Data in Brief 53 (2024) 110163

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A dataset of optical spectra and clinical features acquired on human healthy skin and on skin carcinomas

Thomas Elsen[a], Clément Fauvel[a], Grégoire Khairallah[a,b], Ahmed Zghal[a], Alain Delconte[a], Valentin Kupriyanov[a,c], Walter Blondel[a], Marine Amouroux[a,*]

[a] *Université de Lorraine, CNRS, CRAN UMR 7039, Vandoeuvre-lès-Nancy, 54500, France*
[b] *Metz-Thionville Regional Hospital, Department of plastic, aesthetic, and reconstructive surgery, Ars-Laquenexy, 57530, France*
[c] *Laboratory for Remote Sensing of the Environment, V.E. Zuev Institute of Atmospheric Optics SB RAS, Tomsk, Russia*

ARTICLE INFO

Article history:
Received 2 October 2023
Revised 30 January 2024
Accepted 31 January 2024
Available online 6 February 2024

Dataset link: SPECTROLIVE : OPTICAL SPECTRA (AUTOFLUORESCENCE AND DIFFUSE REFLECTANCE) ACQUIRED ON HUMAN SKIN CARCINOMAS (Original data)

Keywords:
Optical spectroscopy
Spatial resolution
Human skin cancer
Autofluorescence
Diffuse reflectance
Actinic keratoses
Squamous cell carcinomas
Basal cell carcinomas
Healthy skin
Clinical study
In vivo measurement

ABSTRACT

Optical spectroscopy is studied to contribute to skin cancer diagnosis. Indeed, optical spectra are modified along cancer progression and provide complementary information (e.g., on metabolism and tissue structure) to clinical examination for surgical guidance [1,2]. The current original dataset is made of autofluorescence and diffuse reflectance spectra acquired *in vivo* on 131 patients' skin with the SpectroLive device [3,4]. Spatially-resolved spectroscopy measurements were performed using a multi-fiber optic probe featuring 4 distances (0.4–1 mm) between excitation and collection optical fibers: spatial resolution allows spectra acquired at different distances to carry information from different depths in skin tissues. Five types of autofluorescence spectra were acquired using five different wavelength excitations (on the 365–415 nm spectral range) in order to collect information on several skin endogenous fluorophores (e.g., flavins, collagen). A sixth light source (white broadband) was used to acquire diffuse reflectance spectra carrying information about skin scattering properties and skin endogenous absorbers such as melanin and hemoglobin. Patients were proposed to

* Corresponding author.
E-mail address: marine.amouroux@univ-lorraine.fr (M. Amouroux).

https://doi.org/10.1016/j.dib.2024.110163

be included into the clinical trial if they were suspected of suffering from actinic keratoses (precancerous skin lesions) or from basal or squamous cell carcinomas: in all cases, complete diagnostics is provided in the dataset. To increase the interest of the dataset and evaluate the dependence of optical spectra (intensity, shape) not only on pathological states but also on healthy skin features (civil age, skin age, gender, phototype, anatomical site), spectra were acquired for all 131 patients on two so-called "reference" skin sites known to rarely suffer from skin cancer: palm of the hand (featuring a thick skin type) and inner wrist (featuring thin skin). Spectra are available in .tab files: first column displays the spectral range on which intensity spectra were recorded (317–788 nm) and each following column provides an intensity spectrum acquired by each spectrometer for a given combination of light source excitation and distance. Each of the 131 folders corresponding to each of the 131 patients contains a .json file providing patients clinical features: gender, civil age, skin age, phototype score and class. All .tab files names include anatomical site and anatomopathological diagnostics of the skin site on which spectra were acquired: codes were defined to match a letter or an acronym to each diagnostic and anatomical site. To ensure quality control, a spectrum was acquired on the same calibration standard before starting spectra acquisition on each patient. It is therefore possible to follow the impact of the acquisition optical chain ageing during the 4.5 years that the patients were included. This dataset can be used by epidemiologists for the characterization of populations affected by skin cancers (gender ratio, mean age, anatomical sites typically affected, etc.); it may also be used by researchers in artificial intelligence to develop innovative methods to process such data and contribute to non-invasive diagnostics of skin cancers whose incidence is steadily increasing.

Specifications Table

Subject	Biomedical engineering
Specific subject area	Fibered optical spectroscopy for in vivo skin cancer diagnosis
Data format	Raw data A folder entitled "Preprocessing_plotting_programs" contains a program developed in python language (pre_processing.py) to preprocess raw spectra available in .tab files in the dataset.
Type of data	Table: .tab files (dataset with numbers) Json files: .json (dataset with labels)
Data collection	Spectra were acquired from November 2016 to March 2021 on 131 patients by the SpectroLive device developed by our team, patended [4] and fully described in the related research article [3]. Once patients agreed on being included into the SpectroLive clinical trial (national clinical trial identifier from clinicaltrials.gov: NCT02956265), the optical probe was put in gentle contact with patients' skin. On each skin site, three sets of 24 spectra (corresponding to three .tab files) were acquired in a row in approximately 30 s during which the probe was kept in contact with the skin.

