scientific reports

Check for updates

OPEN

Feasibility study of hyperspectral colposcopy as a novel tool for detecting precancerous cervical lesions

Carlos Vega[1✉], Norberto Medina[2], Laura Quintana-Quintana[1], Raquel Leon[1], Himar Fabelo[1,3,4], Jorge Rial[5], Alicia Martín[2] & Gustavo M. Callico[1]

Cervical cancer remains a major global health concern, with a specially alarming incidence in younger women. Traditional detection techniques such as the Pap smear and colposcopy often lack sensitivity and specificity and are highly dependent on the experience of the gynaecologist. In response, this study proposes the use of Hyperspectral Imaging, a pioneering technology that combines traditional imaging with spectroscopy to provide detailed spatial and spectral information. Over a period of six-months, our custom-designed hyperspectral colposcope was used on 62 patients. The gathered data underwent a specialized preprocessing workflow using a PCA-based strategy for unsupervised segmentation of the cervical region. This process extracted spectral signatures from various tissue types, and our subsequent statistical analysis highlighted its ability to detect differences and alterations in the cervical tissue. This offers a promising avenue for improving the precision of cervical lesion diagnosis.

Cervical cancer is recognized as a major global health concern, ranking fourth in terms of incidence and third in mortality, among women worldwide in 2020 and affecting younger populations under 45 years old at an alarming rate[1]. In Europe, the incidence of cervical cancer is 15 cases per 100,000 inhabitants, although there is a high variability between regions[2]. For example, in Canary Islands (Spain), where this study was conducted, the age-standardized incidence of cervical cancer in 2018 was 8.5 cases per 100,000 inhabitants, compared with a national average standardized incidence of 5.4 cases per 100,000 inhabitants[3,4]. These data highlight the necessity for enhanced detection and screening tools for cervical cancer, especially at early stages[4].

The invasive squamous epithelial carcinoma of the cervix is developed from precursor lesions, specifically the high-grade squamous Cervical Intraepithelial Neoplasia (CIN). For this pathology to grow, the patient must have been persistently infected with the Human Papilloma Virus (HPV)[5], although this is not a sufficient condition for its development. The natural history of cervical cancer starts several years or decades earlier in the patient, with the HPV infection in the keratinocytes of the basal stratum of the squamous epithelium. Initially, low-grade benign lesions appear, which progress to squamous cell high-grade lesions with a high potential for malignancy[6]. CIN can be classified on a scale from one to three, depending on how deeply the abnormal cells affect the squamous epithelium. Destruction and involvement of the lower third of the squamous epithelium is called CIN 1, a benign lesion that regresses spontaneously in 60% of cases[6]. In the cases where the degree of involvement of the squamous epithelium affects two-thirds, it is called CIN 2. Finally, if the abnormal cells affect more than two-thirds of the squamous epithelium, it is called CIN 3. Both, CIN 2 and CIN 3 lesions, are identified as high-grade lesions.

The Papanicolaou test (Pap smear)[7] is a primary cervical screening method to detect potential precancerous and cancerous lesions of the cervix and is used in the early detection of possible CIN. The test is performed by collecting cells from the cervix with a cytobrush and examining the samples under a microscope[8]. Cytology is usually combined with an HPV test to determine if the patient is infected with the virus and which type. If the test shows abnormal results, the patient undergoes a colposcopic examination. The colposcope is a microscope with stereoscopic vision, low magnification, and a powerful white light source. To facilitate diagnosis, acetic acid

[1]Research Institute for Applied Microelectronics (IUMA), University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain. [2]Complejo Hospitalario Universitario Insular Materno Infantil (CHUIMI), Servicio Canario de Salud (SCS), Las Palmas de Gran Canaria, Spain. [3]Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Las Palmas de Gran Canaria, Spain. [4]Research Unit, Hospital Universitario de Gran Canaria Dr. Negrin, Las Palmas de Gran Canaria, Spain. [5]OPTOMIC ESPAÑA S.A, Madrid, Spain. ✉email: cvega@iuma.ulpgc.es

or Lugol's iodine may also be used [9]. This evaluation, combined with colposcopically guided biopsies, is critical in determining whether to return the patient to routine screening, more intensive surveillance, or treatment[10]. Although this is an important tool in the diagnosis of preneoplastic lesions, it has limitations. Patient assessment is subjective and operator dependant[11]. This method generally has low sensitivity (between 68.5 and 77%) and specificity (between 75.9 and 82%)[12,13], and often leads to unnecessary biopsies[14].

These challenges have motivated the development of a more objective tool, for example using medical imaging. These images are processed to detect abnormal regions in the cervix since the early 2000s[15,16]. In recent years, advancements in deep learning algorithms have led to noticeable improvements in this area, with systems achieving sensitivity and specificity results over 85%[17]. These new systems are based on advanced deep learning architectures that classify cervical lesions using one or more colposcopy images. These images capture the whitening process of the acetic acid in the cervical tissue, a phenomenon referred to as acetowhitening. Some of these new algorithms, as the one presented by Peng et al.[18], combined two different captures: a pre-acetic and post-acetic image of the cervix[18]. By combining the information from both domains, a deep learning network evaluated the acetowhitening process of the tissue with similar information as a trained clinician. This concept was extended by Yu et al., where seven images were combined for the evaluation: a sequence of five images capturing the acetowhitening process, one image taken with a green filter to enhance vascular patterns, and another image using the iodine test[19]. Based in this data, they proposed a neural network architecture that analysed all this information combined and made a CIN-level prediction. This approach with multiple captures combined with deep learning models achieved noticeable results, with authors claiming results over 95% in sensitivity and specificity using their dataset[19]. In terms of application, these systems present two main limitations, on one hand they force the gynaecologist to keep the acquisition system static during the capturing process, interfering with the normal clinical workflow. Furthermore, they offer limited information in terms of interpretability of the results, which is currently a serious issue in terms of legislation and its application in the clinical practice[20–22].

Multispectral (MS) Imaging (MSI) and Hyperspectral (HS) Imaging (HSI), which combines traditional image and spectroscopy methods to capture spatial and spectral information, have emerged as a novel approach in the medical field to overcome all these challenges, especially for cancer analysis[23]. MS and HS images are represented as three-dimensional cubes, where the spatial dimensions (x, y) form the horizontal planes, while the spectral part introduces the third dimension (λ). The reflectance-based MS studies mostly combine a RGB image with a set of images from narrow wavelengths that extend the amount of information. One recent study by Wang et al. presented a MSI system that combined the RGB with two bands: 450 and 545 nm[24]. Using this approach, the authors achieved a sensitivity of 85.3% and a specificity of 70.8% in the discrimination between diseased cervical tissue and normal tissue. Furthermore, Karthika et al. focused on the development of the technology and its portability[25]. This work combined the RGB with other four bands: 450, 545, 575, and 610 nm into a small endoscopic system. The authors conducted a preliminary study a in a small set of patients, examining the differences between healthy epithelial tissue and abnormal regions. They observed preliminary differences by calculating the pixel-wise ratios of 610 to 450 nm and 610 to 545 nm.

