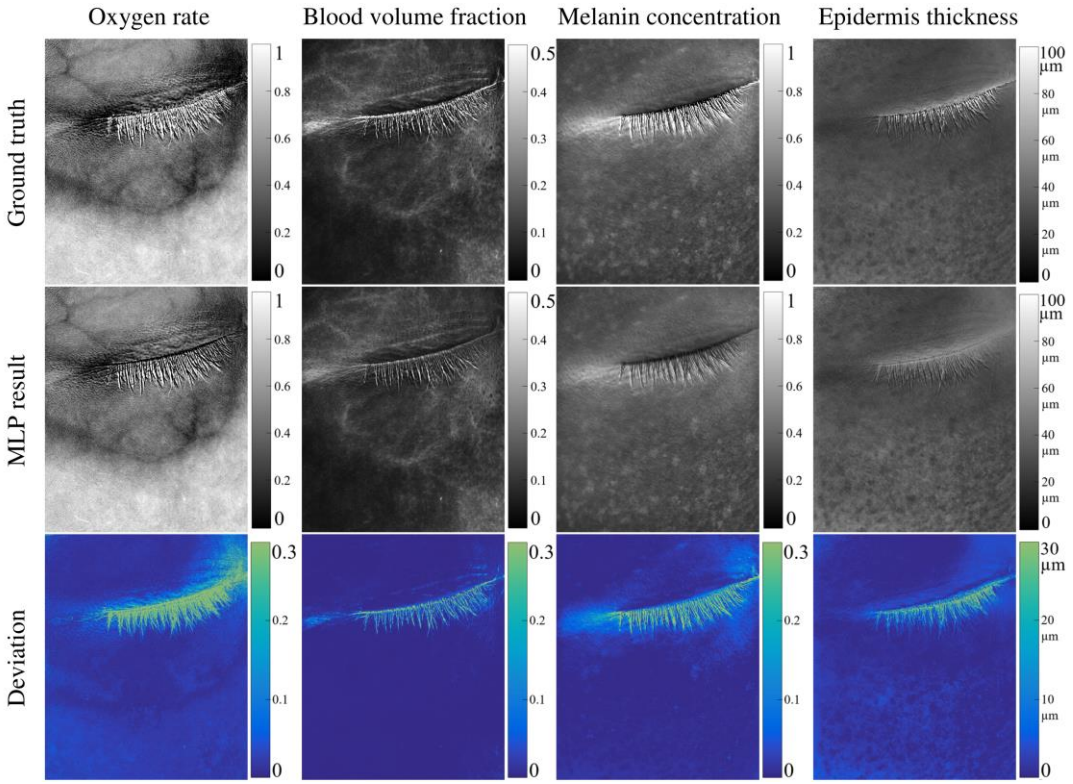

Figure 8. Comparison of skin parameter maps obtained using the optimization-based algorithm (ground truth) and the trained neural network (MLP) on an image acquired of the eye. The deviation between the two results, which corresponds to the absolute difference, is displayed in false colors.

