



Hyperspectral imaging and robust statistics in non-melanoma skin cancer analysis

LLOYD A. COURTENAY,^{1,*}  DIEGO GONZÁLEZ-AGUILERA,¹
SUSANA LAGÜELA,¹ SUSANA DEL POZO,¹ CAMILO RUIZ-MENDEZ,²
INÉS BARBERO-GARCÍA,¹ CONCEPCIÓN ROMÁN-CURTO,^{3,4} JAVIER
CAÑUETO,^{3,4,5} CARLOS SANTOS-DURÁN,³ MARÍA ESTHER
CARDEÑOSO-ÁLVAREZ,³ MÓNICA RONCERO-RIESCO,³ DAVID
HERNANDEZ-LOPEZ,⁶ DIEGO GUERRERO-SEVILLA,⁶ AND PABLO
RODRÍGUEZ-GONZÁLEZ⁷

¹Department of Cartographic and Terrain Engineering, Higher Polytechnic School of Ávila, University of Salamanca, Hornos Caleros 50, 05003 Ávila, Spain

²Department of Didactics of Mathematics and Experimental Sciences, Faculty of Education, Paseo de Canaleja 169, 37008, Salamanca, Spain

³Department of Dermatology, University Hospital of Spain, Paseo de San Vicente 58-182, 37007, Salamanca, Spain

⁴Instituto de Investigación Biomédica de Salamanca (IBSAL), Paseo de San Vicente, 58-182, 37007 Salamanca, Spain

⁵Instituto de Biología Molecular y Celular del Cáncer (IBMCC)/Centro de Investigación del Cáncer (lab 7). Campus Miguel de Unamuno s/n. 37007 Salamanca, Spain

⁶Institute for Regional Development, University of Castilla la Mancha, Campus Universitario s/n, 02071, Albacete, Spain

⁷Department of Mining Technology, Topography and Structures, University of León, Ponferrada, León, Spain

*ladc1995@gmail.com

Abstract: Non-Melanoma skin cancer is one of the most frequent types of cancer. Early detection is encouraged so as to ensure the best treatment, Hyperspectral imaging is a promising technique for non-invasive inspection of skin lesions, however, the optimal wavelengths for these purposes are yet to be conclusively determined. A visible-near infrared hyperspectral camera with an ad-hoc built platform was used for image acquisition in the present study. Robust statistical techniques were used to conclude an optimal range between 573.45 and 779.88 nm to distinguish between healthy and non-healthy skin. Wavelengths between 429.16 and 520.17 nm were additionally found to be optimal for the differentiation between cancer types.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Skin cancer is the most common type of cancer in countries of predominantly light-skinned populations [1–4]. Skin cancer is typically divided between Melanoma and Non-Melanoma Skin Cancer (NMSC). NMSC is up to 20 times more common than malignant melanoma, representing one-third of cancer cases in the US [5,6], while their incidence is still growing yearly [7, 8, *ibid*]. Similarly, high percentages of NMSC incidences have also been reported in other countries, such as Germany [9], and Korea [10].

NMSCs can additionally be divided into different types, with the most common including Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC). While SCCs are not necessarily restricted to appearing on skin, throughout the present study SCC refers only to Cutaneous SCC, sometimes referred to as cSCC. BCC represents approximately 80% of cases [11], with SCC nearly the remaining 20% [12]. Other NSMCs, such as, Merkel-Cell Carcinomas, Adnexal

Tumours, and other primary cutaneous neoplasms, are present in much lower frequencies [7]. Although SCC and BCC are sometimes pooled together, there are clear differences between their characteristics, treatments and outcomes. BCCs are malignant lesions with low metastasis risk. SCCs can both locally invade, metastasize, and cause death in a subset of patients [12]. There are an estimated 15,000 deaths per year from SCC in the United States, which is twice the number of deaths from melanoma. The most important environmental factor for NMSC is exposure to UV radiation [7].

Early detection of skin cancer is vital in order to achieve successful treatments and the best possible outcome. Early diagnosis has mainly relied upon clinical examination, nevertheless, many visual biomarkers are often limited by subjectivity, with reports calculating a margin of human-induced error of up to 15% among experts [13]. Similarly, five year survival rates for SCC patients have been calculated to fall from 95-98% to 60-62% if not detected early [14,15]. In the case of BCC, early detection is fundamental for the prevention of metastasis [16]. BCC metastasis rates are typically below 0.55%, when caught early. Nevertheless, when treatment is delayed, lesions of > 3 cm in diameter present an increase in metastasis risk of up to a 1-2%, while for tumors of < 10 cm this risk can increase up to 50% [16].

While visual inspection can also be facilitated by the use of a dermatoscope with polarized filters (also known as a dermoscope), results have also been noted to be heavily influenced by inter-observer variability [17,18], with a notable drop of up to 23% sensitivity product of analyst experience [19]. Similarly, reports have shown that in the case of small melanomas, sensitivity can be recorded as low as 39% [20]. In light of this, the highest current accuracy in skin lesion diagnosis is through the biopsy and direct histopathological analysis of the lesion. Needless to say, while accurate, these methods are invasive, costly and time-consuming [11]. Similarly, one of the greatest issues with biopsies are frequently found when the procedure is unable to capture the entirety of the lesion. While this does not necessarily effect their diagnostic accuracy, this does condition their effectiveness [21].

In response to these issues, multiple efforts have been made to reduce diagnostic error, screening time, increase efficiency and reduce the invasiveness of these procedures using advanced computer vision techniques. At the forefront of these advances, multimodal, multispectral and hyperspectral images present promising results [22–26]. In particular, these techniques currently present non-invasive alternatives for the detection of melanoma [22,24,26], as well as the distinction between the different types of skin lesion [23]. From this perspective, hyperspectral images provide a detailed spectral signature of each tissue type, composed of a high number of spectral bands, usually beyond the visible range of the spectrum. These approaches thus allows for the combination of spectral information derived from spectroscopy, and metric information provided by image data.

Among the several types of hyperspectral cameras available, sensors can be classified either by their spectral range, the means in which information is acquired, as well as imaging modality. From the spectral perspective, the most common cameras are those that cover the Visible to Near Infrared (VNIR: 400-1000 nm), Near Infrared (NIR: 1000-1700nm) and the Short-Wave Infrared spectrum (SWIR: 1000-2500 nm). Imaging modality can be divided into pushbroom, whiskbroom and frame cameras, among others. Pushbroom linear cameras are highly popular, registering information through a vectorial array of pixels, also known as line-scanning. Moreover, pushbroom linear cameras are more durable than whiskbroom cameras because they have fewer moving parts, while their geometric resolution is higher than frame cameras. Finally, these linear cameras are cheaper and require a much lower number of sensors [25,27].

Several studies have focused on the identification of NMSCs using sections of the NIR spectrum, identifying higher percentages of water in non-melanoma tumours as opposed to healthy skin [28,29]. Part of this success is product of the greater penetrating capabilities electromagnetic frequencies with wavelengths longer than 700 nm have of biological tissues [30]. The use of

VNIR spectroscopy has also proven useful for the detection of melanoma [31], while frequencies between the wide window of 400 to 1700nm have proven useful in some studies for the detection of SCC [32]. Although infrared light seems to be the most common spectral range for detection of SCC and BCC, some studies have also achieved success detecting SCC tumours within the 450 to 900 nm range [33], or 500 to 900 nm for BCC [34].

Needless to say, however, hyperspectral imaging also has its disadvantages, seen in the high complexity and large size of the images obtained. From this perspective, the processing of images of this type can frequently be considered not only difficult, but also computationally expensive. A frequent task in the preprocessing of any dataset, prior to more advanced applications such as image classification or segmentation, consists in feature selection and extraction [35,36]. This process entails the calculation of areas that are most informative, allowing for the removal of redundant variables that may be hindering model performance. While many studies are able to extract valuable information from hyperspectral images of NMSC skin lesions [22–26], most perform these studies directly on entire datasets. In light of this, a preprocessing procedure such as that of feature selection could be considered a valuable step towards optimizing these approaches. While multiple approaches exist for the purpose of feature selection and extraction, some of the most important tools available in the field of data science are those used in advanced statistics. From this perspective, advanced statistical analyses of hyperspectral data can provide a valuable insight into the precise nature of the data at hand, as well as any underlying patterns, thus facilitating the selection of the most important wavelengths for subsequent classification tasks.

The present study develops this perspective for the analysis of NMSCs, using samples of hyperspectral images obtained from both BCC and SCC patients. The present study employs the use of a pushbroom hyperspectral linear camera registering wavelengths 398.08 to 995.20 nm over 270 bands. Using robust statistical approaches, these analyses show a window between 573.45 and 779.88 nm to be particularly powerful for the detection of differences between healthy skin and both types of NMSC lesions. This type of analysis can be considered a fundamental basis upon which more complex computer vision and artificial intelligence-based studies can be built upon.

2. Materials and methods

2.1. Hyperspectral image acquisition

A Headwall Nano-Hyperspec, Visible-Near InfraRed (VNIR) hyperspectral imaging sensor, was used for the purpose of the present study (Table 1). This particular sensor is a pushbroom linear camera, offering a vectorial array of pixels (1×640 px per image). While pushbroom sensors present the disadvantage of a longer measuring time than that of snapshot hyperspectral cameras, pushbrooms offer higher spatial resolution than many other types of hyperspectral sensors. Nevertheless, these sensors require that the Field-Of-View (FOV) be moved, or *pushed*, along the x-axis in order to obtain an entire image of more than one column of pixels. For this purpose, the present study built an ad-hoc platform.

The platform constructed for this study consists of a motorized structure, composed of an 80 cm long aluminium rail with a motorized base. Control of this base can be managed by external hardware that can control the speed and modality of base displacement (single or loop movements). On top of the base, a multifunctional structure was constructed to fit the hyperspectral sensor, alongside a system of illumination, as well as a frame that controls the distance between the object and the scanning window (Fig. 1(a) & 1b). The system of illumination consisted of two 60 watt halogen light sources mounted on either side of the hyperspectral sensor, with a distance of 14 cm between themselves, and 19 cm between the lens and the object to be photographed (Fig. 1(a)).