However, there is still room for improvement, HSI provides hundreds of narrow and contiguous spectral bands, presented as spectral signatures for each pixel of the image, providing specific details about the captured material[26]. The studies of HSI in cervical cancer are not limited to a small number of measurable bands and allow a detailed evaluation of the tissue alterations and different biochemical features[27]. In 2002, before the application of HSI technology, the use of reflectance spectroscopy for in-vivo detection of cervical cancer was studied by Mirabal et al. using a spectrograph in the wavelength range between 355 and 655 nm with a spectral resolution of 2.5 nm[28]. This study achieved preliminary high sensitivity and specificity in the CIN 2 and 3 detections. In 2004, Wang et al. studied the application of spectroscopy in the uterine-cervical epithelial tissue using an approach based on Monte Carlo simulations in the visible spectrum (380–640 nm). The work revealed that the bands with the most significant variability in the spectral signatures of the different tissues were the Soret (420 nm) and the Q-band (540–580 nm)[27]. In 2016, a preliminary study presented the use of multi-scale HSI for detecting cervical neoplasia at tissue and cellular levels[29]. HS images were captured with a monochrome industrial camera, attached to an Acousto-Optic Tuneable Filter (AOTF) with a filtering range between 500 and 900 nm, with an FWHM (Full Width at Half Maximum) for each band of 2 nm. The samples were captured in-vivo, during the patient examination, and ex-vivo, from tissue slices in the microscope. The authors showed favourable results by analysing the spectrum between 600 and 800 nm. These studies have demonstrated the potential of HSI for detecting CIN. Furthermore, the promising results of HSI diagnosing cancer diseases and supporting medical-guided surgery[23] remark the importance of developing new tools for cervical examination with HSI capabilities.

Within this context, our study aims to introduce a new clinical workflow for the in-vivo analysis of cervical lesions, based on the HS colposcope, a HS acquisition system previously designed by this group[30]. Prior to clinical application, this system was optically characterised to validate its spatial and spectral response. The HS colposcope was used in a six-month data acquisition campaign in which 62 patients were examined. The captured HS images underwent a custom preprocessing workflow for its analysis and characterisation of the spectral signatures of the cervical tissue. A technique was proposed for reducing the spectral dimensions based on the representation of the cubes in pseudo RGB (Red, Green, and Blue) format based on the PCA (Principal Component Analysis). This process allows unsupervised detection of the cervical perimeter and segmentation the spectral signatures of the main areas of the cervix, the endocervix and exocervix, and also the outlier pixels due to specular reflections or external elements, as the IUDs (Intrauterine Devices). The invasive cervical cancer tissue was manually segmented to perform a statistical study against the other tissue types. The results prove that this technology is able to extract variations in the tissue present on the cervix with a statistically significant, showcasing the potential of this technology to develop a decision support tool capable of assisting gynaecologists in the detection of cervical lesions during routine clinical practice.

Results

HS colposcope integration and spatial-spectral characterisation

The IMEC Snapscan VNIR camera (IMEC, Belgium) was coupled to a commercial colposcope, the Optomic OP-C2 (OPTOMIC, Spain), whose Infrared (IR) filter was removed to capture information in the Visible and Near-Infrared (VNIR) range and whose light source had to be replaced from a LED (Light-Emitting Diode) model to a halogen model (see "Methods" section). A schematic of the integrated HS colposcope is shown in Fig. 1. The light of the external source is projected (Fig. 1a) onto the sample and then reflected back to the colposcope, following an optical path towards three apertures: HS camera (Fig. 1c); RGB camera (Fig. 1d) and binoculars (Fig. 1e). Internally, the colposcope incorporates multiple lenses, through which light must be corrected and collimated within the internal tube, then split and focused, and ultimately reaching the three apertures. The focus adjustment of the splitter must be modified so that the camera with an extended flange focal distance can focus at the working distance (WD) of the colposcope (26 cm from the cervical surface of the patient). The HS camera employed in this work captures a wider spectral range (470–900 nm) than RGB cameras, which means that further distortion could be introduced due to the interaction between different spectral bands and the lenses. This effect motivates the spatial and spectral characterisation of the HS colposcope (see "Methods" section).

The spatial resolution was determined by capturing a certified ruler (quality resolution chart), containing alternating white and black lines of 1 mm thickness (Fig. S1a in the Supplementary Material) extracting one profile of the mean monochromatic image and calculating its first derivative (Fig. 2a)[31]. Minima and maxima show the change between white and black lines, the mean distance between them was 14.7 pixels for 1 mm lines, thus, pixel size was calculated to be 67.87 µm. Since the HS cube are 1000×900 pixels, the captured FOV is 67.87×61.08 mm. Knowing that the sensor size is 5.50×4.95 mm, the HS colposcope system presents a total magnification power of $12.3 \times$.

The highest spatial resolution that the system can capture is determined by the maximum spatial frequency achieved by an optical system. To measure it, the Siemens Star target (Fig. S1b in the Supplementary Material) was captured with the HS colposcope. The target has an unresolved core radius of 50 µm, since pixel size was calculated to be 67.87 µm, its theoretical maximum reachable frequency (Nyquist Frequency) is 168.84 lp/mm. To determine its empirical maximum frequency, the Modulation Transfer Function (MTF) was calculated at different frequencies (Fig. 2b)[32]. Considering the MTF10[33], results show that the HS colposcope system can keep a contrast in frequencies up to 9.49 LP/mm.

Optical aberration should also be measure due to the nature of the HS colposcope optics. A diffuse reflectance grid distortion target was captured (Fig. S1c in the Supplementary Material). Circles were automatically found from 550 to 785 nm since extreme bands are too noisy for circles' detection. Distortion results over the monochromatic image give a mean barrel distortion of -3.21%, however, when examining each wavelength separately, there is a decrease of the barrel distortion with respect to the wavelength (Fig. 2c). Such small values of distortion usually go unnoticed by the human eye[34]. Furthermore, not only the system has a barrel distortion but also, the light intensity is misaligned with respect to the centre of the optics (Fig. 1). Using an HS image of a white reference, we can see how the light is distributed along the sensor (Fig. 2d). There is no light on the corners because the colposcope optics and the sensor are circular and square shaped, respectively. The maximum light spot was found at pixel (392, 471) of an HS image with a spatial size of 900×1000 pixels.