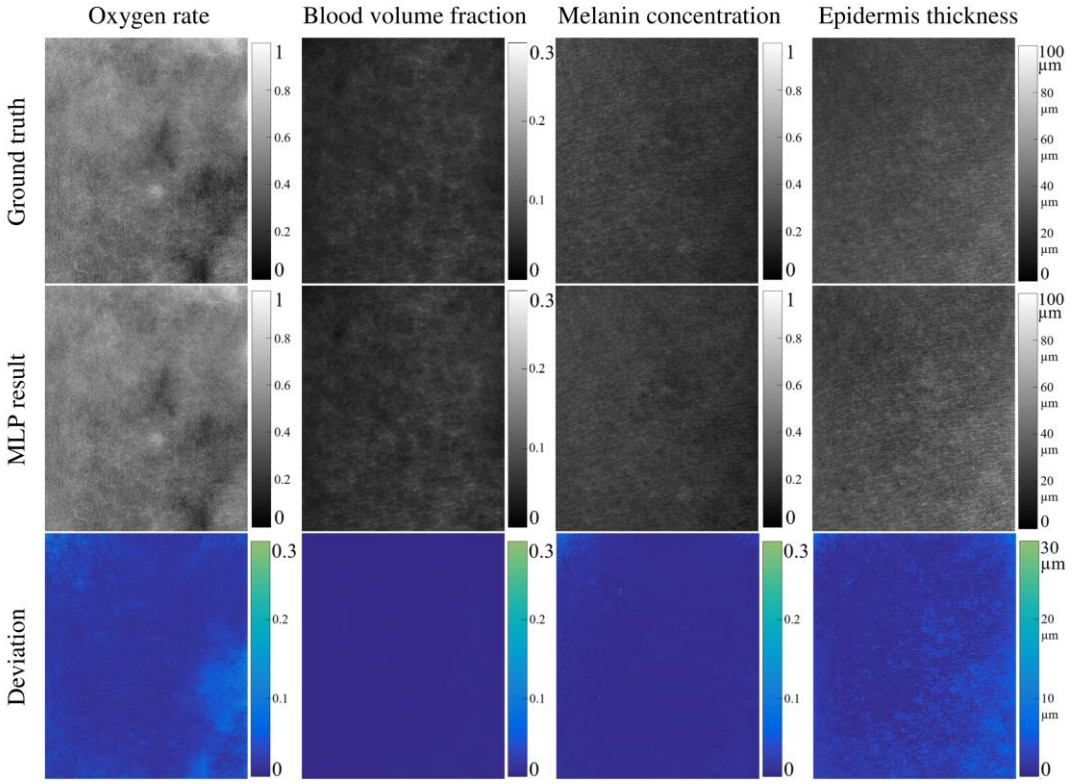

Figure 9. Comparison between skin parameter maps obtained using the optimization-based algorithm (ground truth) and the MLP on an image acquired of the inner forearm, deviation between the two results displayed in false colors.

To the naked eye, the skin parameter maps predicted by the MLP are very similar to the ground truth maps obtained using the optimization-based algorithm. The MLP maps show slightly less contrast than the ground truth maps, however, the difference is barely perceptible except on areas that do not represent "normal" skin, such as on the eyelid (see Figure 8). This is likely due to the fact that the training dataset did not contain any pixels representing it. Deviation is also high on the inside corner of the eye (on the right), a result which was expected since concave areas on the face are affected by interreflections [23], i.e., the fact that light undergoes several successive reflections on different skin areas before forming the radiance captured by the sensor. Spectral radiances issued from concave areas concerned by interreflections present spectral shapes different from the ones issued from flat or convex areas, and the relationship between the two kinds of spectral shapes is strongly non-linear. In absence of model taking explicitly this phenomenon into account, the areas concerned by interreflections cannot be adequately addressed [4].

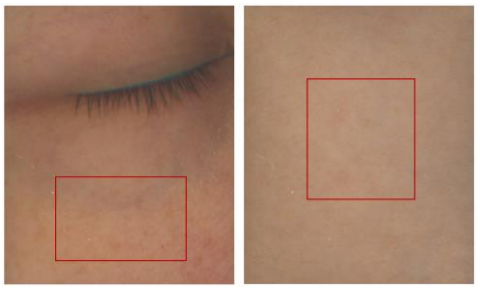

Figure 10. Areas selected under the eye (left) and on the arm (right).

Table 2. MLP results on the area selected underneath the eye (see Figure 10 left): average deviation value (i.e., absolute difference between the MLP prediction and the ground truth), average ground truth value and relative average deviation (i.e., average deviation expressed as a percentage of the average ground truth).