In order to control the platform, an external electronic module device was designed (Fig. 1(c)), synchronising the movement of the platform with the illumination system and the sensor's shutter

Table 1. Specifications for the hyperspectral pushbroom camera, Headwall Nano-Hyperspec, used for the present study.

	Headwall Nano-Hyperspec
Image Modality	Pushbroom
Sensor Type	CMOS
Number of Spectral Bands	270
Wavelength Range	398.08–995.20 nm
Space between Spectral Bands	2.2 nm
Number of Spatial Bands	640
Pixel Size	7.4 μm
Depth of Field	17.6 mm
Sensor Size	76 \times 76 \times 87 mm
Sensor Weight	523 g (+90 g lens)
Platform Push-speed	12 mm/s
Full-Width at Half Maximum	6 nm
Entrance Slit Width	20 μm
Bit-depth	12-bit

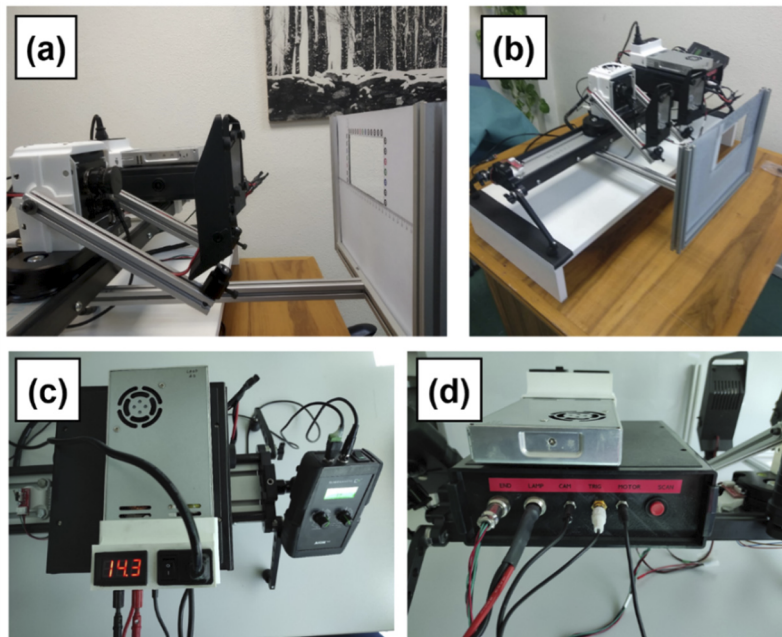


Fig. 1. The hyperspectral pushbroom platform (80 \times 25 cm) and system built for data acquisition purposes in the present study. (a & b) Multifunctional structure, composed of the sensor, halogen light illumination, and calibration marker board and frame. (c) Electronic module controller. (d) Power supply connected to the controller.

speed. This is also highly important so as to ensure a stable displacement speed and thus optimise the generation of a final complete image. With a simple switch, this controller could manage all three elements, as well as simultaneously managing the power source for the entire system (motor, platform control, illumination and the camera; Fig. 1(d)).

For data acquisition, each image was calibrated using the same marker board and frame. For this purpose, the present study used a known reflectance pattern (Spectralon) in order to obtain reflectance values for all 270 bands of the camera (Fig. 2(a)), instead of digital values. Reflectance values were calculated and radiometrically corrected (Eq. (1)). This consists in taking raw uncorrected data (S) from images and calculating % reflectance values (X), taking into consideration the dark current (B) and a “white” reflectance standard (W) [37]. Dark pixel offset values were thus obtained by taking photographs with the lens cap covered [38]. Considering how charge-coupled devices are never capable of measuring an absolute zero value (black), even in cases where no light is available, the present study performed these calibrations so as to calculate the presence of residuals prior to photographing each patient and thus removing dark noise across all bands (Eq. (1), Fig. 2(b)). W values were obtained from a designated region on the Spectralon (Fig. 2(b)).

$$X = 100(S - B)(W - B)^{-1}. \quad (1)$$

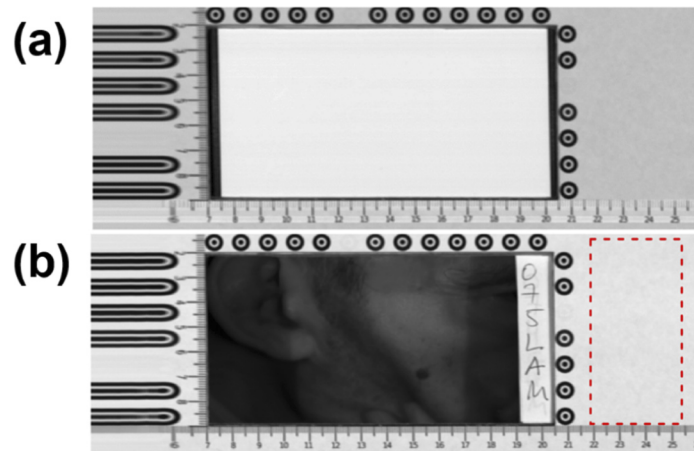


Fig. 2. The marker board used for the calibration and correction of the hyperspectral images. (a) Spectralon reflectance pattern. (b) Region (marked in red) used to obtain reflectance values for all 270 bands of the camera

Upon the Spectralon frame, circular elements were additionally used to calibrate the movement and speed of the platform, whereby the known shape and size of these circles could be used to correct displacement and ensure removal of image distortion (Fig. 2(b)).

In order to perform these calibrations, a software tool was designed to perform radiometric corrections for both eliminating dark current as well as obtaining reflectance images.

Once the entire image had been generated, each image’s size was measured at 640×1785 px, consisting of up to ≈ 308.5 million digital values and occupying ≈ 1.2 Gb of memory. After cropping images to remove the calibration marker board, final images were of size 431×851 px. The first and last five hyperspectral bands were also removed as a safe-measure prior to further processing, as they were occasionally observed to present unusual anomalies. The final spectral window was thus calculated to fall between 409.18 and 984.10 nm. After cropping, final images could therefore be reduced to tensors of size $431 \times 851 \times 260$ (rows, columns and bands) with ≈ 95.4 million numeric values, occupying ≈ 0.3 Gb of memory.

2.2. Sample

A total of 115 patients with observed skin lesions were selected and photographed using the hyperspectral sensor. After images had been acquired, the presence and type of cancer was confirmed after a histopathological analysis and final diagnosis. Of the 115 patients studied, 1 patient was diagnosed with melanoma, 1 patient with merkel-cell carcinoma, 5 with actinic keratoses, 67 with BCC, and 22 with SCC. 19 patients were not diagnosed with malignant skin cancer.

All patients agreed to participate in the study, however due to patient animosity, no further details have been disclosed. All patients were registered and treated at the Institute for Biomedical Research of Salamanca (IBSAL), University Hospital of Salamanca, Spain.

2.3. Hyperspectral signature analysis

Once images had been obtained, careful evaluation of the characteristics of each image was performed. After careful visual inspection, some images were found to be out of focus due to patient movement. Each image was therefore assessed and discarded in cases where quality was found to be insufficient (blurry or incomplete). After careful inspection, 3 SCC patients were unfortunately discarded due to insufficient image quality, while 26 BCC patients were discarded. Final medical samples, therefore, consisted of 60 patients presenting 41 confirmed cases of BCC and 19 cases of SCC.

Once the best images had been obtained, Regions of Interest (ROI) were established to sample pixels of Healthy (H) and pathological skin (BCC & SCC) (Fig. 3) on each of these patients. SCC and BCC ROIs were established directly over the tumour, while H samples were taken from skin farthest away from the tumour so as to avoid possible contamination. After defining ROIs for each of the images, a Python algorithm randomly sampled pixels to extract hyperspectral signatures with as little intervention by a human analyst as possible. Randomness was employed so as to avoid subjective sampling. Sampling was additionally performed until a sufficiently large sample size of pixels had been obtained for H, SCC and BCC. The final selection obtained consisted of 504 hyperspectral signatures for BCC samples, 513 signatures for SCC samples, and 488 signatures for healthy (H) skin samples (total $n = 1,505$).

The statistical power for these sample sizes according to Cohen's δ [39], with an α value of 0.05, were computed at 0.88 ($\delta = 0.2$) and 1 ($\delta < 0.3$), while an α value of 0.003 lowers power to 0.57 ($\delta = 0.2$), 0.96 ($\delta = 0.3$) and 1 ($\delta < 0.4$). From this perspective, the current sample sizes of ≈ 500 observations have a 96% probability of detecting an alternative hypothesis (H_a) when using an α value of 0.003, even if the effect size (i.e. importance of differences) is small ($\delta = 0.3$) [39].

In order to determine the best statistical means of characterising this data, signatures were first subjected to normality testing. The concept of "normality" is a fundamental component in statistics, considering how many statistical tests are conditioned by precise underlying mathematical properties. From this perspective, and in order to select the most reliable statistical tests for comparing samples, the analyst must be aware of the nature of the distributions being studied.