Spectral characterisation is also needed to evaluate the ability of the HS colposcope to accurately capture and reproduce the spectral signatures of the recorded scene. From a capture of the Zenith Polymer SG3333 (SphereOptics GmbH, Germany) target (Fig. S1d in the Supplementary Material), and taking as reference the light centre previously calculated, the mean spectra of two different radial profiles ($r = \{70, 310\}$ pixels) was compared to the certified data provided by the manufacturer (Fig. 2e). The mean Spectral Correlation Measure (SCM) of the spectra against the reference achieved a value of 0.93.
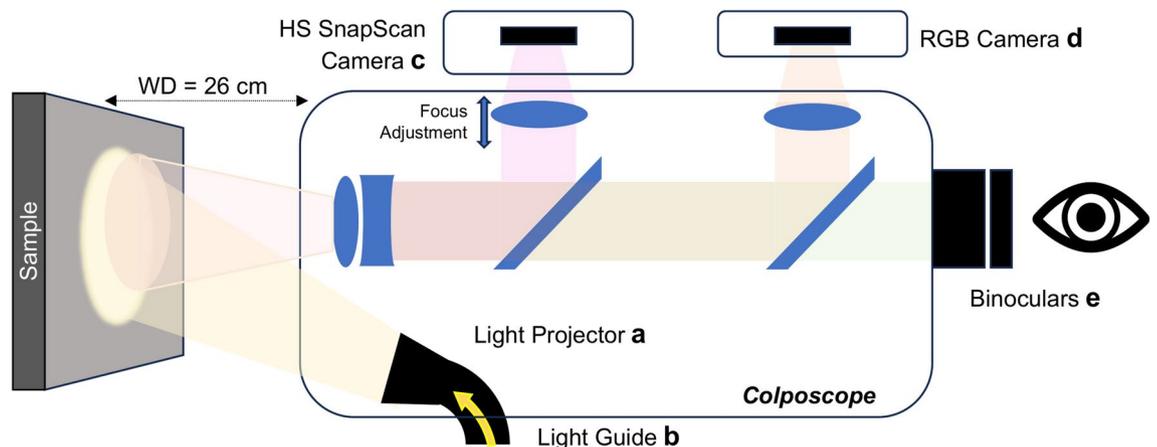

Fig. 1. Schematic of the HS Colposcope optics. (**a**) Light projector over the sample. (**b**) Light guide from the light source. (**c**) HS camera. (**d**) RGB camera. (**e**) Binoculars for gynaecologist vision.
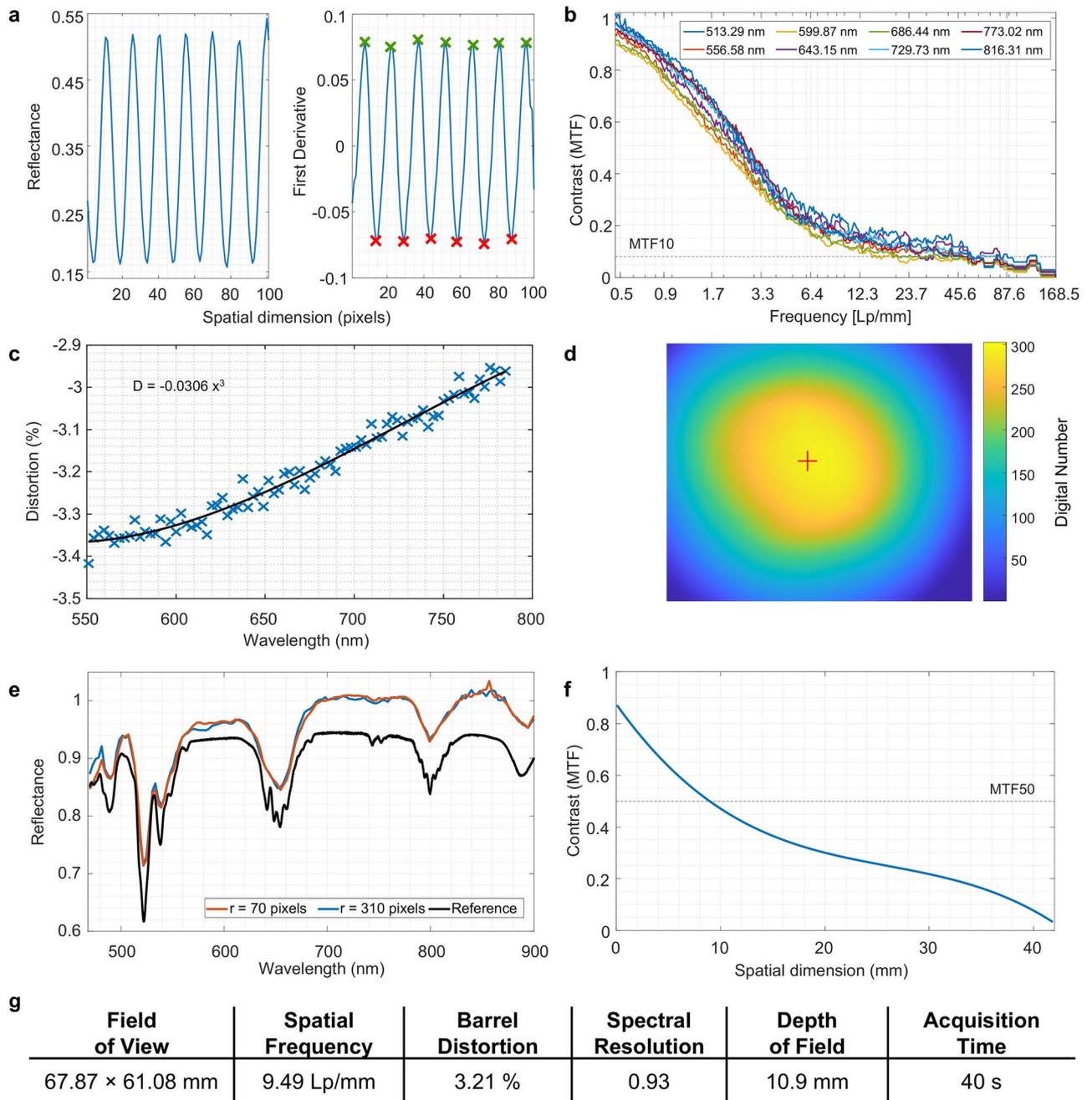
Fig. 2. Results of the HS colposcope characterisation. (**a**) Reflectance profile from the ruler and its first derivative (green and red crosses meaning local maxima and minima, respectively). (**b**) Contrast with respect to frequency for different wavelengths. (**c**) Distortion values with respect to the wavelength, showing the characterised function of the graph. (**d**) Light distribution on the sensor and maximum light spot. (**e**) Mean spectra from two different radial profiles and a certified reference. (**f**) Contrast at different working distances, showing optimal DoF at MTF50. (**g**) Summary of characterization parameters.