Output	Average deviation value	Average ground truth value	Relative average deviation
A	0.0186	0.647	2.9%
C_{blood}	0.0019	0.092	2.0%
C_{mel}	0.0056	0.309	1.8%
H (μm)	1.42	33.5	4.3%

Table 3. MLP results on the area selected on the inner forearm (see Figure 10 right): average deviation value, average ground truth value and relative average deviation.

Output	Average deviation value	Average ground truth value	Relative average deviation
A	0.0145	0.453	3.2%
C_{blood}	0.0017	0.062	2.8%
C_{mel}	0.0054	0.236	2.3%
H (μm)	1.45	29.9	4.9%

The average deviation calculated on the selected areas (see Tables 2 and 3) allows us to quantify the method's accuracy. In Tables 2 and 3, deviation is also expressed as a percentage: the relative average deviation corresponds to the average deviation divided by the average value of ground truth on the same area. We will focus on this relative deviation to assess the accuracy of the MLP prediction, as it is primarily relative variations on the skin parameter maps, such as between areas of the image or between acquisitions taken at different times, rather than absolute values that are of interest to cosmetology studies. The MLP predictions are satisfactory for oxygen rate, blood volume fraction and melanin concentration, with a relative deviation of no higher than 3.2% across the results. Predictions are also rather satisfactory for epidermis thickness, with less than 5% of relative deviation. The deviation rates calculated on these images are of similar magnitude to those calculated on the test and validation datasets (see Table 1 in Section 4.3). It confirms that the MLP has been correctly trained, as it is able to predict as accurately as possible images that resemble the training dataset. A visual comparison between the predicted maps and the ground truth maps corroborates the low deviation percentages obtained on these results.

5.2 Predicted skin parameter maps on healthy skin

The results presented in Section 5.1 allow an evaluation of the accuracy of the MLP on images that are similar to the training dataset. To evaluate the ability of the MLP to generalize, we applied it on an image representing healthy skin, but differing more markedly from the images in the training dataset. The image selected is of the area under the eye, as for the images in Figure 8, but has been acquired by a different operator, using a different SpectraCam® device, on a person of different skin color. The skin parameter maps corresponding to the ground truth, the MLP predictions as well as the absolute difference between the two are shown in Figure 11. The average deviation between the ground truth images and the MLP images on an area of interest selected underneath the eye (see Figure 12) is given in Table 4.

Table 4 shows the relative deviation obtained for each map on this image, which is roughly twice that of the deviation percentage calculated from the images that are similar to the training dataset. Nevertheless, relative deviation remains below 10%, and below 5% for all quantities but epidermis thickness, which is not too aberrant.