For these purposes, Shapiro-Wilk tests were first passed over each band for each of the samples [40]. Shapiro-Wilk testing was additionally complemented with a visual inspection of each distribution using density plots and quantile-quantile plot calculations [41]. From a similar perspective, mean residual counts from a linear Gaussian model were calculated for each band to visualize areas of greatest deviations from the mean. Residual calculations were performed so as to assess areas where a simple linear model is less likely to capture the general trend of the distribution in question. Calculations for sample skewness and kurtosis were also performed, so as to better define the nature of each distribution. With regards to skewness, in cases where samples are normally distributed, sample skewness would be expected to be close or equal to 0, indicating a symmetric distribution. As for kurtosis calculations, distributions that are

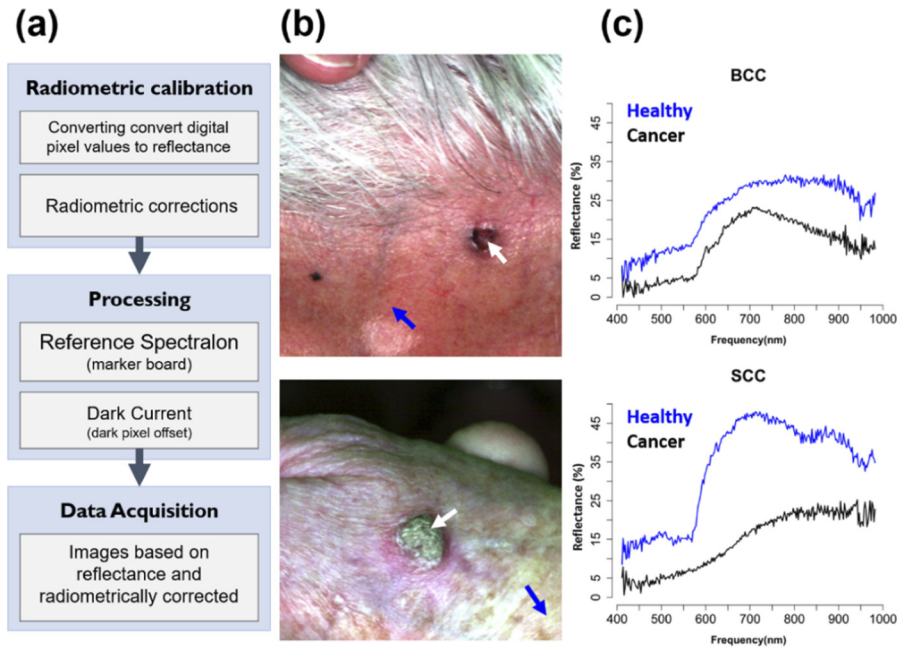


Fig. 3. Examples of sampled hyperspectral curves from different patients. (a) A graphical description of the image acquisition workflow. (b) An example of (upper) Basal Cell Carcinoma, found under the hairline of the frontal portion of a male patient's head, and (lower) a Squamous Cell Carcinoma found on the back of a female patient's hand. (c) Examples of the hyperspectral signatures obtained for (upper) the Basal Cell Carcinoma patient (a-upper) and the Squamous Cell Carcinoma patient (a-lower). Faces and distinguishing features have been excluded from these figures to ensure patient confidentiality.

normally distributed typically present kurtosis values close to 0, indicating neither an excessive concentration of information (kurtosis > 0), nor a wide spread of values (kurtosis < 0).

Upon accepting or rejecting the null hypothesis, H_0 , of normally distributed data, different statistical approaches were used to define the hyperspectral "signature" of each sample.

For descriptive statistics, central tendency was calculated using either the mean or median for Gaussian and non-Gaussian distributed data respectively [41–46]. Likewise, calculations for sample variance were either calculated using the standard deviation or the Square Root of the Biweight Midvariance ($\sqrt{\text{BWMV}}$) (Eq. (2)-5) [45–47]. $\sqrt{\text{BWMV}}$ values are calculated in accordance with the Median Absolute Deviation (MAD, Eq. (2)), which takes the absolute difference of each value (x) to the sample median (\tilde{x}). When non-symmetric measures of variance were required, robust quantile calculations were performed using 95% confidence intervals [41].

$$\text{MAD} = \tilde{x}(|x_i - \tilde{x}|), \quad (2)$$

$$\text{BWMV} = \frac{n \sum_{i=1}^n a_i (x_i - \tilde{x})^2 (1 - U_i^2)^4}{\left(\sum_{i=1}^n a_i (1 - U_i^2) (1 - 5U_i^2) \right)^2}, \quad (3)$$

$$a_i = \begin{cases} 1, & \text{if } |U_i| < 1 \\ 0, & \text{if } |U_i| \geq 1 \end{cases}, \quad (4)$$

$$U = \frac{x_i - \tilde{x}}{9MAD}. \quad (5)$$

For hypothesis testing, tests were performed to analyse homoscedasticity and thus determine areas of important differences in variance. For parametric testing of homoscedasticity, the Bartlett's test was used [48], while non-parametric tests employed Levene's test [49]. Both the Bartlett and Levene test assume H_0 to infer samples have equal variance. For multivariate analyses, a Multivariate Analysis of Variance (MANOVA) was performed, using either the Hotelling-Lawley test statistic for parametric approaches [50], while in cases where distributions proved to be non-homogeneous, a pairwise Wilcoxon test was performed [51]. In each of these tests, H_0 assumes samples to be similar.

In addition to hypothesis testing, and as a means of comparing differences and similarities between probability distributions, the Jensen-Shannon Distance (JSD) was computed using the Kullback Leibler divergence [52,53]. This method was used to measure the similarity between different distributions across the spectrum, thus finding areas of greatest separations between samples in accordance with mutual information theory. Samples that are considered similar would thus produce distance calculations closer to 1, while values closer to 0 indicate greater differences between sample distributions.

All statistical applications were performed in the R programming language (v.4.0.4) [54]. The Python programming language (v.3.7.4) was also employed for hyperspectral image processing and signature extraction.

2.4. Evaluation of hypothesis test results

Recent years have seen a rise of criticism on the "blind" use of p -values for withdrawing scientific conclusions, especially with regards to the use of $p < 0.05$ for hypothesis testing [55,56]. In light of this, the present study has made a particular effort to avoid the misuse of p -values, so as to ensure the highest possible validity of the presented conclusions.

In accordance with the most recent recommendations set forth by the *American Statistician* [55,56], p -values were not evaluated using the more traditional $p < 0.05$ as a threshold for defining statistical significance. Likewise, the term "significant" has been avoided throughout the present study. Instead, frequentist p -values were evaluated in accordance with calibrated Bayesian statistical approaches, converting each p -value into Bayesian Factor Bound values, following the suggestions by Benjamin and Berger [57]. In each of these cases, the upper bound on the posterior probability of the alternative hypothesis, H_a , were calculated ($P^U(H_a|p)$; Eq. (6) & (7));

$$BF \leq BFB \equiv \frac{1}{-e p \log(p)}, \quad (6)$$

$$P^U(H_a|p) = \frac{BFB(p)}{1 + BFB(p)}. \quad (7)$$

Similarly, where necessary, the Bayes Factor Bound (BFB) derived from Eq. (6) was used to determine posterior odds of H_a to H_0 for each of the tests. For the majority of tests, each calibration was performed using prior probabilities indicative of complete randomness (prior = 0.5), as suggested by Colquhoun [58]. Nevertheless, for the construction of confidence intervals around these calculations, prior probabilities of 0.8 and 0.2 were also considered and reported.

To assess and account for possible Type I statistical errors among hypothesis tests [58], each of these p and $P^U(H_a|p)$ values were also accompanied by calculations of the False Positive Risk (FPR; Eq. (8) & (9));

$$L_{10} = \frac{P(x|H_a)}{P(x|H_0)}, \quad (8)$$

$$FPR = \frac{1}{1 + L_{10} \frac{P(H_a)}{1 - P(H_a)}}. \quad (9)$$

While a number of different formulae have been proposed as a means of defining the likelihood ratio of H_0 against H_a , in other words L_{10} in Eq. (8) [58,59], the present study employs the Sellke-Berger approach [57,60], which is the equivalent of using Eq. (6) for the calculation of FPR (Eq. (9)). Additionally, considering observations by Courtenay et al. [61], where necessary, a complementary calculation of FPR for deriving the probability of H_0 ($P(H_0)$) was also performed (Eq. (10) & (11));

$$IFPR = \frac{1}{1 + L_{10} \left(1 - \left(\frac{P(H_a)}{1 - P(H_a)} \right) \right)}, \tag{10}$$

$$P(H_0) = \begin{cases} FPR(p), & \text{if } p \leq 0.3681 \\ 1 - IFPR(p), & \text{if } p > 0.3681 \end{cases}. \tag{11}$$

$P(H_0)$ is used as a means of calibrating p-values above $p = 0.3681$, considering how observations by Courtenay et al. [61] found this value as a point of maximal curvature in p -value calibration curves using equations 7 & 9. From this perspective, the inverse of FPR (Eq. (10)) can be used to ensure each p -value between 0 and 1 have their own unique calibration values [61].

For the interpretation of these calibrated metrics, BFB values indicate the data-based odds of H_a being true to H_0 , whereby high values of BFB support H_a . So as to facilitate the interpretation of these odds, the $P^U(H_a|p)$ calculation ensures this number is reported as a percentile falling between 0.5 and 1. The posterior odds of H_a to H_0 are interpreted the same as BFB, where high values support H_a . FPR, on the other hand, returns a decimal value between 0 and 0.5, with 0.5 indicating a high chance of the concluding hypothesis to be a Type I statistical error. The $P(H_0)$ thus ensures that FPR values are returned between 0 and 1, with 1.00 indicating a 100% probability of incorrectly concluding H_a to be true.

In light of these calibrations, p -values were evaluated in accordance not only with their corresponding $P^U(H_a|p)$, FPR and $P(H_0)$, but also using a more robust and conservative p values $< 3\sigma$ from the mean (0.003) as a threshold for more conclusive results. For ease of comparison and calibration, Table 2 presents different p -values calibrated using metrics BFB, $P^U(H_a|p)$, FPR and H_a to H_0 ratios. Table 3 presents different p -values calibrated using $P(H_0)$. As can be seen, the traditional $p < 0.05$ (2σ) presents a very low BFB value of 2.5 to 1, with a 28.9% probability of being a Type I statistical error using prior probabilities of 0.5 in support of the alternative hypothesis. This indicates that $p < 0.05$ is not a reliable threshold to define conclusive evidence. $p < 0.003$ (3σ), on the other hand, results in BFB values of 21 to 1, with only a 4.5% chance of being a Type I statistical error using the same prior probabilities.