Moreover, the Depth of Field (DoF) of the system provides valuable information about the optimal range of distances over which objects appear acceptably sharp in the captured images. The contrast of the 15 lp/mm profile from the captured DoF 5–15 target (Fig. S1e in the Supplementary Material) was analysed. DoF was found to be 10.9 mm following MTF50 (Fig. 2f).

Acquisition time for each cube is 45 s with an integration time is 18 ms. Although this acquisition time is significantly longer compared to traditional RGB systems, which typically have an acquisition time of about 16 ms, the sensor in this system must capture at least 256 frames and combine them into a cube of 192 bands, of which the HS image is composed of 158 bands.

Finally, a summary of all the characterization parameters is presented in Fig. 2g.

Study population

Women over the age of 18 who were treated at the *Complejo Hospitalario Universitario Insular Materno Infantil* in Las Palmas de Gran Canaria, Spain, were eligible for the study. Patients were recruited during screening and diagnostic evaluation by a gynaecologist specialising in cervical cancer. The gynaecologist conducted a regular consultation, collected the patient's sociodemographic and clinical information and captured the cervix using the HS colposcope. The study adhered to ethical guidelines (see Methods), obtaining informed consent from all participants. Over a six-month period, 124 HS images were acquired from 62 different patients. The lesions were diagnosed according to the clinical procedure, where patients first underwent a colposcopic examination, followed by liquid-based cytology (ThinPrep Pap Test PreservCyt™ Solution) and a HPV testing (Cobas 4800). A biopsy was performed in cases of pathological colposcopic findings. The final diagnosis was based on biopsy (abnormal colposcopies) or cytology (normal colposcopies). The latter usually identifies low-grade lesions.

A summary of the patient demographics and clinical data is presented in Table 1. Ages ranged from 28 to 58 years, with a median age of 43 years. Of these 62 patients, 19 (31%) were healthy patients (not infected with HPV and not affected by any lesion). Regarding the HPV test, 21 (34%) were HPV negative, while 13 (21%) were infected with the HPV-16, 2 (3%) were infected with the HPV-18, 4 (6%) were infected with the HPV-16 and others, and 22 (38%) were infected with HPV other than HPV-16 and HPV-18. The latter group, although not considered high risk, had a significant number of advanced lesions (8 CIN 3 lesions and 1 case of invasive cancer). After analysing the cytology results, 50% of the cases ($N = 31$) were diagnosed as normal, with no cervical lesion present. The remaining 50% presented High-Grade Squamous Intraepithelial Lesion (HSIL) and Low-Grade Squamous Intraepithelial Lesion (LSIL) (26%, $N = 16$ and 23%, $N = 14$, respectively) and one patient (2%, $N = 1$) was diagnosed as invasive cervical cancer. Finally, colposcopy results showed that only 67% ($N = 8$) of patients with a HSIL were correctly diagnosed, while invasive cervical cancer was correctly identified in 62% of patients ($N = 5$). The correlation between colposcopy and cytology results is low and correlates to previous studies[35], thus supporting the need for a more precise diagnostic tool. The statistical analysis of the clinical data (Table 1) shows that the population in this study presents a notorious diversity and a large variation between the different variables. The significance of each variable to the diagnosis of the patient was studied by the Chi-Square test of independence. The results show a high correlation ($p < 0.01$) between the *HPV test*, *cytology*, and *colposcopy* and the definitive diagnosis of the patient.

Cervix region segmentation using a PCA-based approach

In this study, only the cervical tissue was analysed and for this reason each HS image must be delineated to exclude the data outside this region. To perform the initial analysis, the HS image was reduced from 158 bands to a 3-channel image using the three first components in the PCA domain. This strategy allows the system to allocate the 89.2% of the information contained in the HS image using only 2.3% of its data size for the representation, a significant reduction compared to the original 158 bands. Figure 3 shows the synthetic RGB of three tissue samples (cervix area was manually surrounded by an expert with a blue line) and their corresponding 3-channel images (RGB), where a higher contrast is present between the different tissue areas respect to the synthetic RGB image. The 3-channel image was used to automatically segment the cervix area (Fig. 3) by a technique based on the Spectral Angle Mapper (SAM) (see "Methods" section). The results for lesions lower than CIN 3 (Fig. 3a,b) provided an accurate delineation that included the whole cervical area. However, less precision is presented for patients with invasive cervical cancer (Fig. 3c) likely due to increased variability arising from excessive bleeding and significant tissue alterations, which may lead to deviations in the spectral signature of cancerous tissue compared to the reference cervix sample. Masks of the cases with very advanced lesions were manually corrected.

Unsupervised segmentation of cervical tissue types

After delineation of the cervix region by using the PCA-based approach, this area was automatically segmented into three groups (*endocervix*, *exocervix*, and *outliers*) by applying an unsupervised technique based on the SAM algorithm (see "Methods" section). The segmentation of each HS image was visually assessed a posteriori and manually corrected if necessary. Figure 4 shows the synthetic RGB images with the endocervix and exocervix regions manually surrounded by an expert in green and blue colours, respectively. Moreover, the results of the automatic segmentation map obtained with the proposed approach are shown, highlighting the different regions with the same colours and including the outlier class in black colour.

The segmentation results were positive, with an accuracy of approximately 75% percent for the HS images with lesions up to CIN 3. The HS images with invasive cervical cancer samples from the 6 patients were manually segmented using a custom-designed labelling tool, since there were no other tissue classes to segment.