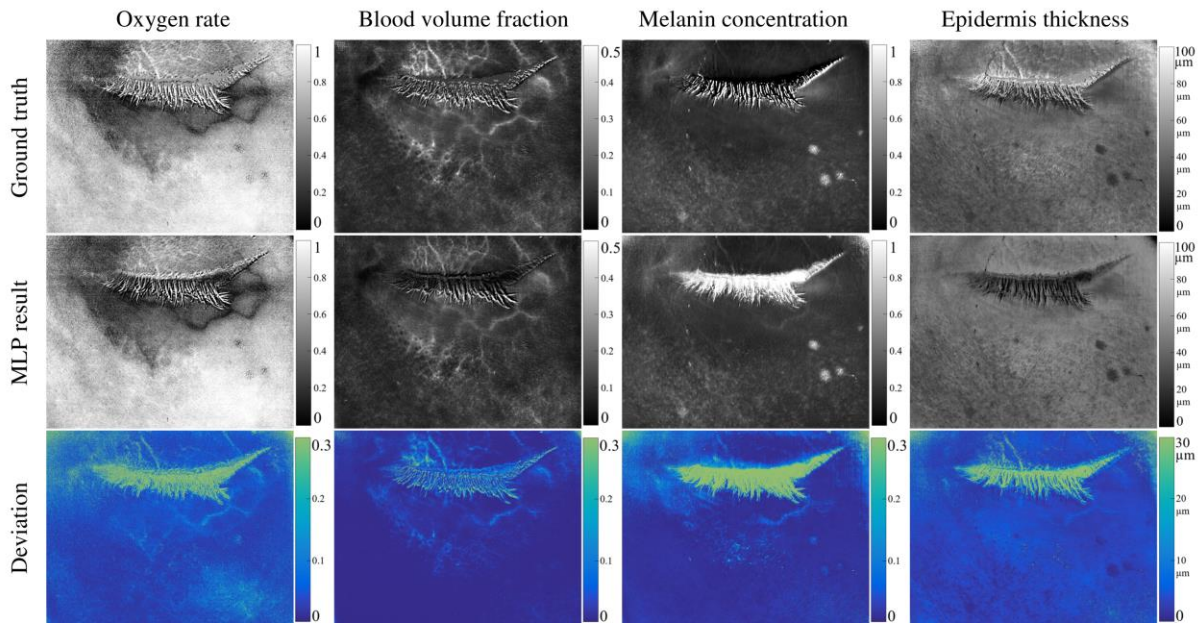

Figure 11. Comparison between the maps of skin parameter obtained using the optimization-based algorithm (Ground truth) and the MLP on an image acquired of the eye, deviation between the two results displayed in false colors.

Regarding the visual observation of the skin parameter maps (see Figure 11), there is little perceptible difference between the ground truth maps and the MLP maps. The level of deviation appears higher on pixels corresponding to parameter values that are either drastically higher or lower than the typical values: for example, deviation is high on the blood vessels located directly underneath the eye, where the blood volume fraction is high and the oxygen rate is low. On the map showing deviation for melanin concentration, we recognize underneath the eye a shape that can be identified as blood vessels. This suggests that on these pixels, the MLP interpreted a signal corresponding to blood as being linked to melanin. This kind of artefact can be problematic for certain applications and represents a limitation to using the MLP method when a very high degree of accuracy is required.

Figure 12. Area selected under the eye.

Table 4. MLP results on the area selected under the eye under the eye (see Figure 12): average deviation value, average ground truth value and relative average deviation.

Output	Average deviation value	Average ground truth value	Relative average deviation
α	0.0436	0.842	5.2%
C_{blood}	0.0073	0.129	5.6%
C_{mel}	0.0141	0.260	5.4%
H (μm)	3.26	39.3	8.3%

5.3 Full face reconstruction and robustness to irradiance drifts

The MLP training dataset was assembled and processed with the aim of training the MLP to be as robust as possible to irradiance drifts. The hyperspectral cameras used in this work relies on black and white reference tiles to calibrate the acquired signal and obtain skin spectral reflectance. When the skin surface is not flat however, this calibration step does not yield spectral reflectance but spectral reflectance multiplied by a factor that is determined by the orientation of the surface [4]. This factor is constant over the spectrum but can vary from one pixel to another, according to the surface shape, a phenomenon that we call irradiance drift. If the skin analysis method were not robust to irradiance drifts, variations in the shape of the surface could be erroneously interpreted as variations in chromophore concentration. Areas that are darker on the acquired image because they receive less light or reflect less light toward the camera, such as the side of the nose for example, could be erroneously interpreted as containing more