Table 2. Bayes Factor Bounds (BFB, eq. (6)), Posterior Probability of H_a values ($P^U(H_a|p)$, eq. (7)), Posterior Odds of H_a to H_0 values ($H_a:H_0$), and False Positive Risk (FPR, eq. (9)) values, for a number of their corresponding p -values with different prior odds.

prior	p	0.368	0.100	0.050	0.010	0.005	0.003	0.001	0.0001
	BFB	1.000	1.598	2.456	7.988	13.89	21.11	53.26	399.4
	$P^U(H_a p)$	0.500	0.615	0.710	0.889	0.933	0.955	0.982	0.998
0.2	$H_a:H_0$	0.200	0.320	0.491	1.600	2.778	4.222	10.65	79.88
0.2	FPR	0.800	0.715	0.715	0.334	0.224	0.159	0.070	0.010
0.5	$H_a:H_0$	0.500	0.799	1.220	3.994	6.943	10.55	26.63	199.7
0.5	FPR	0.500	0.385	0.289	0.111	0.067	0.045	0.018	0.002
0.8	$H_a:H_0$	0.800	1.278	1.964	6.391	11.11	16.89	42.60	319.5
0.8	FPR	0.200	0.135	0.092	0.030	0.177	0.012	0.005	0.001

Table 3. $P(H_0)$ values (eq. (11)) for a number of their corresponding p-values using different priors.

p	0.99	0.8	0.5	0.368	0.1	0.050	0.010	0.005	0.003
0.2 Prior	0.993	0.892	0.809	0.800	0.715	0.620	0.334	0.224	0.159
0.5 Prior	0.974	0.673	0.515	0.500	0.385	0.289	0.111	0.067	0.045
0.8 Prior	0.902	0.340	0.210	0.200	0.135	0.092	0.030	0.018	0.012

3. Results

3.1. Descriptive statistics

Hyperspectral curves for all three samples (H, BCC & SCC) presented highly inhomogeneous distributions across most of the spectrum (Fig. 4 & 5). This was especially evident for frequencies below 699.97 nm (Fig. 5). While frequencies above this threshold presented increasingly more Gaussian-like distributions in the case of BCC and H samples (central $w = 0.99$, $p = 0.03$, $P^U(H_a|p) = 0.77$), SCC samples were found to be highly inhomogeneous throughout the entire spectrum (central $w = 0.98$, $p = 1.8e-06$, $P^U(H_a|p) = 0.99$). Furthermore, in the case of SCC, the probability of this observation being a Type I statistical error was calculated at $0.006 \in [0.0016, 0.026]\%$ (Fig. 5).

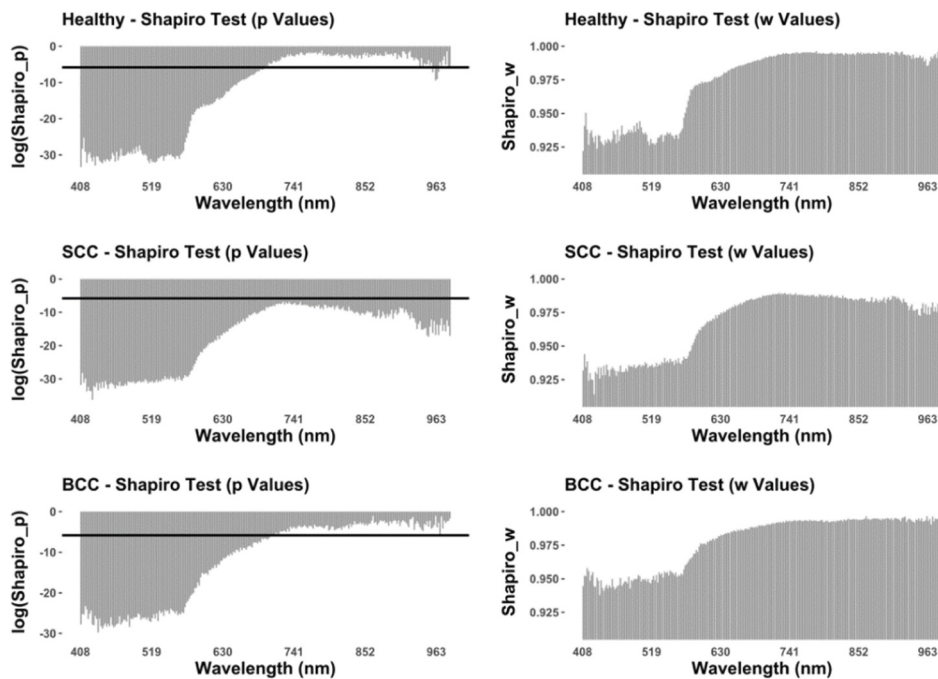


Fig. 4. Graphs presenting the logarithm of Shapiro-Wilks p-values as well as test statistics (w) for each of the samples across each of the bands. The solid horizontal line in each of the left-hand panels mark the $p = \log(0.003)$ threshold, that is, all $\log(p)$ values that fall below this line have less than a 5% chance of being false positives, and can thus be considered strong deviations from the normal distribution.

Calculations regarding residuals when fitting linear models onto the data additionally reveal a notable decrease in residuals towards the NIR regions (>702.19 nm) of the spectrum (Fig. 6). Combined with observations regarding sample skewness, it can be seen how regions of lower frequencies present great positive skewness (skewness > 0 , Fig. 7), contributing to the lack of

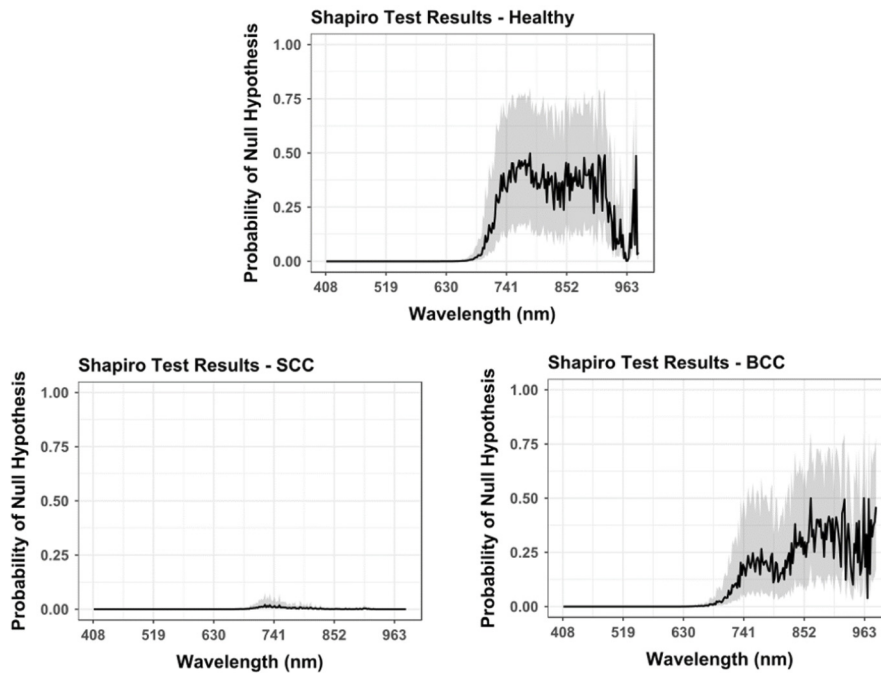


Fig. 5. Graphs presenting $p(H_0)$ calibrations for each of the Shapiro-Wilks p -values in Fig. 4. Central values were calculated using 1:2 prior probabilities while confidence intervals mark upper bound 3:10 prior probabilities in favour of H_0 and lower bound 7:10 prior probabilities in favour of H_0 .

sample normality, while skewness values drop for most samples. In the case of kurtosis, SCC samples are seen to have a very wide spread (kurtosis < 0), while H and BCC have very high concentrations of information in the shorter wavelength frequencies of the visible light spectrum (kurtosis > 0 , Fig. 7). Nevertheless, as noted by the lack of normality in SCC samples, positive skew remains high throughout.

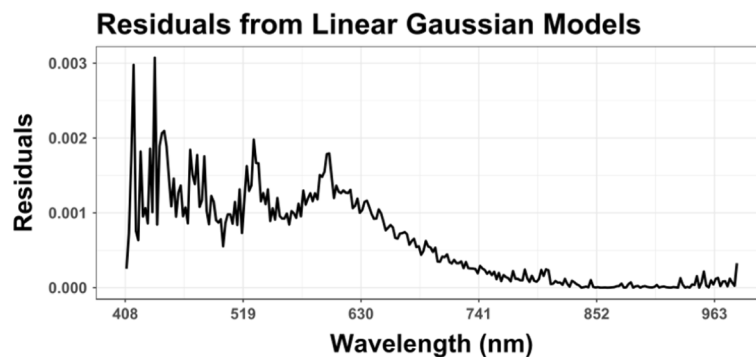


Fig. 6. Calculated residuals for fitted linear models across the entire spectrum analysed.

In light of each of these observations, it can be seen how SCC is strongly characterised by a highly inhomogeneous and skewed distribution throughout the spectrum, while BCC and H hold a more Gaussian-like nature towards the end of the visual light spectrum. Upon plotting central tendency and variance curves, it can be seen how BCC samples reflect the least amount

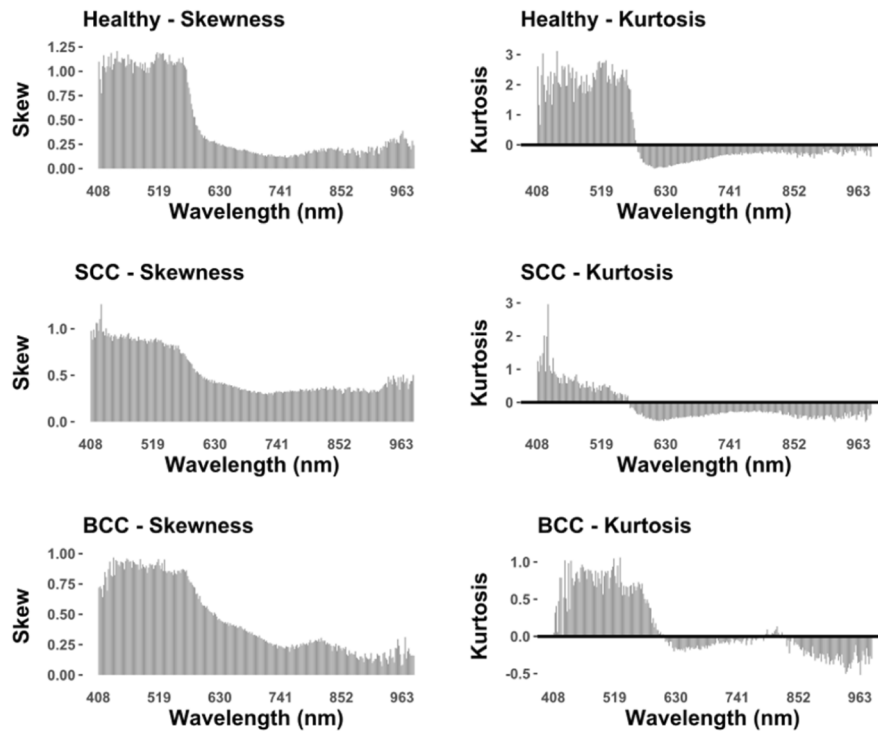


Fig. 7. Sample skewness and kurtosis calculations across the entire spectrum analysed.

of light across all frequencies, while SCC reflects the most light, especially between the ranges of 606.74 and 862.02 nm. Healthy skin samples, on the other hand, appear to have signatures midway between the two samples, appearing more similar to SCC samples. Nevertheless, great overlapping is observed across most samples (Fig. 8), especially between H and SCC samples below 600 nm. This is especially evident when observing central tendency calculations. The greatest differences are observed when considering the variability of sample distributions ($\sqrt{\text{BWMV}}$), highlighting the importance of robust statistical approaches in this type of analysis.