Spectral characterisation of cervical tissue

Statistical analysis was undertaken to assess the capacity of the HS image to detect variations between different tissue types in the cervix and the invasive cancerous tissue. Figure 5a presents the mean reflectance spectrum of cervical cancer (red line), exocervix (blue line), endocervix (green line) and outlier (black line) pixels. A remarkable difference is observed between the outlier mean spectrum with respect to the other classes, which supports the exclusion of the outliers from the rest of the analysis. The three tissue classes exhibit increased differences between them in the range between 470 and 550 nm and over 750 nm. The endocervix spectral signature presents a noticeable lower reflectance in the whole spectral range compared to the exocervix and cancer classes, specifically 1.29 and 1.52 times respectively. The absorption spectra (Fig. 5b) exhibits equivalent results to the reflectance mode, with also a noticeable difference between the endocervix tissue to the cancer and exocervix root mean square error (RMSE) of 0.28 and 0.25, respectively.

Since the data had a non-normal distribution, evaluated with the Lilliefors test (see "Methods" section), the differences between the spectral signatures of the tissue types were statistically analysed using a two-sided

Feature	Category/range	Total		Normal		LSIL (CIN 1)		HSIL (CIN 2–3)		Cancer		Chi²p-value
		N	%	N	%	N	%	N	%	N	%	
Age	28–36	26	42	15	58	5	19	3	12	3	12	0.518
	37–43	15	24	7	47	3	20	4	27	1	7	
	44–51	16	26	7	44	2	12	6	38	1	6	
	52–58	5	8	3	60	1	20	0	0	1	20	
Parity	= 0	29	47	13	45	9	31	4	14	3	10	0.457
	= 1	19	31	12	63	1	5	6	32	0	0	
	> 2	14	23	7	50	1	7	3	21	3	21	
Smoker	No	46	74	24	52	6	13	10	22	6	13	0.881
	Yes	16	26	8	50	5	31	3	19	0	0	
Contraceptive	Barrier	10	16	4	40	4	40	2	20	0	0	0.763
	Copper IUD	3	5	2	67	0	0	1	33	0	0	
	Hormonal IUD	3	5	1	33	1	33	1	33	0	0	
	Anovulators	23	37	14	61	4	17	2	9	3	13	
	Hormone implant	1	2	1	100	0	0	0	0	0	0	
	Tubal ligation	2	3	1	50	0	0	0	0	1	50	
	No	19	31	9	47	2	11	7	37	1	5	
	Partner's vasectomy	1	2	0	0	0	0	0	0	1	100	
Age at First Intercourse	< = 15	10	16	5	50	0	0	4	40	1	10	0.628
	16–18	46	74	23	50	10	22	8	17	5	11	
	> 18	6	9	4	66	1	17	1	17	0	0	
Number of Sexual Partners	< 5	29	47	14	48	5	17	7	24	3	10	0.138
	6–10	10	16	8	80	2	20	0	0	0	0	
	> 10	23	37	10	43	4	17	6	26	3	13	
Menopause	No	58	94	30	52	10	17	13	22	5	9	0.947
	Yes	4	6	2	50	1	25	0	0	1	25	
Transfer Area	Zone type 1	32	52	19	59	5	16	7	22	1	3	0.404
	Zone type 2	17	27	8	47	4	24	4	24	1	6	
	Zone type 3	13	21	5	38	2	15	2	15	4	31	
HPV Test	HPV 16	13	21	4	31	1	8	3	23	5	38	**0.000**
	HPV 16 and Others	4	6	1	25	1	25	2	50	0	0	
	HPV 18	2	3	2	100	0	0	0	0	0	0	
	HPV Negative	21	34	19	90	2	10	0	0	0	0	
	HPV Others	22	35	6	27	7	32	8	36	1	5	
Cytology	HSIL	16	26	0	0	0	0	11	69	5	31	**0.000**
	LSIL	14	23	1	7	11	79	2	14	0	0	
	Normal	31	50	31	100	0	0	0	0	0	0	
	Invasive Cancer	1	2	0	0	0	0	0	0	1	100	
Colposcopy Result	Invasive Cancer	8	13	1	12	1	12	1	12	5	62	**0.001**
	Grade 2	12	19	2	17	2	17	8	67	0	0	
	Grade 1	6	10	2	33	2	33	1	17	1	17	
	Normal	36	58	27	75	6	17	3	8	0	0	

Table 1. Statistical study of the clinical variables of the database. CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; IUD, intrauterine device; LSIL, low-grade squamous intraepithelial lesion; N, number. Significant are in value [bold].