(*continued on next page*)

Data source location	Data collection • Institution: Metz-Thionville Regional hospital • City: Ars-Laquenexy • Country: France. • Latitude and longitude: 49.084499,6.240191 Data storage • Institution: Université de Lorraine • City: Nancy • Country: France • Latitude and longitude: 48.696048, 6.1763405
Data accessibility	Repository name: Dorel Data identification number: https://doi.org/10.12763/EYVX3P Direct URL to data: https://dorel.univ-lorraine.fr/dataset.xhtml?persistentId=doi:10.12763/EYVX3P [5].
Related research article	W. Blondel, A. Delconte, G. Khairallah, F. Marchal, A. Gavoille, M. Amouroux, Spatially-Resolved Multiply-Excited Autofluorescence and Diffuse Reflectance Spectroscopy: SpectroLive Medical Device for Skin In Vivo Optical Biopsy, Electronics. 10 (2021) 243. https://doi.org/10.3390/electronics10030243. [3]

1. Value of the Data

- The dataset will be useful for researchers in artificial intelligence: it may be used as a training set for machine-learning-enabled analytical approaches. Researchers in biomedical engineering, like teams involved in the development of optical biopsy methods, will benefit from this dataset at it includes a wide range of different types of spectroscopic information acquired at four different distances between excitation and collection optical fibers using several excitation wavelengths. This allows to retrieve information from several endogenous skin compounds. Researchers in epidemiology will take advantage of the complete set of clinical features used to label the dataset: gender, skin and civil age, phototype score and class, anatomical site, complete diagnostics.
- These data are useful because they were acquired in vivo both on patients' healthy skin and on pathological skin sites:
  - Data acquired on healthy skin sites are useful for studying skin optical properties dependence on anatomical sites (skin composition and thickness greatly vary from one site to another) and on individual physiological variations such as phototype, ageing and gender. They can also be used for studying inter-individual variation within a subpopulation: e.g., variation of spectra acquired on hand palm within phototype II females aged 80–90.
  - Data acquired on pathological skin sites (carcinomas) are useful in developing data processing methods for non-invasive skin cancer diagnostics. Those data greatly benefit from corresponding data acquired on healthy skin to carry studies on optical spectra standardization which is at stake for clinical development of optical methods.
- The clinical cohort used to acquire data is made of 131 patients that display typical characteristics of a population affected with skin carcinomas in terms of age, phototype and anatomical location of skin carcinomas. A valuable study in the field [6] evaluated optical spectroscopy diagnostic accuracy on a set of 57 carcinomas (138 available in the current dataset) and 14 actinic keratoses (48 in the current dataset).

2. Background

The primary objective for creating this dataset was to gather the most complete possible set of optical spectra carrying information on three types of interactions between light and skin tissues: endogenous fluorescence, scattering and absorption. To do so, data were acquired using six different light sources and four different distances between excitation and collection fibers. To our knowledge, this dataset is the most complete available in the community of research teams working on skin cancer diagnosis using optical spectroscopy.

Table 1
Clinical characteristics of the population included in the clinical study (131 patients, 229 lesions).

Characteristics	Number	Percentage
Gender		
Male	91	69
Female	40	31
Age		
30-50	2	2
50-60	8	6
60-70	23	18
70-80	35	27
80-90	48	37
90-100	15	11
Phototype		
I	15	13
II	77	66
III	23	20
IV	2	2
V-VI	*None*	0
Diagnostics		
Carcinomas		
Basal Cell Carcinomas (BCC)	86	38
Squamous Cell Carcinomas (SCC)	52	23
Actinic keratoses (AK1, AK2 and AK3)	48	21
Trichoblastic carcinomas	3	1
Healthy (inflammatory cicatricial changes, elastosis, etc.)	27	12
Others (seborrheic keratoses, malignant melanoma, etc.)	13	6
Anatomical sites		
Face (cheeks, forehead, nose, ears, chin, temples, etc.)	167	73
Vertex (men only)	29	13
Trunk (back or abdomen)	16	7
Limbs (upper or lower)	17	7