In either case, the notable peaks in reflectance between ca. 600 nm and 850 nm indicate all three samples to be strongly characterised by greater reflectance of orange, red, and the lower frequencies of NIR light (Fig. 8), than any other part of the visible spectrum.

3.2. Univariate hypothesis testing

From a more analytical perspective, Levene tests reveal most samples to reject the terms of homoscedasticity, especially when comparing H and BCC, as well as between SCC and BCC (Fig. 9). Nevertheless, comparisons reveal high similarities in the variances between SCC and BCC (minimum $F = 4.5e-09$, $p = 0.17$, $P^U(H_a|p) = 0.55$), with posterior odds of at most 1 to 0.61 $\in [0.24, 0.85]$ in favour of the alternative hypothesis.

In the case of comparing H with BCC, Levene's test reveals important divergences in the range 571.23 to 651.14 nm (Fig. 9; central $F = 11.4$, $p = 0.0007$, $P^U(H_a|p) = 0.987$), with posterior odds of at most 1 to 36.7 $\in [14.7, 58.7]$ in favour of the alternative hypothesis, and a 1.3 $\in [0.4, 5.2]\%$ chance of this observation being a false positive. When comparing both types of cancer, deviation also occurs beyond 571.23 nm (Fig. 9), however, in this case divergences are prolonged throughout the spectrum until 691.09 nm, with the exception of 7 bands within this 54-band range. Furthermore, while divergences between SCC and BCC are slightly less marked than

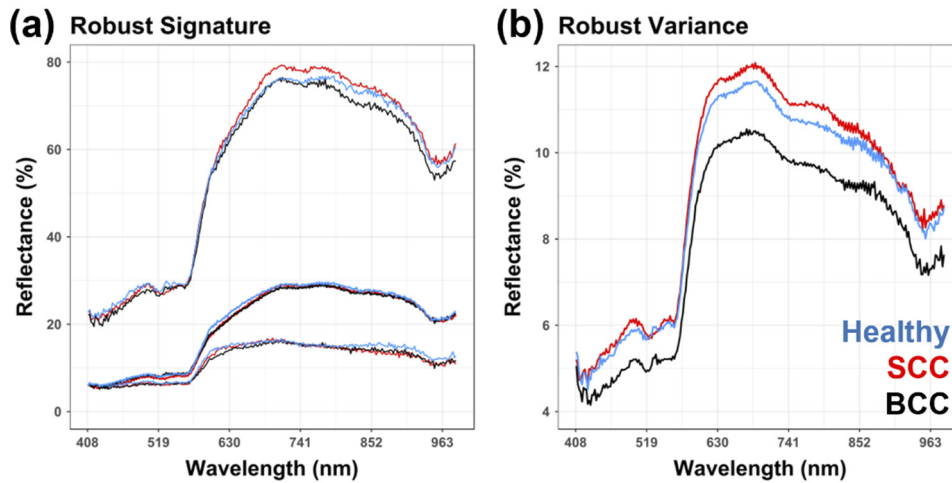


Fig. 8. Hyperspectral signatures for each of the samples. (a) Robust signature marking the central tendency as well as 5% and 95% quantile confidence intervals (lower lines and upper lines respectively). (b) $\sqrt{\text{BWMV}}$ calculations representing robust sample variance.

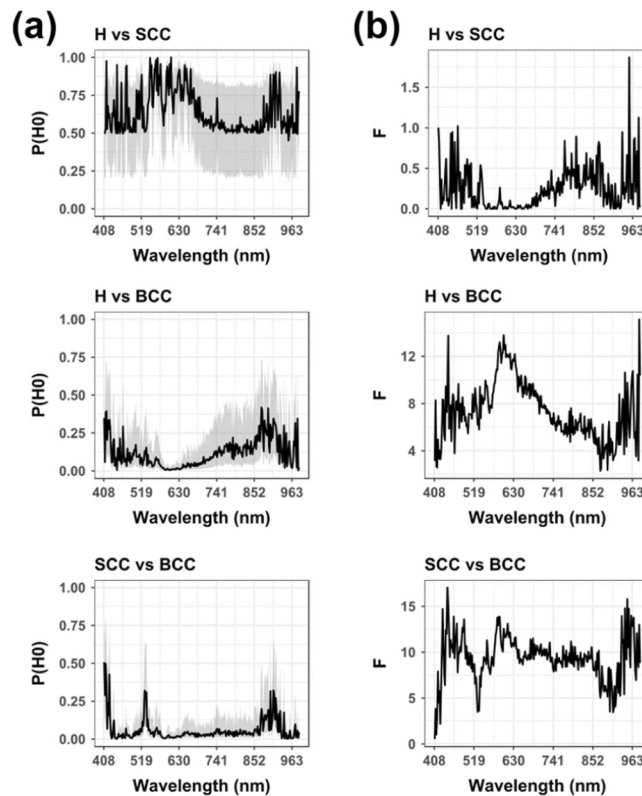


Fig. 9. Univariate hypotheses test results comparing each of the samples across each of the hyperspectral bands using the Levene test for homoscedasticity. (a) Probability of Null Hypotheses ($P(H_0)$) values, calibrated for each p-value using priors of 1:2 to mark the central tendency, while confidence intervals mark upper bounds using 3:10 prior probabilities in favour of H_0 and lower bounds using 7:10 prior probabilities in favour of H_0 . (b) Test statistic calculations for each of the corresponding hypothesis tests.

those observed between H and BCC, they are still of substantial importance (central $F = 11.0$, $p = 0.0009$, $P^U(H_a|p) = 0.982$), with posterior odds of at most 1 to $29.1 \in [11.7, 46.6]$ in favour of the alternative hypothesis, and a $1.8 \in [0.5, 6.9]\%$ chance of this observation being a false positive.

In addition to these windows, BCC and SCC were also found to differ in an additional window towards the violet, blue and cyan regions of the visible spectrum. From this perspective, Levene's test found notable divergences between 440.25 and 502.41 nm (Fig. 9; central $F = 11.2$, $p = 0.0009$, $P^U(H_a|p) = 0.984$). In this case, only 2 bands proved an exception to this rule in the 28 band window. This window is additionally associated with posterior odds of at most 1 to $30.4 \in [12.1, 48.6]$ in favour of H_a . FPR values were calculated at $1.6 \in [0.4, 6.2]\%$.

Table 4 presents a summary of these results and the windows where the greatest differences have been found.

Table 4. Description of the hyperspectral frequency ranges where univariate hypotheses testing found notable differences between samples. FPR values have been calculated using the worst-case scenario with prior odds of 2:10 against the alternative hypothesis.

Comparison	Frequency Range (nm)	N° Bands Included	Test Used	Test Statistic	$P^U(H_a p)$	FPR
H vs BCC	571.23 - 651.14	36	Levene	11.4	0.987	5.2%
SCC vs BCC	571.23 - 691.09	54	Levene	11.0	0.982	6.9%
SCC vs BCC	440.25 - 502.41	28	Levene	11.2	0.984	6.2%

3.3. Jensen-Shannon distances

Similarity measures in accordance with JSD ranged from highly similar probability distributions ($JSD = 0.62$) to areas of notable divergences ($JSD = 0.001$). Nevertheless, the only window presenting convergence of differences across all samples was found to be located approximately between 582.32 to 748.81 nm (JSD median = 0.011, range = [0.004, 0.041]). When considering differences between cancer samples and healthy skin, the greatest divergence of BCC from H was found at 735.49 nm ($JSD = 0.006$), while SCC from H was found at this frequency as well ($JSD = 0.004$). In the case of separating between different types of cancer, the greatest differences for BCC and SCC were found at 673.33 nm ($JSD = 0.006$).

Interestingly SCC and H are clearly differentiable from most points beyond 580.10 nm, with only 3 peaks in similarity in the NIR spectrum at ca. 890.87, 943.04 and 948.59 nm (Fig. 10).

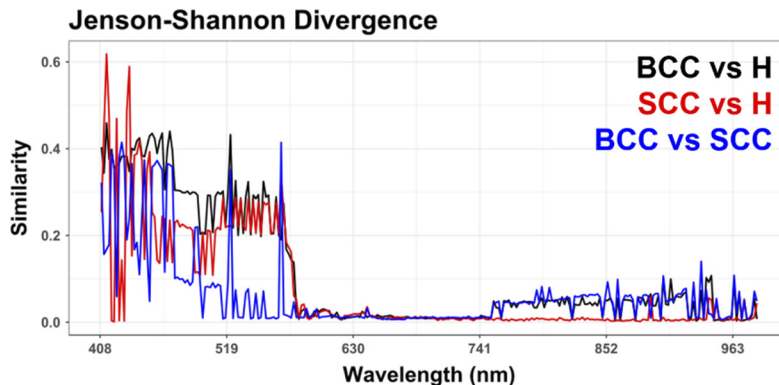


Fig. 10. Comparisons of samples using calculations of distribution similarities via Jensen-Shannon distance metrics.

3.4. Multivariate testing

When comparing all the different results obtained throughout this study, a final window where most major differences seem to be located can be established between 573.45 and 779.88 nm, occupying a substantial proportion of the visible light spectrum, including the spectral colours yellow, orange and red, as well as the beginning of the NIR light spectrum (Fig. 11).