Wilcoxon Rank-Sum test at the 5% significance level. The study in the reflectance mode (Fig. 5c) was performed by comparing classes in pairs: the exocervix vs endocervix, exocervix vs cancer, and endocervix vs cancer. For all the comparisons sufficient evidence were appreciated to reject the null hypothesis ($p < 0.05$), suggesting that the spectral signatures of the different tissues originate from a source with different distribution and median. Just the wavelength of 610 nm (indicated in Fig. 5c with a magenta arrow) exhibits an equivalent result between the cancerous tissue and the exocervix tissue with a $p = 0.29$, indicating an area of high similarity between tissues. The study in the absorbance mode (Fig. 5d) followed the same methodology of analysis by pairs. Equivalent results as in the reflectance analysis were achieved, supporting the hypothesis that the spectral signatures of the different tissues originate from a source with different distribution and median. The results from this study
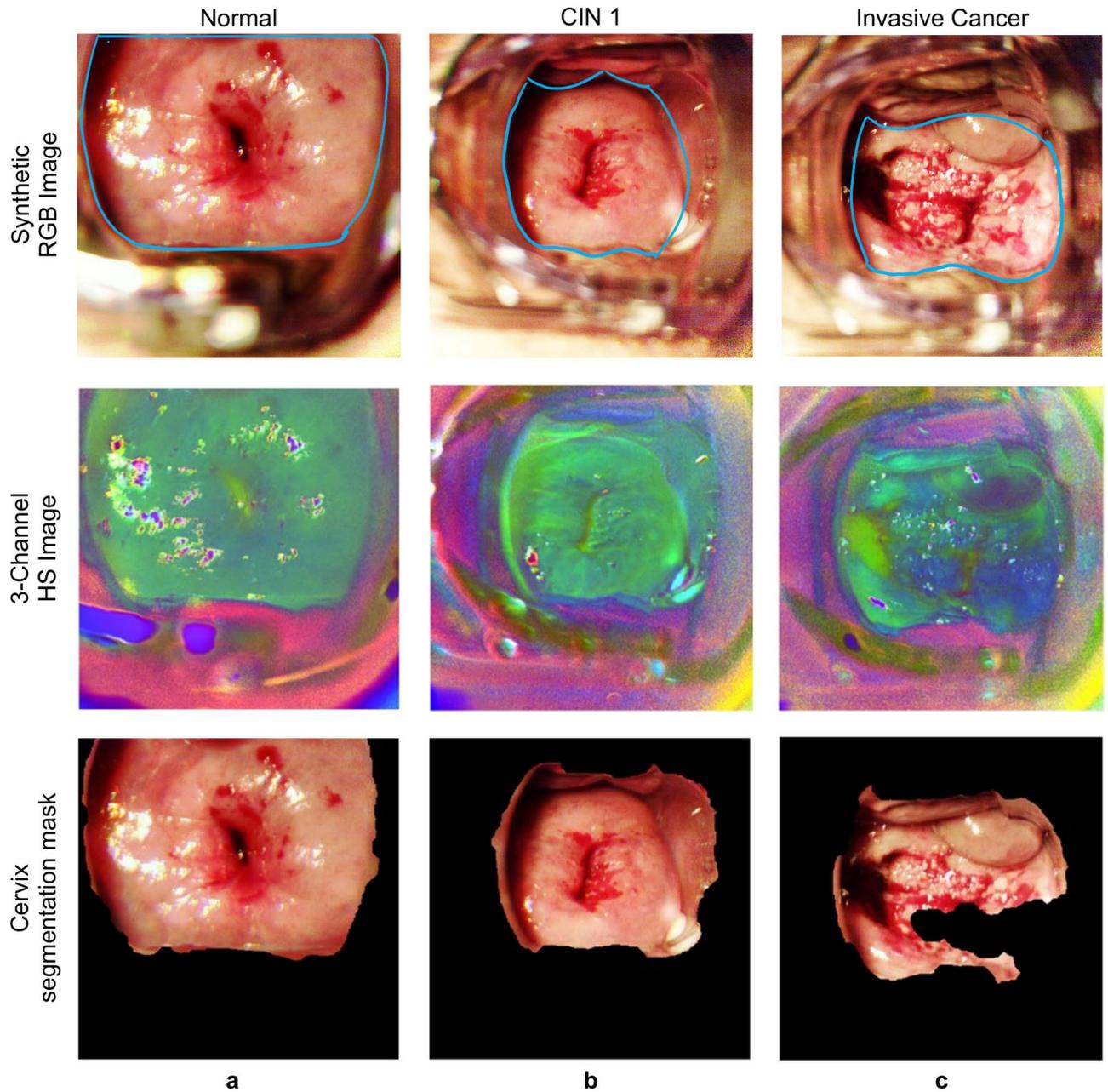
Fig. 3. Result of the PCA-based cervix segmentation approach in different tissue and lesion types. (**a**) Normal cervix. (**b**) CIN 1. (**c**) Invasive cervical cancer. Blue line represents cervix area.

suggest that the spectral signatures of the different tissue types present sufficient measurable differences that would allow its differentiation in clinical practice.

Further analysis was performed based in the most relevant absorption peaks in our acquisition range from oxyhaemoglobin (HbO_2) at 546 and 576 nm and deoxyhaemoglobin (HbO) at 750 nm[36], which are indicated with the grey areas in Fig. 5c,d. The endocervix tissue presents an absorbance 1.5 times larger than exocervix tissue in the 546 and 576 nm peaks of HbO_2, which suggests a higher blood volume in its glandular tissue. This observation could align with the endocervix's rich vascularization and characteristics. Analysing the cancer spectral signature, it is closer to exocervix rather than to endocervix tissue (RMSE of 0.09 and 0.25, respectively). This is coherent, as the most common type of cancer is the squamous cell carcinoma, which mainly affect to the squamous cells that are present in the cervix[37]. In terms of HbO_2, it is noticeable a slightly increase in the absorbance of the 546 and 576 nm bands, with a 4.3% higher absorbance in the cancer tissue compared with the exocervix tissue. The contribution of HbO in the cancer tissue, which is usually in hypoxia, is expected to be higher than in the exocervix tissue, as stated by state-of-the-art MRI studies[38,39]. Absolute absorbance values show that exocervical tissue is 20% more absorbent than cancer tissue at the HbO absorbance wavelength (750 nm). However, the cancer absorbance spectra exhibit a pronounced peak, whereas the exocervical spectra
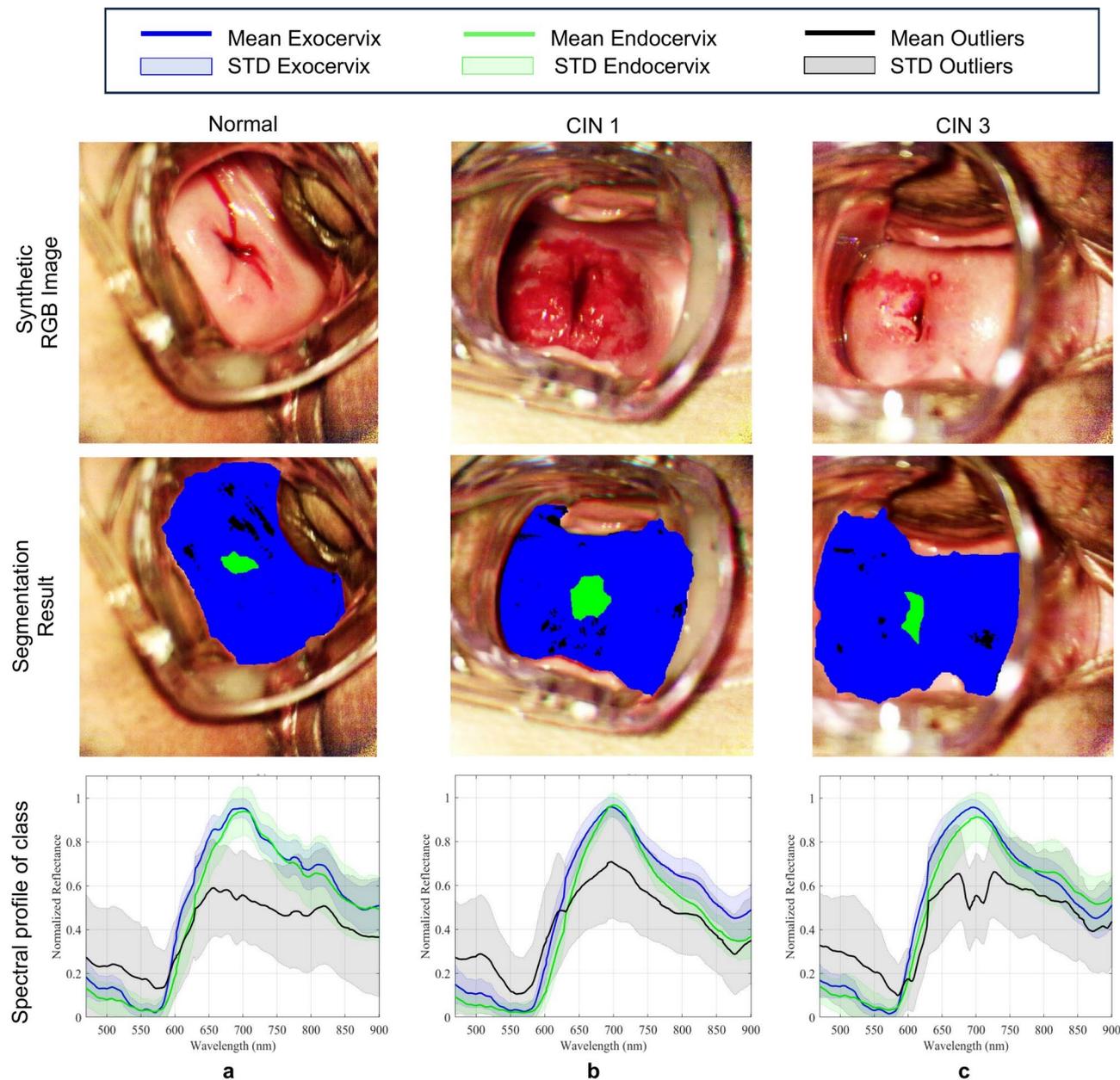
Fig. 4. Results of the unsupervised segmentation of the cervical tissue into exocervix (blue), endocervix (green), outliers (black) and spectral signatures for different lesions grades. (**a**) Normal cervix. (**b**) CIN 1. (**c**) CIN 3.