3. Data Description

3.1. Dataset clinical description

The dataset was put in the form of a relational database in SQLite format linking all patients' features given in Table 1 to spectroscopic data contained in .tab files: gender, age and phototype are provided in a .json file located inside each patient folder. All numbers mentioned in Table 1 were automatically retrieved from the database. Table 1 provides clinical characteristics of the population included into the clinical trial. About two thirds of the population made of 131 patients were males (91 males); mean age was 77 years (median: 79, min: 32 and max: 98 years). Most patients were fair skinned (79% featured phototypes I or II): the total number of phototypes mentioned in Table 1 is 117 (<131) because some patients were not able to answer questions from the Fitzpatrick questionnaire used to determine phototype. In such cases, phototype was simply not recorded. Although skin carcinomas rarely appear on phototype IV patients, two of them were included into the study: such two patients were taking post-transplant immunosuppressive treatment at the time they were included into the study. A total of 229 lesions were included into the study as most patients suffered from more than one lesion suspected of being a skin carcinoma (mean: 1.7 lesion/patient): 61% of lesions were diagnosed as carcinomas, 21% as actinic keratoses (precancerous lesions) and 12% were diagnosed as healthy. Indeed, some patients included into the study had already undergone biopsy performed by a dermatologist. After dermatologists received diagnostics mentioning either carcinomas or actinic keratosis, they referred patients to the plastic surgeon for complementary resection especially if the carcinoma was located on anatomical sites that require specific surgical methods (e.g., skin graft). Sometimes the biopsy performed was enough to remove the entire cancerous lesion; sometimes

it was not, but local anesthesia performed prior to the biopsy may have inhibited the growth of remaining cancer cells [7]. Therefore, for 12% of the patients included into the study, no cancerous tissue was left in the skin sample removed by the surgeon and the sample was classified as "healthy" (i.e., no cancerous feature could be detected); complementary diagnostic information is though provided in each case when inflammatory cicatricial changes (ICC) due to a previous biopsy or elastosis (EL) were noticed by the anatomopathologist.

3.2. Folders organization

The dataset contains 132 folders:

- 131 patients' folders identified with a unique combination (e.g., 100_059) of a number identifying the folder created on the device each time a new patient was included into the study (e.g., 100) and the inclusion number of the patient into the clinical study (e.g., 059), and
- one folder containing preprocessing and plotting programs.

Patients' folders typically contain 6 subfolders (Fig. 1):

- EX1_C1_STD_530 contains intensity spectra acquired on the calibration standard Cal_STD (Ref. SRS-99-010, Labsphere) used for quality insurance and diffuse reflectance calculation.
- EX1_L1_Cl_524 contains three files: 524_[1,2,3]_L_Cl_iBCC_100_059.tab acquired on the suspected lesion skin site (L) on the patient's left temple (corresponding letter to temple: C and "l" for left side); Diagnostics is: invasive basal cell carcinoma (iBCC).
- EX1_L1_Cl_527 contains three files: 527_[1,2,3]_NL_Cl_H_100_059.tab acquired on the first non-lesional (NL) skin site located close to the previously mentioned iBCC and confirmed as healthy (H) by anatomopathology.
- EX1_L1_Cl_528 contains three files: 528_[1,2,3]_NL_Cl_H_100_059.tab acquired on the second non-lesional (NL) skin site located close to the previously mentioned iBCC.
- EX1_R1_V_522 folder contains three files: 522_[1,2,3]_Ref_V_100_059.tab acquired on the first (R1) reference skin site (V: palm hand).
- EX1_R2_T_523 folder contains three files: 523_[1,2,3]_Ref_T_100_059.tab acquired on the second (R2) reference skin site (T: inner wrist).
- All .tab files are named according to the same nomenclature: databaseID_spectrumID_measurement-site_anatomical-site_diagnosis_fileID
- databaseID: a meaningless unique identifier attributed to each folder in the database (e.g., 524, Fig. 1),
- spectrumID: 1, 2 and 3 identifies each of the three files that correspond to each of the three spectroscopic acquisitions acquired successively in a row (the optical probe was kept in contact with the skin surface for all three acquisitions) on the same skin site, within 30 seconds approximately.
- Measurement-site: NL or L
  - NL: spectra were acquired on non-lesional (NL) skin sites (Fig. 4, b)
  - L: spectra were acquired on a lesion (L) skin site i.e., suspected of being a carcinoma (Fig. 4, a and b).
- Anatomical site: each letter corresponds to an anatomical site as defined in our self-developed code (Tables 2 and 3).
- Diagnostics: each acronym corresponds to a specific diagnostic as subjectively defined in our self-developed dictionary (Tables 4 and 5). In the current example iBCC means: invasive basal cell carcinoma.