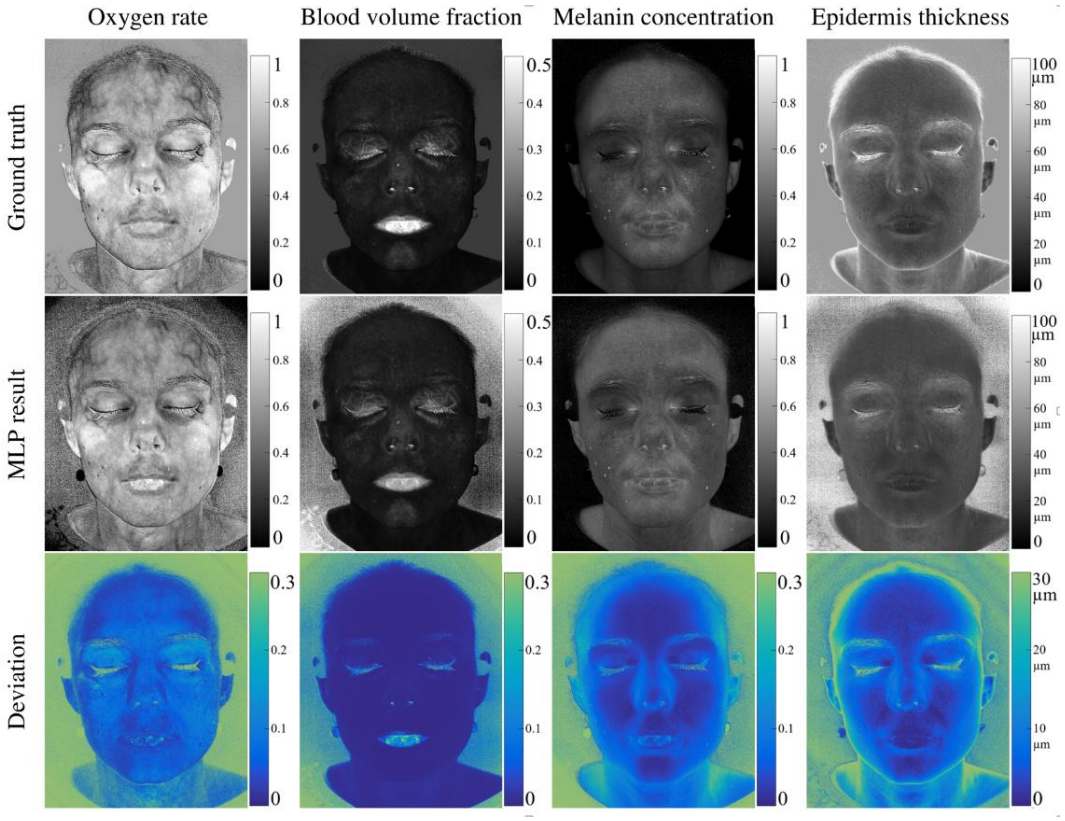

Figure 13. Comparison between skin parameter maps obtained using the optimization-based algorithm (Ground truth) and the MLP on a full-face image, deviation between the two results displayed in false colors.

melanin. The robustness of the MLP to irradiance drifts is especially critical for the analysis of full-face hyperspectral images, as the complex shape of the face makes irradiance drifts unavoidable. For these full-face images, which can comprise up to 4 megapixels, computation time using the optimization-based algorithm is especially high (around five hours). An MLP method able to perform a skin analysis within a short time therefore represents a huge benefit.

The MLP implemented in this work, however, has been trained only on data representing flat areas of skin, with very little irradiance variation in comparison to those which can be observed on a full-face. The training dataset has been supplemented by data on which irradiance drifts have been artificially added, but this provides no guarantee that the MLP can correctly predict skin parameter maps from full face images. The results presented below provide some clarity as to whether the MLP predictions are satisfactory on a full-face image and in which contexts the method can be applied.

The full-face image used in this section, acquired using the SpectraFace®, comprises 1548 × 1948 pixels. Analysis of the image using the MLP took 2 seconds using a GPU and 8 seconds using a CPU (Intel®Core™ i7-6700 CPU 3.40 Ghz), which represents a considerable time gain compared to the optimization-based algorithm analysis.