remain relatively constant. This alteration is even more pronounced in the reflectance spectrum, appearing as a distinct valley in the signal. This fact could be related to the more presence of HbO in cancer tissue rather than in the healthy exocervix. Further research is needed in this area, as the numerous different absorbers in the tissue may distort the quantitative analysis of the spectra. Fig. S2 in the supplementary material shows different HS images, organized by its lesion grade for further evaluation, at the significant absorption bands of HbO_2 and HbO.

Discussion

The predominant approach for the diagnosis of preneoplastic lesions in the cervix is the combined use of colposcopy with cervical biopsies. However, this conventional method is inherently subjective and relies on the expertise of the practitioner, leading to a considerable variability in results[11]. Previous studies encounter low sensitivity (between 68.5 and 77%) and specificity (between 75.9 and 82%) using this method[12,13], often leading to unnecessary biopsies[14]. To overcome these challenges, HSI, which combines traditional image and spectroscopy methods to capture spatial and spectral information, has emerged as a novel approach in the medical field, especially for cancer detection and tissue analysis[23].
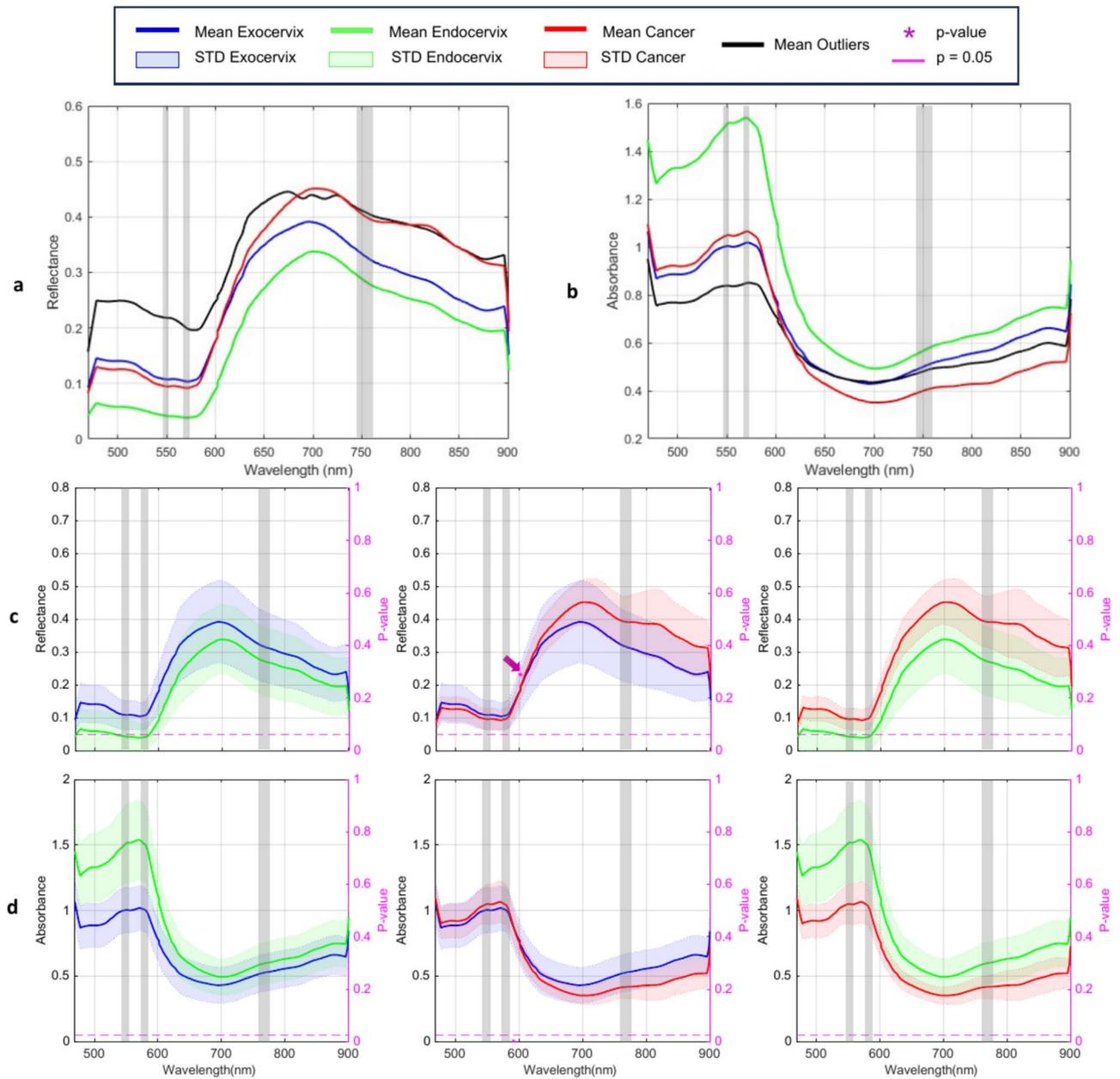
Fig. 5. Spectral comparison of the different cervical tissue types. Mean (**a**) reflectance and (**b**) absorption spectra for each of the different classes. Statistical analysis performed to evaluate the differences between the different tissue types for (**c**) reflectance and (**d**) absorption spectra. The darker grey areas represent the wavelengths related to the most relevant absorption peaks of haemoglobin in 546, 576 and 750 nm. The magenta arrow highlights the 610 nm wavelength where cancer and exocervix tissue present equivalent distribution.