Diagnostics were provided by the anatomopathology department of the hospital where the clinical study took place according to the clinical routine procedure already widely described in the literature. After fixation in formaldehyde for at least 24 hours, the skin samples were cut macroscopically in order to identify the sites of spectroscopic measurements in order to match

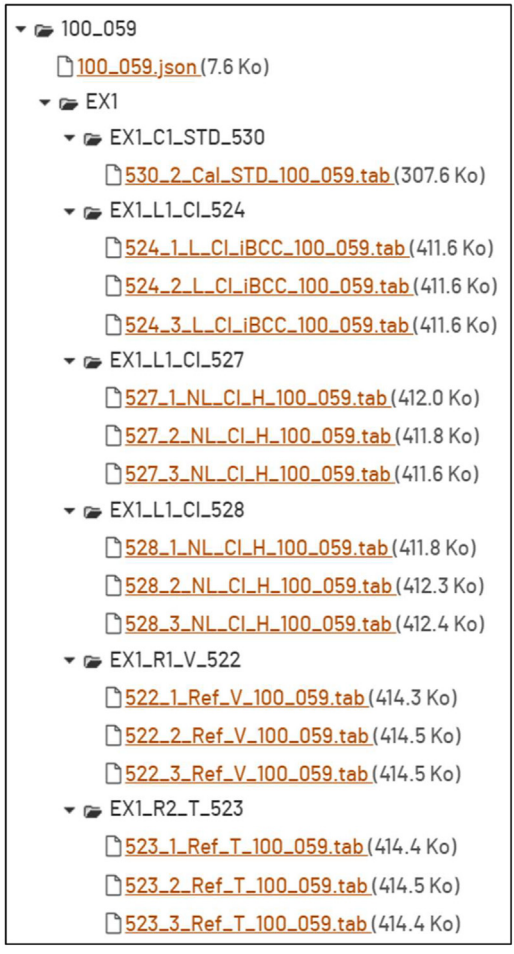

Fig. 1. Typical organization of a patient folder (e.g., 100_059) containing one .json file providing patient's clinical features and 6 subfolders: one containing spectra acquired on the calibration standard (Cal_STD), three subfolders containing spectra acquired on the surgically-removed skin sample (lesion: L site and two non-lesional skin sites: NL for the first lesion to diagnose: EX1) and two subfolders containing spectra acquired on reference skin sites (R1: palm of the hand and R2: inner wrist).

Table 2
Anatomical sites codification.

Code	Location	Code	Location	Code	Location
A	Vertex	J	Neck anterior side	U	Dorsal hand
AA	Scalp area	K	Neck posterior side	V	Palmar hand
B	Forehead	L	Shoulder and clavicle	W	Anterior thigh
C	Temple	M	Upper third back	X	Posterior thigh
CC	Ear	N	Lower third back	Y	Half proximal anterior leg
D	Eyelid	O	Thorax upper-anterior third	Z	Half proximal posterior leg
E	Nose	P	Thorax two thirds inf-abdomen	ZA	Half distal anterior leg
F	Cheek	Q	Front arm	ZB	Half distal posterior leg
G	Upper lip	R	Dorsal arm	ZC	Foot ankle dorsum
H	Lower lip	S	Dorsal forearm	ZD	Foot ankle plantar
I	Chin	T	Palmar forearm		

Table 3
Anatomical site (secondary location).

Code	Secondary location (English)
L	Left
R	Right
I	Inferior
S	Superior
O	Old
N	New
T	Top
B	Bottom

Table 4
Diagnostic (level 1).

Code	Diagnostic level 1
BCC	Basal cell carcinoma
isSCC	Squamous cell carcinoma in situ
invSCC	Infiltrative Squamous cell carcinoma
K	Keratosis
TR	Trichoblastic carcinoma
H	Healthy
ELSE	Other

Table 5
Diagnostic (level 2).

Code	Diagnostic level 2	Code	Diagnostic level 2
iBCC	Infiltrative BCC	AK	Actinic keratosis
nBCC	Nodular BCC	AK1	Actinic keratosis stg.1
supBCC	Superficial BCC	AK2	Actinic keratosis stg.2
sclBCC	Sclerodermiform BCC	AK3	Actinic keratosis stg.3
mdSCC	Moderately Differentiated SCC in situ	SK	Seborrheic keratosis
wdSCC	Well Differentiated SCC in situ	CC	Cicatricial changes
mdISCC	Moderately Differentiated Infiltrative SCC	ICC	Inflammatory cicatricial changes
wdISCC	Well Differentiated Infiltrative SCC	EL	Elastosis

the corresponding diagnostics. Once cut, the skin samples were dehydrated in alcohol baths of increasing concentrations, then embedded in paraffin before being cut into 5 micrometer-thick sections. Once placed on a glass slide, the sections were placed in an automated machine for being stained with hematoxylin and eosin (H&E). The stained slides were then examined under a microscope by a pathologist who provided the diagnosis indicated in the files names of the dataset.

- FileID: as already described, 100_059 corresponds to the main folder identification dedicated to each patient made of deviceID (100) and inclusionID (059).

In some cases, spectra could be acquired only on one NL skin site, for instance when the suspected lesion was located on a temple close to hair or on the forehead close to eyebrows: since hairs behave as an optical screen, spectra were not acquired on the NL site located in hairs. In such cases (e.g., folder 187_122), only one NL subfolder appears and the patient folder contains five subfolders, not six.