Figure 13 shows a comparison of the skin parameter maps obtained using the MLP and the ground truth. Visually, the skin parameter maps predicted by the MLP are rather similar to the ground truth maps, with the exception of the map illustrating epidermis thickness. Significantly, there are no

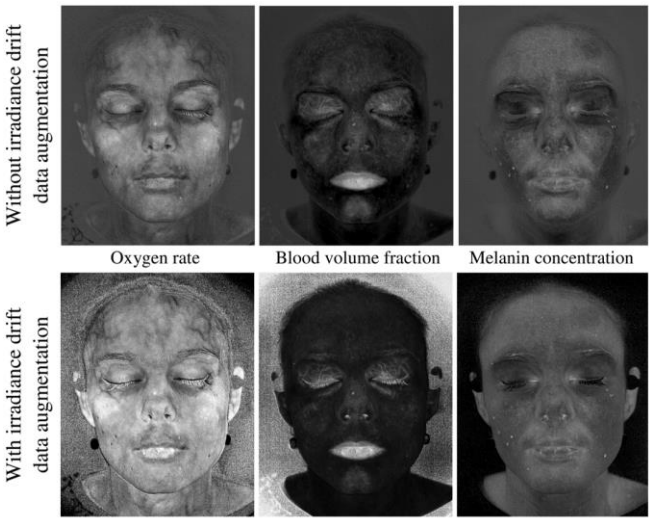

Figure 14. Estimated maps of skin parameter from a full-face hyperspectral image, for a MLP trained on a dataset that has not been augmented by data affected by irradiance drifts (top), compared to the current MLP (bottom), for which irradiance drift data augmentation has been performed.

abrupt variations in the middle of the face, indicating that incidence of irradiance drift has not been mishandled by the MLP. As a comparison, Figure 14 (top) shows skin parameter maps obtained using another MLP network trained on the initial database that was not supplemented with data artificially altered by irradiance drifts: a sudden variation correlated with the shape of the face is clearly visible on the blood volume fraction and melanin concentration maps. The absence of such artefacts from the maps shown in Figure 13 (or Figure 14 bottom) encourages us to think that training the MLP on an augmented dataset data did improve the robustness of the MLP to irradiance drifts.

We can also observe on the melanin concentration and epidermis thickness deviation maps that deviation is higher on the sides of the face. A close look at the ground truth maps shows a diminution of these two quantities on the sides of the face that is not present on the MLP results. As these variations do not correspond to any potential variation in skin properties, they are likely to be artefacts. This indicates that the optimization-based algorithm considered to be the ground truth is, surprisingly, less faithful to reality than the MLP method on the sides of the face. A potential explanation for this phenomenon is that the signal-to-noise ratio is low on those pixels affected by strong irradiance drifts, and the optimization-based algorithm is less robust to noise than the MLP. If it is the case, the MLP method would have an additional advantage over the optimization-based method..

Figure 15. Areas selected on the forehead (A), the cheek (B) and the chin (C).

Table 5. MLP results the three selected areas corresponding to the forehead (A), the cheek (B) and the chin (C) (see Figure 15): average deviation value, average ground truth value and relative average deviation.

	Output	Average deviation value	Average ground truth value	Relative average deviation
Area A	α	0.0610	0.619	9.9%
	C_{blood}	0.0015	0.063	2.0%
	C_{mel}	0.0147	0.355	4.2%
	$H\ (\mu m)$	2.32	30.4	7.6%
Area B	α	0.0840	0.824	10.2%
	C_{blood}	0.0019	0.067	2.1%
	C_{mel}	0.0209	0.298	7.0%
	$H\ (\mu m)$	3.46	29.3	10.6%
Area C	α	0.0700	0.707	9.9%
	C_{blood}	0.0018	0.0976	1.9%
	C_{mel}	0.0137	0.343	4.0%
	$H\ (\mu m)$	1.63	27.3	6.0%

We chose three areas to evaluate relative deviation, illustrated on Figure 15: the middle of the forehead (Area A), the lower part of the cheek (area B), and the middle of the chin (Area C). The average deviation calculated on these areas, shown in Table 5, are not drastically higher than the deviation on a small area shown in Table 4, as their maximal values are around 10%, which is encouraging given the presence of irradiance drifts on full face images. The MLP predictions are even close to the ground truth values for blood volume fraction, with relative deviation below 2%, which is the same as on the test dataset.