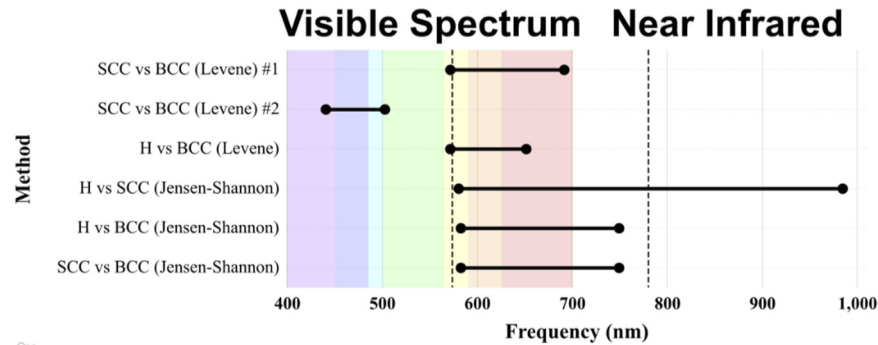


Fig. 11. Optimally defined windows as established by multiple methods within the present study. Area delimited by dotted vertical lines marks the final window of interest between 573.45 and 779.88 nm.

Performing multivariate statistical analyses using the Wilcoxon test across this region reveals important differences between H and SCC ($p < 2.0e-16$, $P^U(H_a|p) \approx 1$), as well as between H and BCC ($p = 1.5e-12$, $P^U(H_a|p) \approx 1$). In the case of H vs SCC, this corresponds to a posterior probability of at most 1 to $2.5e+13$ with prior odds of 1:2, and thus a worst case scenario of a posterior probability of at most 1 to $1.0e+13$ when considering prior odds of 2:10. In the case of H vs BCC, a worst case scenario's posterior probability can thus be calculated at 1 to $1.8e+09$. Needless to say, on both accounts, the probability that this observation is a Type I statistical error is $1.1e-08\%$.

For the case of BCC and SCC, multivariate testing unfortunately reveals these samples to be indistinguishable within this region ($p = 0.17$, $P^U(H_a|p) = 0.55$). Nevertheless, when taking into

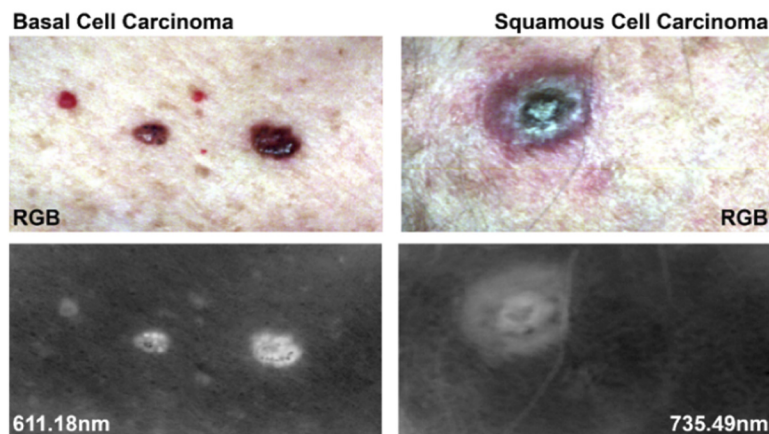


Fig. 12. Visualisation of different skin lesions via single bands (611.18 nm and 735.49 nm) of hyperspectral images, as well as their corresponding RGB images. Channel bandwidths have been measured at 2.2 nm.

consideration the region of 429.16 to 520.17 nm, as defined by the Levene test, Wilcoxon results reveal great differences between BCC and SCC ($p = 6.4e-10$, $P^U(H_a|p) \approx 1$), with posterior probability H_a to H_0 ratios of $1.3e+7 \in [5.4e+6, 2.2e+7]$ using prior odds of 1:2 (FPR = $3.7e-06 \in [1.5e-05, 9.2e-07]\%$).

Needless to say, when processing the entire spectrum, all 270 bands present important multivariate differences between H and both cancer samples. Nevertheless, p -values are slightly higher ($p < 4.5e-12$), while BCC and SCC are still indistinguishable ($p = 0.3$, $P^U(H_a|p) = 0.5$).

Finally, when these bands are used to visualise skin lesions, it can clearly be seen how certain frequencies have greater potential of isolating cancerous skin cells over others, helping in determining areas of subclinical invasion, which may be a valuable tool for future classification tasks as well as for applications in surgical removal of these types of lesions (Fig. 12).

4. Discussion

While cancer is often perceived to be a disease of modernity, studies have shown neoplasms to have great antiquity [62–64], having an effect on most living things at least 255 million years ago (Mya) in mammals [65], and 1.7 Mya in humans [66]. Nevertheless, the severity and increase in malignancy in these pathological phenomena is of increasing concern to modern-day society, a fact that is strongly conditioned by our way of life. NMSC is the most frequent type of cancer in humans [67], representing a major health problem in fair skinned elderly people, while being associated with elevated health costs [68]. Moreover, beyond this, cancer is known to have a significant emotional impact on not only the patient but their families too [69].

Timely detection of skin cancer is fundamental for suitable treatment and improving patient survival rates. The present study has revealed important statistical differences between the multiple samples analysed within the VNIR spectrum using a linear hyperspectral camera. Robust statistical techniques, as well as feature selection algorithms, have been able to highlight these differences particularly within the range of 573.45 and 779.88 nm; occupying a considerable portion of yellow, orange and red visible light, as well as a more reduced proportion of the NIR spectrum. Moreover, the probability that observations separating healthy skin from cancer are a false positive has been calculated at less than 1% based on the present data. While differences between both cancer types were limited within these frequencies, a secondary window was found to be important for these samples between 429.16 to 520.17 nm; occupying portions of violet, blue, cyan and green visible light.

Multiple studies have shown how substances such as melanosomes, collagen, blood and water, affect the spectral signatures of skin and other biological tissues [70]. Similarly, the content of these substances varies among different parts of the body, causing great natural variability within healthy skin hyperspectral signatures. Most studies concur that one of the greatest conditioning factors in the morphology of skin spectral signatures is product of melanin's red light absorption rate, explaining the large peaks and variances in the range of 600 to 800 nm [22,24,26,33,70]. While this is evidently a greater biomarker for the study of melanoma, the present study also highlights portions of this range for the detection of NMSCs as well. Likewise, substances such as haemoglobin have been noted to condition reflectance within the range 530 to 600 nm, while pagetoid growth is known to affect the lower end of the visible spectrum, between 400 and 500 nm.

Halicek et al. [33] describe these phenomena in a study of SCC patients, noting haemoglobin to reflect less light in SCC than healthy skin. While the signatures between 530 and 600 nm are relatively similar in the present study, SCC is indeed seen to reflect 0.65% less light than H samples, while BCC presents a difference of 0.16% less light reflected. While the present study has a smaller number of SCC patients, each of these differences have been reported here as minute, revealing no statistical data that supports using this region of the spectrum as a diagnostic

biomarker. Moreover, a similar lack of statistical differences was noted by Gareau et al. [71] and Hosking et al. [26] in the case of melanoma.

Halicek et al. [33] also describe SCC samples to reflect more light from 600 to 900 nm than healthy skin, concurring with observations by Pardo et al. [24]'s study of melanoma patients. These observations were additionally attributed to melanin's greater absorption rates of light. Via robust statistical techniques, Pardo et al. [24] additionally prove this a valuable biomarker. The present study confirms both these observations, concluding SCC to have a slightly higher absorption rate than healthy skin (0.53% less reflected light), while BCC is strongly characterised by a greater absorption of this type of light (0.85%), much similar to melanoma.

Observing the differences found within the present study for SCC and BCC between 429.16 and 520.17 nm, in a study on melanoma, Hosking et al. [26] note blue light to be particularly susceptible to differences in the dermoepidermal junction. From this perspective, it can be argued that this region of the spectrum is useful for containing information regarding the specific variability of skin cancer types. Considering how SCC and BCCs affect different types of cells, particular atypia and possible variances in pagetoid spread may be contained within this region of the spectra, proving a possible starting point for future investigation in hyperspectral diagnostics.

Nevertheless, the present study still has some limitations. From one perspective, the characteristics of the platform designed, while proving useful, present room for improvement. Considering most patients were of elderly age, the patient's ability to stay still during the 15 s exposition time was greatly reduced. In cases where skin lesions were found on patient's hands, possible tremors meant many images presented a "wavy" pattern or appearance. Similarly, when lesions were found on the face or top of the head, many patients commented on a lack of comfort trying to hold their heads in position. As a consequence, a number of images were unfortunately discarded. In order to overcome this, future efforts should consider the use of a more ergonomic platform, designed with a resting pad where the patient can place their arm or head in a more comfortable position during image acquisition processes.

From a similar perspective, considering the natural variability most lesions presented, due to within-group typological variances among NMSCs, a larger sample of different types of SCC and BCC tumours should be obtained. Moreover, a large number of the patients sampled here were at advanced stages of both tumour growth and spread, especially in the case of SCC patients. From this standpoint, research into earlier stages of carcinoma development and metastasis would be a valuable step towards ensuring, not only efficient diagnoses, but also early detection. It is also strongly recommendable that research goes into premalignant lesions -such as actinic keratosis-, as well as non-malignant nevi, so as to provide a better diagnostic tool.

Similarly, although the present study concurs with data provided by other authors, here artificial skin pigmentation prior to hyperspectral image acquisition was avoided, making some comparisons with other studies that did use pigmentation difficult. Likewise, the natural variability of human skin pigmentation is likely to be an important factor when performing future studies. From this perspective, it can be predicted that healthy skin signatures for fairer-skinned patients will present greater differences than those observed here, with trends towards more reflected light in the 530 to 600 nm range, while less light would theoretically be reflected between 600 and 800 nm.

Needless to say, the application of robust statistical measurements was still able to reveal important features of SCC and BCC signatures, presenting promising possibilities for applications with larger patient sample sizes. Similarly, the differentiation between H, SCC and BCC samples, in specific regions of the spectrum, allows for efficient feature selection. The most recent advances in the integration of hyperspectral imagery to medical applications have shown these tools to greatly increase the accuracy when delimiting skin lesions as opposed to human analysts [24,34]. While the present study has not taken this step, a detailed statistical analysis and characterisation of these types of skin lesions is fundamental for more advanced applications. In sum, and based

on the present data, it can clearly be seen how robust statistical analyses of this nature provide a basis upon which more complex computational learning applications can be built from, especially using artificial intelligence.