Each patient folder (e.g., 100_059) contains a .json file (Fig. 2) providing patients features:

- Patient_Tag: e.g., 100_059,
- Number of exams (EX): 1, 2 or 3,
- Gender: male (M) or female (F),

Fig. 2. Example of a .json file providing patients' clinical features: patient tag (unique identifier in the dataset), number of exams corresponding to the number of lesions being treated, gender (M: male), age (= inclusion year – birth year), skin age (evaluated using 6 skin criteria), Fitzpatrick score (total score obtained for the questionnaire) and corresponding phototype.

- Age: this score is obtained by subtracting the birth year from the inclusion year e.g., in the following example (Fig. 2), the patient was born in 1935 and he was included into the study in 2021. Therefore, his "civil age" is 86 (86 = 2021 – 1935).
- Skin age [8,9]: this corresponds to what is also called "Apparent skin age" in the scientific literature.

Skin age was determined using the Merz scale [10]. Six criteria were chosen by the plastic surgeon in charge of the skin age evaluation according to his experience of different anatomical sites typical skin ageing: neck volume, forehead lines, hands, upper cheek fullness, nasolabial folds and lip wrinkles. In order to provide quantitative results, evaluations on a 5-level scale (0–4) for each of the 6 criteria were combined. A specific weight was assigned to each level assigned to a criterion: 2.7 for the first level (0) and 13.5 to the fifth level (4). Therefore, if all criteria from a patient were scaled at the first level (all scores were scaled as 0), the patient's biological age was determined as 16 years old ($2.7 \times 6 = 16.2$). If all criteria from a patient were scaled at the fifth level (all scores were scaled as 4), the patient's skin biological age was determined as 81 years old [11].

- Total Fitzpatrick score and phototype were evaluated thanks to the Fitzpatrick questionnaire [12].

A score between 0 and 4 (0, 1,2,3 or 4) is awarded to each answer provided by the patient to each of the nine questions of the questionnaire. Patients who obtain a total score from 0 to 7 belong to phototype I, a score 8-16 corresponds to phototype II, 17-25 to phototype III and 26-30 to phototype IV (a score greater than 30 corresponds to phototypes V or VI but it was not the case of any of the patients included in the current study).

3.3. Spectra files organization

Each file contains 1822 rows and 29 columns. Rows are organized as follows:

- The first row displays headers (e.g., wavelengths, AF1_1) that each identifies one of the 29 columns,
- Each of the next 1821 rows provides the wavelength (from 341 to 784 nm, in 0.24 nm-steps) at which light intensity is measured by spectrometers.

Columns are organized as follows:

Table 6
Light sources codification used in .tab files headers. AF excitation light sources were LEDs providing narrow light band (featuring 10 nm-full- width at half maximum).

Code	Light source
Dark	All light sources switched off
AF1	365 nm
AF2	385 nm
AF3	395 nm
AF4	405 nm
AF5	415 nm
White	Broadband white light (340-785 nm)

- Column 1 provides wavelengths expressed in nanometers (nm) from 341 to 784 nm,
- Columns 2–5 provide dark electronic noise recorded on each spectrometer: Dark1 to Dark4,
- Columns 6–9 provide autofluorescence (AF) spectra measured when using the first light source (AF1 = 365 nm, Table 6) by each of the 4 spectrometers (at 4 different distances from the excitation optical fiber): AF1_1 to AF1_4,
- Columns 10–25 provide AF spectra measured when using the 4 other light sources (AF2, AF3, AF4 and AF5 respectively) by each of the 4 spectrometers: AFX_1 to AFX_4 (X = 2, 3 or 4),
- Columns 26–29 provide spectra measured when using the white light source by each of the 4 spectrometers: White1-White4.

3.4. Preprocessing and plotting programs

The preprocessing and plotting folder contains programs (pre_processing.py that refers to another called listshandling.py) developed in Python language. In order to make it easily accessible (readable) in any numerical environment, the code is made available in three formats: html, pdf and ipynb.

Two .tab files are also available in this folder:

- Spectral_correction_coefficients.tab provides correction coefficients used for preprocessing raw data according to each light source S (S1: 365 nm, S2: 385 nm, S3: 395 nm, S4: 405 nm and S5: 415 nm) and each distance D from excitation optical fiber to collection optical fibers (D1: 400 µm, D2: 600 µm, D3: 800 µm and D4: 1000 µm.
- [Pre_Processed]-27_L_Al_wdlSCC_32_18.tab provides an example of a preprocessed file chosen among the 2767 tab files providing raw data in the dataset.

Preprocessing and plotting programs can be used to plot raw and preprocessed data (Fig. 3).