The results presented in Section 5 allow us to validate the use of the MLP method for pre-visualization purposes. The percentages of relative deviation calculated on the images outside the training dataset are rather promising given the simplicity of this dataset, which is certainly not representative of the diversity of skin spectral reflectance that can be found on different people. However, the percentage of deviation is not negligible compared to the percentage of variation sometimes observed on the skin parameters in cosmetology studies [5]. For this reason, the current MLP cannot replace the optimization-based algorithm when accuracy is needed.

6. Conclusion

The implementation of a neural network method for the estimation of skin parameter maps can drastically reduce calculation time, allowing for the visualization of results immediately after acquisition. The skin parameter maps obtained using this method closely approximate those computed using a model-based skin analysis method, and they could conceivably be used for the purposes of previewing results for skin phototypes I to III. However, the accuracy of the implemented MLP does not yet meet the needs of image analysis for cosmetology studies. The selection of the dataset seems crucial to ensuring the accuracy of the neural network's predictions. A potential improvement of the method would be to train the network on data acquired on a larger number of patients of different ages, skin color, and living conditions.

In this work, we have demonstrated the importance of accounting for irradiance drifts when working on full-face images, and provided a solution by augmenting the training dataset with data artificially altered by irradiance drifts. The extended training dataset is likely more representative of the spectral reflectance measured by the hyperspectral camera on a full-face. An alternative solution could be to normalize spectral reflectance as a pre-processing, which would cancel the effects of irradiance drifts.

In our approach, the neural network was not trained to model light-skin interactions, but only to solve an inversion problem more quickly than by using optimization-based algorithm. The approach illustrated in Figure 1.a, using a training dataset based on a Monte Carlo method [7–10,24] and a much more detailed model of skin than our two-layer model, could be investigated. In this case, the neural network would provide a model for skin spectral reflectance and solve the inverse problem as well. However, we cannot be guaranteed that this alternative approach would yield satisfactory results on noisy spectral reflectance corresponding to real skin. Conversely, our current approach has the advantage of accounting for real data, acquired with a real optical system that also introduces some noise on the data.

The work presented in this article paves the way to future developments of hyperspectral imaging for skin analysis. A possibility for further developing the acquisition system is to reduce the complexity of the measured data, using a neural network to replace the optimization-based analysis. Currently, the skin analysis method requires high-resolution skin spectral reflectance measurements as an input, as skin parameters are estimated using the shape of the spectrum. High spectral resolutions are thus required to distinguish small variations in the shape of the spectrum, corresponding to small variations in skin optical properties. However, it is possible that not all acquired wavelengths contain useful information, just as it is possible that two different wavelengths could be correlated, making it redundant to measure both. If such is the case, the quantity of acquired data could be reduced by ignoring certain wavelengths, and the analysis method could be replaced by a neural network. This would allow us to reduce acquisition time or even to consider using a multispectral camera rather than a hyperspectral camera. The latter would have many advantages, as a multispectral camera is generally lighter, less fragile, and cheaper than a hyperspectral camera. Such a simplification would require a careful selection of the acquired wavelengths, retaining those that carry the most relevant information to ensure good performances for the neural network method.