5. Conclusion

In this study, robust statistical tests were employed on hyperspectral data in the VNIR spectrum in order to identify the hyperspectral differences between carcinomas (SCC & BCC) and healthy skin (H). The optimal spectral ranges for discrimination between SCC, BCC, and H, as well as between BCC and SCC, have been defined using robust statistics. The results are especially promising for the discrimination between cancerous and non-cancerous areas, paving the way for future research in NMSC diagnostics using hyperspectral images.

A study with larger patient samples should be carried out in the future, allowing for a wider representation of lesion variability (including different stages of growth and stage). The same can be said considering the variability of healthy skin (including features such as non-cancerous-lesions). Similarly, while the presented ad-hoc platform has been designed for its implementation in clinical practice, especially as a tool to support medical practitioners, further improvements in the image acquisition platform should also be made in order to make it more ergonomic and facilitate data acquisition. This can be considered of great importance for more practical scenarios, especially for elderly patients.

Funding. Gerencia Regional de Salud de Castilla y León (GRS 2139/A/20); Spanish Ministry of Science, Innovation and Universities (PRE2019-089411); Instituto de Salud Carlos III (PI18/00587); Iberdrola Spain; Junta de Castilla y León (GRS 1837/A/18).

Acknowledgements. This project was funded by the Junta de Castilla y León, under the title project HYPER-SKINCARE (Ref. GRS 1837/A/18). Lloyd Austin Courtenay is funded by the Spanish Ministry of Science, Innovation and Universities with an FPI Predoctoral Grant (Ref. PRE2019-089411) associated to project RTI2018-099850-B-I00 and the University of Salamanca. Susana Lagüela and Susana del Pozo are both funded by the Iberdrola Spain through the initiative Cátedra Iberdrola VIII Centenario of the University of Salamanca. Javier Cañueto is partially supported by the PI18/00587 (Instituto de Salud Carlos III cofinanciado con fondos FEDER) and GRS 2139/A/20 (Gerencia Regional de Salud de Castilla y León).

We would like to thank all the patients who agreed to participate in the present study. We are also very grateful to the two anonymous reviewers who took the time to revise our manuscript and improve its presentation. We would like to greatly acknowledge the efforts made by health care professionals throughout the world in the fight for cancer. The authors are grateful for the IBSAL for their efforts in the field, while the corresponding author is especially grateful for the Marie Curie organization's contributions to this cause. The corresponding author would like to dedicate this work to Ginger Courtenay.

Disclosures. The authors declare that there are no conflicts of interest related to this article

Data availability. All data used in the current study are located at the corresponding author's GitHub repository [54]. All numeric values used to create figures have also been included within this repository. All code used is also located within this repository.

References

1. H. W. Rogers, M. A. Weinstock, A. R. Harris, M. R. Hinckley, S. R. Feldman, A. B. Fleischer, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer in the United States, 2006," *Arch. Dermatol.* **146**(3), 283–287 (2010).
2. A. Lomas, J. Leonardi-Bee, and F. Bath-Hextall, "A systematic review of worldwide incidence of nonmelanoma skin cancer," *Br. J. Dermatol.* **166**(5), 1069–1080 (2012).
3. N. Eisemann, A. Waldmann, A. C. Geller, M. A. Weinstock, B. Wolkmer, R. Greinert, E. W. Breitbart, and A. Katalinic, "Non-melanoma skin cancer incidence and impact of skin cancer screening on incidence," *J. Invest. Dermatol.* **134**(1), 43–50 (2014).
4. A. Brunssen, A. Waldmann, N. Eisemann, and A. Katalinic, "Impact of skin cancer screening and secondary prevention campaigns on skin cancer incidence and mortality: a systematic review," *J. Am. Acad. Dermatol.* **76**(1), 129–139.e10 (2017).
5. T. L. Diepgen and V. Mahler, "The epidemiology of skin cancer," *Br. J. Dermatol. Supplement* **146**(61), 1–6 (2002).
6. American Cancer Society, "Cancer Facts & Figures," American Cancer Society, Atlanta. (2021). <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf>

7. V. Madan, J. T. Lear, and R. M. Szeimies, "Non-melanoma skin cancer," *Lancet* **375**(9715), 673–685 (2010).
8. T. K. Nikolouzakakis, L. Falzone, K. Lasithiotakis, S. Krüger-Krasagakis, A. Kalogeraki, M. Sifaki, D. A. Spandidos, E. Chrysos, A. Tsatsakis, and J. Tsiaoussis, "Current and future trends in molecular biomarkers for diagnostic, prognostic and predictive purposes in non-melanoma skin cancer," *J. Clin. Med.* **9**(9), 2868 (2020).
9. U. Leiter, U. Keim, T. Eigentler, A. Katalinic, B. Holleczek, P. Martus, and C. Garbe, "Incidence, mortality, and trends of nonmelanoma skin cancer in Germany," *J. Invest. Dermatol.* **137**(9), 1860–1867 (2017).
10. C. M. Oh, H. Cho, Y. J. Won, H. J. Kong, Y. H. Roh, K. H. Jeong, and K. W. Jung, "Nationwide trends in the incidence of melanoma and non-melanoma skin cancers from 1999 to 2014 in South Korea," *Cancer. Res. Treat.* **50**(3), 729–737 (2018).
11. C. Liu, B. Wu, B. L. A. Sordillo, S. Boydston-White, V. Sriramoju, C. Zhang, H. Beckman, L. Zhang, Z. Pei, L. Shi, and R. R. Alfano, "A pilot study for distinguishing basal cell carcinoma from normal human skin tissues using visible resonance Raman spectroscopy," *J. Cancer Metastasis Treat.* **2019**(4), 1–14 (2019).
12. P. S. Karia, J. Han, and C. D. Schmults, "Cutaneous squamous cell carcinoma: estimated incidence of disease, nodal metastasis, and deaths from disease in the United States, 2012," *J. Am. Acad. Dermatol.* **68**(6), 957–966 (2013).
13. G. Merlino, M. Herlyn, D. E. Fisher, B. C. Bastian, K. T. Flaherty, M. A. Davies, J. A. Wargo, C. Curiel-Lewandrowski, M. J. Weber, S. A. Leachman, M. S. Soengas, M. McMahon, J. W. Harbour, S. M. Swetter, A. E. Aplin, M. B. Atkins, M. W. Bosenberg, R. Dummer, J. Gershenwald, A. C. Halpern, D. Herlyn, G. C. Karakousis, J. M. Kirkwood, M. Krauthammer, R. S. Lo, G. V. Long, G. McArthur, A. Ribas, L. Shuchter, J. A. Sosman, K. S. Smalley, P. Steeg, N. E. Thomas, H. Tsao, T. Tueting, A. Weeraratna, G. Xu, R. Lomax, S. Martin, S. Silverstein, T. Turnham, and Z. A. Ronai, "The state of melanoma: challenges and opportunities," *Pigm. Cell Melanoma Res.* **29**(4), 404–416 (2016).
14. T. A. Warren, B. Panizza, S. V. Porceddu, M. Gandhi, P. Patel, M. Wood, C. M. Nagle, and M. Redmond, "Outcomes after surgery and postoperative radiotherapy for perineural spread of head and neck cutaneous squamous cell carcinoma," *Head Neck.* **38**(6), 824–831 (2016).
15. A. S. Weinberg, C. A. Ogle, and E. K. Shim, "Metastatic cutaneous squamous cell carcinoma: An update," *Dermatol. Surg.* **33**, 885–899 (2007).
16. I. Hoorens, K. Vossaert, K. Ongenaes, and L. Brochez, "Is early detection of basal cell carcinoma worthwhile? Systematic review based on the WHO criteria for screening," *Br. Br. J. Dermatol.* **174**(6), 1258–1265 (2016).
17. H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *Lancet Oncol.* **3**(3), 159–165 (2002).
18. M. E. Vestergaard, P. Macaskill, P. E. Holt, and S. W. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting," *Br. Br. J. Dermatol.* **159**, 669–676 (2008).
19. D. Piccolo, A. Ferrari, K. Peris, R. Daidone, B. Ruggeri, and S. Chimenti, "Dermoscopic diagnosis by a trained clinician vs a clinician with minimal dermoscopy training vs computer-aided diagnosis of 341 pigmented skin lesions: A comparative study," *Br. Br. J. Dermatol.* **147**(3), 481–486 (2002).
20. R. J. Friedman, D. Gutkowitz-Krusin, M. J. Farber, M. Warycha, L. Schneider-Kels, N. Papastathis, M. C. Mihm, P. Googe, R. King, V. G. Prieto, A. W. Kopf, D. Polsky, H. Rabinovitz, M. Oliviero, A. Cagnetta, D. S. Rigel, A. Marghoob, J. Rivers, R. Johr, J. M. Grant-Kels, and H. Tsao, "The diagnostic performance of expert dermoscopists vs a computer-vision system on small-diameter melanomas," *Arch. Dermatol.* **144**(4), 476–482 (2008).
21. E. Stiegel, C. Lam, M. Schowalter, A. K. Somani, J. Lucas, and C. Poblete-Lopez, "Correlation between original biopsy pathology and Mohs intraoperative pathology," *Dermatol. Surg.* **44**(2), 193–197 (2018).
22. I. Kuzmina, I. Diebele, D. Jakovels, J. Spigulis, L. Valeine, J. Kapostinsh, and A. Berzina, "Towards noncontact skin melanoma selection by multispectral imaging analysis," *J. Biomed. Opt. Lett.* **16**(6), 060502 (2011).
23. L. Lim, B. Nichols, M.R. Migden, N. Rajaram, J.S. Reichenberg, M.K. Markey, M.I. Ross, and J.W. Tunnell, "Clinical study of noninvasive in vivo melanoma and nonmelanoma skin cancers using multimodal spectral diagnosis," *J. Biomed. Opt.* **19**(11), 117003 (2014).
24. A. Pardo, J. A. Gutiérrez-Gutiérrez, I. Lihacova, J. M. López-Higuera, and O. M. Conde, "On the spectral signature of melanoma: a non-parametric classification framework for cancer detection in hyperspectral imaging of melanocytic lesions," *Biomed. Opt. Express* **9**(12), 6283–6301 (2018).
25. J.A. Gutiérrez-Gutiérrez, A. Pardo, E. Real, J. M. López-Higuera, and O. M. Conde, "Custom scanning hyperspectral imaging system for biomedical applications: modeling, benchmarking, and specifications," *Sens.* **19**(7), 1692 (2019).
26. A. M. Hosking, B.J. Coakley, D. Chang, F. Talebi-Liasi, S. Lish, S. W. Lee, A. M. Zong, I. Moore, J. Browning, S. L. Jacques, J. G. Krueger, M. K. Kelly, K. G. Linden, and D. S. Gareau, "Hyperspectral imaging in automated digital dermoscopy screening for melanoma," *Lasers Surg. Med.* **51**(3), 214–222 (2019).
27. J. E. Fowler, "Compressive pushbroom and whiskbroom sensing for hyperspectral remote-sensing imaging," *IEEE Int. Conf. Image Process.* 684–688 (2014).
28. E. Salomatina, B. Jiang, J. Novak, and A. N. Yaroslavsky, "Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range," *J. Biomed. Opt.* **11**(6), 064026 (2006).
29. B. C. Q. Truong, H. D. Tuan, and H. T. Nguyen, "Near-infrared parameters extraction: a potential method to detect skin cancer," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 33–36 (2013).
30. A. Spreinat, G. Selvaggio, L. Erpenbeck, and S. Kruss, "Multispectral near infrared absorption imaging for histology of skin cancer," *J. Biophotonics*, **13**(1), 1–8 (2020).