4. Experimental Design, Materials and Methods

The clinical trial was led by a plastic surgeon in the Metz-Thionville Regional Hospital (France) from November 3rd, 2016 to March 25th, 2021 on patients included into the SpectroLive clinical trial. Inclusions criteria of patients were age (18 years old minimum), autonomy (being in full possession of civil rights), social insurance and skin affection suspected of being a carcinomatous-like skin lesion. The approximative duration of a spectroscopic acquisition was about 3 minutes for each patient. This duration included 3 acquisitions in a row of a 24-spectra set on the calibration standard (Ref. SRS-99-010, Labsphere®) and on 5 skin sites: on two reference skin sites as well as on three skin sites (one L and two NL) located in the surgically resected skin sample (Fig. 4). During spectroscopic acquisition, light was switched off in the surgery room to ensure that no stray light could interfere with light collected by the optical fiber bundle tip put in gentle contact with each patient's skin surface. When clicking on the "acquisition" button on the control software, four "dark" spectra were automatically acquired (one for each of the

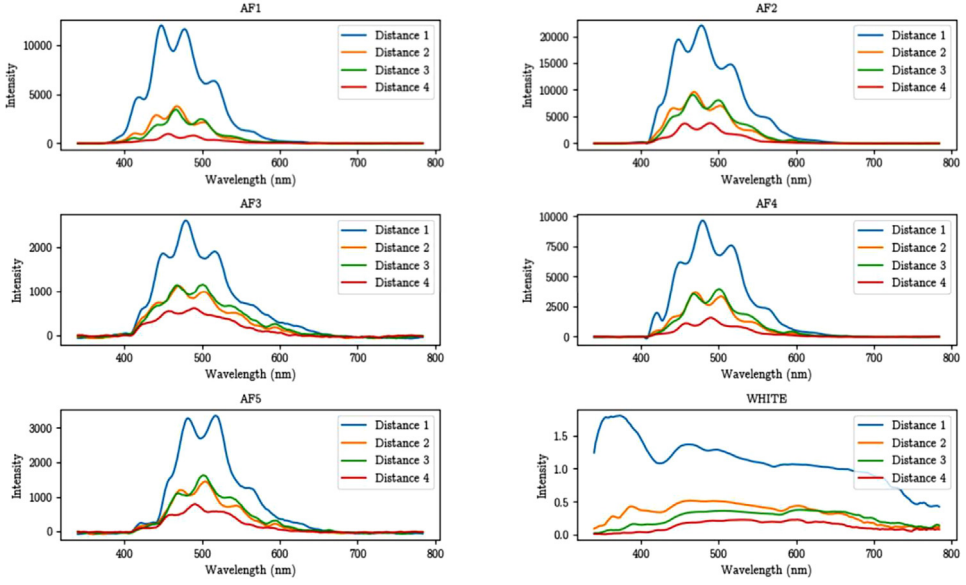

Fig. 3. Preprocessed data plots for first (AF1: 365 nm) and second (AF2: 385 nm) excitation light sources (first row); Third (AF3: 395 nm) and fourth (AF4: 405 nm) excitation light sources (second row); fifth (AF5: 415 nm) and white light sources (third row). Blue: distance 1 (D1=400 µm), yellow: D2 = 600 µm, green: D3 = 800 µm, and red: D4 = 1000 µm

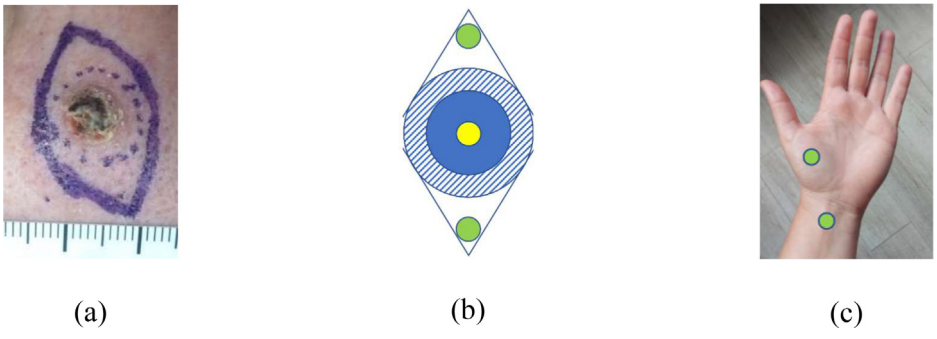

Fig. 4. Picture (a) and scheme (b) of a typical surgical spindle used to resect skin carcinomas. Spectroscopic measurements provided in the dataset were acquired on lesions (L site: yellow disk), non-lesional (NL) parts of the spindle (green disks) and on reference skin sites (c) : palmar hand (R1) and inner wrist (R2).