References

1. P. J. Matts, B. Fink, K. Grammer, and M. Burquest, “Color homogeneity and visual perception of age, health, and attractiveness of female facial skin,” Journal of the American Academy of Dermatology, 57, 977–984 (2007).
2. G. Lu and B. Fei, “Medical hyperspectral imaging: a review,” Journal of Biomedical Optics, 19, 010901 (2014).
3. P. Seroul, M. Hébert, M. Cherel, R. Vernet, R. Clerc, and M. Jomier, “Model-based Skin Pigment Cartography by High-Resolution Hyperspectral Imaging,” Journal of Imaging Science and Technology, 60 (6), 60404-1 (2017).
4. L. Gevaux, C. Adnet, P. Séroul, R. Clerc, A. Trémeau, J. L. Perrot, and M. Hébert, “Three-dimensional maps of human skin properties on full face with shadows using 3-D hyperspectral imaging,” Journal of Biomedical Optics, 24, 1 (2019).
5. A. Nkengne, J. Robic, P. Seroul, S. Gueheunneux, M. Jomier, and K. Vie, “SpectraCam ® : A new polarized hyperspectral imaging system for repeatable and reproducible in vivo skin quantification of melanin, total hemoglobin, and oxygen saturation,” Skin Research and Technology, 24, 99–107 (2018).
6. L. Wang, X. Zhao, and S. L. Jacques, "Computation of the optical properties of tissues from light reflectance using a neural network," Laser-Tissue Interaction V; and Ultraviolet Radiation Hazards (International Society for Optics and Photonics), 2134, 391–400 (1994).
7. M. Jäger, F. Foschum, and A. Kienle, “Application of multiple artificial neural networks for the determination of the optical properties of turbid media,” Journal of Biomedical Optics, 18, 057005 (2013).
8. S. J. Wirkert, H. Kenngott, B. Mayer, P. Mietkowski, M. Wagner, P. Sauer, N. T. Clancy, D. S. Elson, and L. Maier-Hein, “Robust near real-time estimation of physiological parameters from megapixel multispectral images with inverse Monte Carlo and random forest regression,” International Journal of Computer Assisted Radiology and Surgery, 11, 909–917 (2016).
9. S. Panigrahi and S. Gioux, “Machine learning approach for rapid and accurate estimation of optical properties using spatial frequency domain imaging,” Journal of Biomedical Optics, 24, 1 (2018).
10. E. Zherebtsov, V. Dremin, A. Popov, A. Doronin, D. Kurakina, M. Kirillin, I. Meglinski, and A. Bykov, “Hyperspectral imaging of human skin aided by artificial neural networks,” Biomedical Optics Express, 10, 3545 (2019).
11. Y. Garini, I. T. Young, and G. McNamara, “Spectral imaging: Principles and applications,” Cytometry Part A, 69A, 735–747 (2006).
12. P. Kubelka and F. Munk, “An Article on Optics of Paint Layers,” Z. Tech. Phys, 12, 593-601 (1931).
13. S. L. Jacques, “Optical properties of biological tissues: a review,” Physics in Medicine and Biology, 58, R37–R61 (2013).
14. S. L. Jacques, <https://omlc.org/spectra/melanin/>.
15. S. Prahl, <https://omlc.org/spectra/hemoglobin/>.
16. P. Kubelka, “New Contributions to the Optics of Intensely Light-Scattering Materials Part II: Nonhomogeneous Layers*,” Journal of the Optical Society of America, 44, 330 (1954).
17. J. L. Saunderson, “Calculation of the Color of Pigmented Plastics,” Journal of the Optical Society of America, 32, 727–736 (1942).
18. R. Yuhas, A. Goetz, and J. Boardman, “Discrimination among semiarid landscape endmembers using the spectral angle mapper (sam) algorithm,” JPL Publication, 147–149 (1992).
19. A. K. Jain, J. Mao, and K. M. Mohiuddin, “Artificial Neural Networks: A Tutorial,” Computer, 3, 31–44 (1996).
20. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," arXiv:1502.03167 [cs] (2015).
21. M. W. Gardner and S. R. Dorling, “Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences,” Atmospheric Environment, 32, 2627–2636 (1998).
22. P. Werbos, “Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences”, PhD Diss (1974).
23. D. Forsyth and A. Zisserman, “Mutual illumination,” Proceedings CVPR'89; IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 466–473 (1989).
24. R. Deeb, J. Van de Weijer, D. Muselet, M. Hebert, and A. Tremeau, “Deep spectral reflectance and illuminant estimation from self-interreflections,” Journal of the Optical Society of America A, 36, 105 (2019).