31. R. Leon, B. Martinez-Vega, H. Fabelo, S. Ortega, V. Melian, I. Castaño, G. Carretero, P. Almeida, A. Garcia, E. Quevedo, J.A. Hernandez, B. Clavo, and G.M. Callico, "Non-invasive skin cancer diagnosis using hyperspectral imaging for in-situ clinical support," *J. Clin. Med.* **9**(6), 1662 (2020).
32. S. G. Brouwer de Koning, P. Weijtmans, M. B. Karakullukcu, C. Shan, E. J. M. Baltussen, L. A. Smit, R. L. P. van Veen, B. H. W. Hendriks, H. J. C. M. Sterenborg, and T. J. M. Ruers, "Toward assessment of resection margins using hyperspectral diffuse reflection imaging (400–1,700 nm) during tongue cancer surgery," *Lasers Surg. Med.* **52**(6), 496–502 (2020).
33. M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, L. Myers, B. D. Sumer, and B. Fei, "Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning," *Cancers* **11**(9), 1367 (2019).
34. M. Salmivuori, N. Neittaanmäki, I. Pölönen, L. Jeskanen, E. Snellman, and M. Grönroos, "Hyperspectral imaging system in the delineation of ill-defined basal cell carcinomas: a pilot study," *J. Eur. Acad. Dermatol. Venereology*, **33**(1), 71–78 (2019).
35. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
36. C. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
37. P. Geladi, J. Burger, and T. Lestander, "Hyperspectral imaging: calibration problems and solutions," *Chemom. Intell. Lab. Syst.* **72**(2), 209–217 (2004).
38. S. Del Pozo, P. Rodríguez-González, D. Hernández-López, and B. Felipe-García, "Vicarious radiometric calibration of a multispectral camera on board an unmanned aerial system," *Remote Sens.* **6**(3), 1918–1937 (2014).
39. J. Cohen, *Statistical Power Analysis for Behavioural Sciences* (Routledge, 1988).
40. N. M. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *J. Stat. Model. Anal.* **2**(1), 21–33 (2011).
41. J. Höhle and M. Höhle, "Accuracy assessment of digital elevation models by means of robust statistical methods," *ISPRS J Photogramm. Remote Sens.* **64**(4), 398–406 (2009).
42. A. Hasan, P. Pilesjö, and A. Persson, "The use of LIDAR as a data source for digital elevation models – a study of the relationship between the accuracy of digital elevation models and topographical attributes in northern peatlands," *Hydrol. Earth Syst. Sci.* **8**(3), 5497–5522 (2011).
43. M. Herrero-Huerta, R. Lindenbergh, and P. Rodríguez-González, "Automatic tree parameter extraction by a mobile LiDAR System in an urban context," *PLoS One* **13**(4), e0196004 (2018).
44. F. J. Ariza-López, J. Rodríguez-Avi, D. González-Aguilera, and P. Rodríguez-González, "A new method for positional accuracy control for non-normal errors applied to airborne laser scanning data," *Appl. Sci.* **9**, 3887 (2019).
45. M. Rodríguez-Martín, P. Rodríguez-González, E. Ruiz de Oña Crespo, and D. González-Aguilera, "Validation of portable mobile mapping system for inspection tasks in thermal and fluid-mechanical facilities," *Remote Sens.* **11**(19), 2205–2219 (2019).
46. L.A. Courtenay, D. Herranz-Rodrigo, R. Huguet, M.Á. Maté-González, D. González-Aguilera, and J. Yravedra, "Obtaining new resolutions in carnivore tooth pit morphological analyses: a methodological update for digital taphonomy," *PLoS One* **15**(10), e0240328 (2020).
47. E. Nocerino, F. Menna, F. Remondino, I. Toschi, and P. Rodríguez-González, "Investigation of indoor and outdoor performance of two portable mobile mapping systems," in *Proceedings Volume 10332, Videometrics, Range Imaging, and Applications XIV* (2017). <https://doi.org/10.1117/12.2270761>
48. M. S. Bartlett, "Properties of sufficiency and statistical tests," *Proc. R. Soc. Lond. A* **160**(901), 268–282 (1937).
49. H. Levene, "Robust tests for equality of variances," in *Contributions to Probability and Statistics*, I. Olkin and P. Alto, eds. (Stanford University Press, 1960), pp. 278–292.
50. H. Hotelling, "A generalized T Test and measure of multivariate dispersion," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, ed., (University of California Press, 1951), pp. 23–41.
51. M. Hollander and D.A. Wolfe, *Nonparametric Statistical Methods* (John Wiley & Sons, 1973) pp. 27–75.
52. J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory* **37**(1), 145–151 (1991).
53. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *IEEE Trans. Inf. Theory* **49**(7), 1858–1860 (2003).
54. L. A. Courtenay, "Code and Data for the HYPER-SKINCARE project and paper titled 'Hyperspectral Imaging and Robust Statistics in Non-Melanoma Skin Cancer Analysis,'" GitHub (2021), https://github.com/LACourtenay/HyperSkinCare_Statistics
55. R. L. Wasserstein and N. A. Lazar, "The ASA Statement on p -Values: Context, process and purpose," *Am. Stat.* **70**(2), 129–133 (2016).
56. R. L. Wasserstein, A. L. Schirm, and N. A. Lazar, "Moving to a world beyond " $p < 0.05$ "", *Am. Stat.*, **73**(sup1), 1–19 (2019).
57. D. J. Benjamin and J. O. Berger, "Three recommendations for improving the use of p -values," *Am. Stat.* **73**(sup1), 186–191 (2019).
58. D. Colquhoun, "The False Positive Risk: a proposal concerning what to do about p -values," *Am. Stat.* **73**(sup1), 192–201 (2019).
59. D. Colquhoun, "The reproducibility of research and the misinterpretation of p values," *R. Soc. Open Sci.* **4**(12), 171085 (2017).

60. T. Sellke, M. J. Bayarri, and J. O. Berger, "Calibration of p values for testing precise null hypotheses," *Am. Stat.* **55**(1), 62–71 (2001).
61. L. A. Courtenay, D. Herranz-Rodrigo, D. González-Aguilera, and J. Yravedra, "Developments in Data Science Solutions for Carnivore Tooth Pit Classification," *Sci. Rep.* **11**(1), 10209 (2021).
62. A. G. Nerlich, H. Rohrbach, B. Bachmeier, and A. Zink, "Malignant tumors in two ancient populations: an approach to historical tumor epidemiology," *Oncol. Rep.* **16**, 197–202 (2006).
63. M. Binder, C. Roberts, N. Spencer, D. Antoine, and C. Cartwright, "On the antiquity of cancer: evidence for metastatic carcinoma in a young man from Ancient Nubia (1200BC)," *PLoS One* **9**(3), e90924 (2014).
64. G. Fornaciari, "Histology of ancient soft tissue tumors: a review," *Int. J. Paleopathol.* **21**, 64–76 (2018).
65. M. R. Whitney, L. Mose, and C. A. Sidor, "Odontoma in a 255-million-year-old mammalian forebear," *JAMA Oncol.* **3**(7), 998–1000 (2017).
66. E. J. Odes, P. S. Randolph-Quinney, M. Steyn, Z. Throckmorton, J.S. Smilg, B. Zipfel, T.N. Augustine, F. Beer, J.W. Hoffman, R.D. Franklin, and L.R. Berger, "Earliest hominin cancer: 1.7-million-year-old osteosarcoma from Swartkrans Cave, South Africa," *S. Afr. J. Sci.* **112**(7/8), 1–5 (2016).
67. J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Meryl, M. Piéros, A. Znaor, I. Soerjomatram, and F. Bray, "Global cancer observatory: cancer today," International Agency for Research on Cancer, Lyon. (2020) <https://gco.iarc.fr/today>
68. G.P. Guy, S.R. Machlin, D.U. Ekwueme, and K.R. Yabroff, "Prevalence and costs of skin cancer treatment in the US, 2002-2006 and 2007-2011," *Am. J. Prev. Med.* **48**(2), 183–187 (2015).
69. C. Vrinten, L. M. McGregor, M. Heinrich, C. Wagner, J. Waller, J. Wardle, and G. B. Black, "What do people fear about cancer? A systematic review and meta-synthesis of cancer fears in the general population," *Psycho-Oncology* **26**(8), 1070–1079 (2017).
70. S. L. Jacques, "Optical properties of biological tissues: a review," *Phys. Med. Biol.* **58**(11), R37–R61 (2013).
71. D. S. Gareau, J. C. Rosa, S. Yagerman, J. A. Carucci, N. Gulati, F. Hueto, J. L. DeFazio, M. Suárez-Fariñas, A. Marghoob, and J. G. Krueger, "Digital imaging biomarkers feed machine learning for melanoma screening," *Exp.Dermatol.* **26**(7), 615–618 (2017).