spectrometers): dark spectra correspond to spectrometers electronic noise i.e., intensity measured when no light is injected on the spectrometer's entrance slit. Such dark intensity is used in the preprocessing program made available online with the dataset. After spectra acquisition, the skin spindle was drawn on the patient's skin surface (so ink did not interfere with optical spectra acquisition) then surgically resected and evaluated by an anatomopathologist who provided the complete diagnosis. As shown in Fig. 4 a), in most cases, surgical resection was made according to the shape of a spindle made of three types of skin regions: the lesion itself, the safety margin and the edge of the spindle that is clinically evaluated as "healthy" skin and that is removed for the sake of a clean surgical suture. For all patients, spectroscopic data were first acquired on the lesion suspected of being a skin carcinoma, which was called the lesion (L) site. Spectroscopic data were then acquired on both sides of the surgical spindle, called non-lesional (NL) sites. For sake of standardization, spectroscopic data were also acquired for all patients on thenar eminence (palm of the hand) and on inner wrist (Fig. 4, c) that were not biopsied.

```
% L_n: light sources; n =1 to num_light;
% S_m : spectrometers; m = 1 to num_spectro;
% P_j : filter wheel position;
% j = 1: 364 nm-high pass filter; j = 2: 405 nm-high pas filter; j = 3: 442 nm-high pass filter; j = 4: neutral density;
num_light = 6;
num_spectro = 4;
Sm_integration_time = 1s;
for every patient included do
    increment folder ID: X;                          % e.g. X=100
    patient's inclusion number manual fill in: Y;    % e.g. Y=059
    create folder: X_Y;                              % e.g. 100_059
    i=1;                                             % : number of measurements for the patient
    for every measurement launched do
        create file 'X_Y_measurement_i';             % e.g. 100_059_measurement_1.csv
        fill in column 1 with wavelengths
            line 1: header 'Wavelengths (nm)';
            lines 2 to 1822: 341 to 784.76 nm;
        end
        for m = 1 to num_spectro do
            record dark noise detected by Sm in column (m + 1);
        end
        for n = 1 to num_light
            switch on Ln;
            if n = 1 do P1;
            if n = 2,3,4 do P2;
            if n = 5 do P3;
            If n = 6 do P4;
                for m = 1 to num_spectro
                    record back-reflected light intensity in columns 1 + 4n + m
                end
            switch off Ln;
        end
        record file.
    end  i = i + 1;
end
```

Fig. 5. Custom script algorithm of the home-made software developed in Delphi Pascal language to control optical materials and record spectroscopic measurements. L_n: light sources; n = 1 to 6; each light source emits a specific narrow light band (LED for n=1 to5) or a braodband white light (n=6). L_1: 365 nm; L_2: 385 nm; L_3: 395 nm; L_4: 405 nm; L_5: 415 nm; L_6: broadband white light; S_m; m = 1 to 4; spectrometers; each spectrometer records light at a given distance from the excitation optical fiber. S_1: 400 µm; S_2: 600 µm; S_3: 800 µm; S_4: 1000 µm.

A dedicated home-made software was developed in Delphi Pascal language to control optical materials (spectrometers integration time, light sources switch on and off, etc.), record spectroscopic measurements and create corresponding files and folders. Fig. 5 displays the algorithm of the custom script used for data acquisition: the custom script itself is available on the repository with the dataset in a file entitled "SpectroLive_custom_script_data_acquisition.pdf".

Limitations

Not applicable.

Ethics Statement

The SpectroLive clinical trial was carried out in accordance with the Declaration of Helsinki: relevant informed consent was obtained from each patient before inclusion into the study.

The SpectroLive clinical trial (NCT0295626) was approved by the French National Drug Agency (ANSM: approval number DMDPT-RIAL/MM/2016-A00608-43) and by the ethical committee (CPP Est III: approval number 16.06.03).

Data Availability

Spectrolive: Optical spectra (autofluorescence and diffuse reflectance) acquired on human skin carcinomas (Original data) (Dataverse).

CRediT Author Statement

Thomas Elsen: Software, Data curation; **Clément Fauvel:** Conceptualization, Methodology, Software; **Grégoire Khairallah:** Methodology, Investigation; **Ahmed Zghal:** Conceptualization, Methodology, Software; **Alain Delconte:** Resources; **Valentin Kupriyanov:** Data curation; **Walter Blondel:** Supervision, Funding acquisition, Data curation, Writing – review & editing; **Marine Amouroux:** Writing – original draft, Project administration, Methodology, Data curation.

Acknowledgements

Authors acknowledge Thomas Jouneau and Laetitia Bracco from ADOC Lorraine department in charge of Open Science at Université de Lorraine (Direction de la Documentation) for advice on dealing with the dataset publication and on the data paper writing. Authors also acknowledge the Metz-Thionville regional hospital staff: Catherine Fery for clinical trial administrative support (Plateforme d'Appui à la Recherche Clinique, PARC), Célia Carbillet for technical assistance in histopathology macroscopy cuts, medical staff from the histopathological department for clinical routine diagnostics, nurses (Patricia Canteri, Séverine Collado, Chantal Gauthier, Marie-Josée Deback, Nathalie Concas) in the surgery room for constant help during spectroscopic measurements.

This work was supported by the Metz-Thionville Regional hospital under the grant "Prix de la Recherche clinique" 2017. The SpectroLive device belongs to the PhotoVivo platform which is part of the France Life Imaging network. The PhotoVivo platform was funded by La Ligue Contre le Cancer and by Contrat de Plan Etat-Région "Innovations Technologiques, Modélisation et Médecine Personnalisée" Grand Est 2015-2020 (CPER IT2MP, plateforme IMTI) including the European Regional Development Fund (FEDER). This work was supported by the French National Research Agency (ANR) under the Investments for the Future program (Programme d'Investissements d'Avenir PIA) bearing the reference ANR-20-SFRI-0009.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N. Rajaram, T.J. Aramil, K. Lee, J.S. Reichenberg, T.H. Nguyen, J.W. Tunnell, Design and validation of a clinical instrument for spectral diagnosis of cutaneous malignancy, Appl. Opt. 49 (2010) 142–152.

[2] M. Amouroux, G. Diaz-Ayil, W.C.P.M. Blondel, G. Bourg-Heckly, A. Leroux, F. Guillemin, Classification of ultraviolet irradiated mouse skin histological stages by bimodal spectroscopy: multiple excitation autofluorescence and diffuse reflectance, J. Biomed. Opt. 14 (2009) 014011, doi:10.1117/1.3077194.

[3] W. Blondel, A. Delconte, G. Khairallah, F. Marchal, A. Gavoille, M. Amouroux, Spatially-resolved multiply-excited autofluorescence and diffuse reflectance spectroscopy: spectrolive medical device for skin in vivo optical biopsy, Electronics 10 (2021) 243, doi:10.3390/electronics10030243.

[4] M. Amouroux, W. Blondel, A. Delconte, Medical device for fibred bimodal optical spectroscopy, US2018348056 (A1), n.d. https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20181206&DB=&locale=fr_EP&CC=US&NR=2018348056A1&KC=A1&ND=4 (accessed July 19, 2020).
[5] T. Elsen, C. Fauvel, A. Zghal, V. Kupriyanov, A. Delconte, G. Khairallah, W. Blondel, M. Amouroux, SpectroLive : optical spectra (autofluorescence and diffuse reflectance) acquired on human skin carcinomas, (2023). doi:10.12763/EYVX3P.
[6] L. Lim, B. Nichols, M.R. Migden, N. Rajaram, J.S. Reichenberg, M.K. Markey, M.I. Ross, J.W. Tunnell, Clinical study of noninvasive in vivo melanoma and nonmelanoma skin cancers using multimodal spectral diagnosis, J. Biomed. Opt. 19 (2014) 117003, doi:10.1117/1.JBO.19.11.117003.
[7] R. Kim, A. Kawai, M. Wakisaka, T. Kin, Current status and prospects of anesthesia and breast cancer: does anesthetic technique affect recurrence and survival rates in breast cancer surgery? Front. Oncol. 12 (2022) accessed September 19, 2023, doi:10.3389/fonc.2022.795864.
[8] C. Guinot, D.J.-M. Malvy, L. Ambroisine, J. Latreille, E. Mauger, M. Tenenhaus, F. Morizot, S. Lopez, I. Le Fur, E. Tschachler, Relative contribution of intrinsic vs extrinsic factors to skin aging as determined by a validated skin age score, Arch. Dermatol. 138 (2002) 1454–1460, doi:10.1001/archderm.138.11.1454.
[9] G. Dobos, A. Lichterfeld, U. Blume-Peytavi, J. Kottner, Evaluation of skin ageing: a systematic review of clinical scales, Br. J. Dermatol. 172 (2015) 1249–1261, doi:10.1111/bjd.13509.
[10] E. Stella, A. Di Petrillo, Standard evaluation of the patient: the Merz scale, in: M. Goisis (Ed.), Inject. Aesthetic Med. Atlas Full-Face Full-Body Treat., Springer Milan, Milano, 2014, pp. 33–50, doi:10.1007/978-88-470-5361-8_3.
[11] A. Zghal, C. Fauvel, V. Kupriyanov, T. Elsen, G. Khairallah, W. Blondel, M. Amouroux, Human skin phototype and apparent age classification based on machine learning methods of autofluorescence and diffuse reflectance spectroscopic data acquired in vivo, in: Adv. Biomed. Clin. Diagn. Surg. Guid. Syst. XXI, SPIE, 2023, pp. 60–64, doi:10.1117/12.2649661.
[12] T.B. Fitzpatrick, The validity and practicality of sun-reactive skin types I through VI, Arch. Dermatol. 124 (1988) 869–871, doi:10.1001/archderm.1988.01670060015008.