This study introduces a HS colposcope, a new tool for performing the spectral characterisation of the cervical tissue based on HSI technology. This tool has the potential to detect and delineate alterations in the cervix, such as CIN, during routinary revisions and surgical operations. The optical characterisation of the HS colposcope revealed high optical resolution (14.7 pixels/mm), a noticeable maximum spatial frequency (9.49 lp/mm), low optical aberration with a mean barrel distortion of -3.21% (between 550 and 785 nm), and a depth of field of 10.9 mm. Furthermore, the spectral accuracy of the system was evaluated with a calibration target (SG3333, Edmund Optics) achieving a spectral correlation of 0.93, which validates the developed system for the precise characterisation of biological tissue. The HS images obtained have a resolution of 1000×900 pixels, with 158 bands which covers from 470 to 900 nm.

The developed system was integrated into a gynaecologist's practice, allowing the capture of cervical tissue during routine colposcopy. A database was generated with information from 62 patients, including 19 healthy patients, 13 infected with HPV and no lesion, 24 with different CIN grades, and 6 with cervical cancer. The

acquired database was statistically analysed to assess the correlation between the initial colposcopy observation and the final diagnosis derived from the biopsy. The colposcopy results showed that only 67% ($N = 8$) of patients with a HSIL were correctly diagnosed. Furthermore, invasive cervical cancer was correctly identified in 62% of patients ($N = 5$). The correlation between colposcopy and final diagnosis results is low and it is similar to previous studies, highlighting the urgent need for a more accurate diagnostic tool in this medical field.

Segmentation of the HS images into exocervix and endocervix was performed in two steps. First, the cervix area was delineated and then the different tissues were segmented. A novel approach was applied in the preprocessing and segmentation of the cervical tissue based on the dimensional reduction of the HS image by using PCA. This approach allowed compressing the 89.2% of the information present in the HS image (composed of 158 channels) into 3-channels that, which, with posterior regularisation of each channel, allowed quantifying the results to 8 bits for its RGB conversion. From this 3-channel images, the cervical region was automatically extracted from the HS image by using the SAM algorithm, requiring only a manual intervention on the samples with CIN 3 and cervical cancer. Segmentation of cervical pixels into exocervix, endocervix, outliers and cervical cancer was performed by the SAM algorithm and resulted in accurate results in most of cases and, hence, reducing the need for manual segmentation to just a few samples.

A statistical study was performed using the spectral signatures from the different groups. The study revealed a significant difference between the spectral signatures of the exocervix, endocervix and cancer tissue, with an increased difference between them in the range over 750 nm and between 470 and 550 nm. This validates, preliminarily, the capability of the system to differentiate alterations in the tissue, which is essential for the detection of precancerous CIN lesions.

Conclusions

This study contributes to the field of in-vivo analysis of cervical lesions using HSI technology. The characterisation of the acquisition system, together with the evaluation of the spectral and spatial characteristics, provides valuable insights regarding the performance and limitations of the system. The results highlight the potential of HSI for cervical lesion analysis but also emphasize the need for further refinement and optimization of the system. With continued research and development, HSI has the potential to improve the detection of precancerous lesions and cervical cancer at its early stages, when the human eye is unable to distinguish small tissue variations caused by cancer. Further works will focus on developing new classification algorithms for the discrimination between cancer cells, neoplastic lesions and normal tissue in the cervix. Furthermore, we will explore the integration of alternative tests, such as the iodine test and multiple acetowhitening images, to improve the annotations. This could improve screening strategies and provide gynaecologists with accurate tissue information for minimally invasive cervical surgery.

Methods

The methodology proposed in this work (Fig. 6) is composed of four main stages: a) HS colposcope hardware and software design, b) HS colposcope characterisation, c) preliminary data collection, d) data processing and analysis.

HS colposcope hardware and software design

The HS colposcope used in this study was developed using a commercially available colposcope, a halogen illumination system, and an HS camera (Fig. 6a) with the aim of being usable into the gynaecologist's current workflow and assessing the feasibility of the system. The Optomic OP-C2 (Optomic, Spain) colposcope was used, where the IR filter was explicitly removed by the manufacturer to allow the HS camera to capture data beyond 750 nm. Moreover, the image splitter also had to be customised to be compatible with the HS camera standard C-Mount. Then, the HS camera (4 in Fig. 6a) was coupled to the splitter (3 in Fig. 6a) by an external bracket designed to support the higher weight of the HS camera with respect to the original RGB camera mounted on the system. The HS camera Snapscan VNIR (IMEC, Belgium) was selected, which covers the 470–900 nm spectral range, capturing 158 spectral bands. This HS camera is based on a spatio-spectral scanning technology called *Snapscan*[40], which is a linescan sensor on a platform that slides inside the camera. The Optomic OP-C2 colposcope includes a light source based on LED technology and a green filter. However, its narrow spectral range of emission did not proportionate good quality HS images. Thus, a halogen light source (OSRAM, 64634 HLX EFR) was included in the system. A custom Graphical User Interface (GUI) was developed to prevent non-expert users from entering low-level configuration parameters and to simplify the HS image acquisition process (Fig. 6a). The GUI is also used to annotate the positions of the biopsies performed on the patient correlated to the HS image.

Spatial and spectral HS colposcope characterisation

The developed system was characterised in terms of spatial resolution, maximum spatial frequency, and spatial distortion. Spatial resolution refers to a camera's capability to faithfully capture intricate details and distinguish between different entities. This attribute specifies the minimum dimension of an object that can be registered[41]. The spatial resolution of the colposcope HS system was assessed through calculation of its field of view (FOV). It can be calculated following Eq. (1), encompassing variables such as pixel size (Ps), pixel count (N), magnification (M) and sensor dimensions (S_s). The procedure involves capturing a certified ruler (Fig. S1a in the Supplementary Material), which is a quality resolution chart with alternating white and black lines, each of 1 mm thick. Then, one profile line is extracted from the captured image, its first derivative is taken and the mean difference between local minima and maxima determines the empirical pixel size. The FOV of the HS colposcope will be the empirical pixel size by the number of pixels on the sensor. Afterwards, to determine the
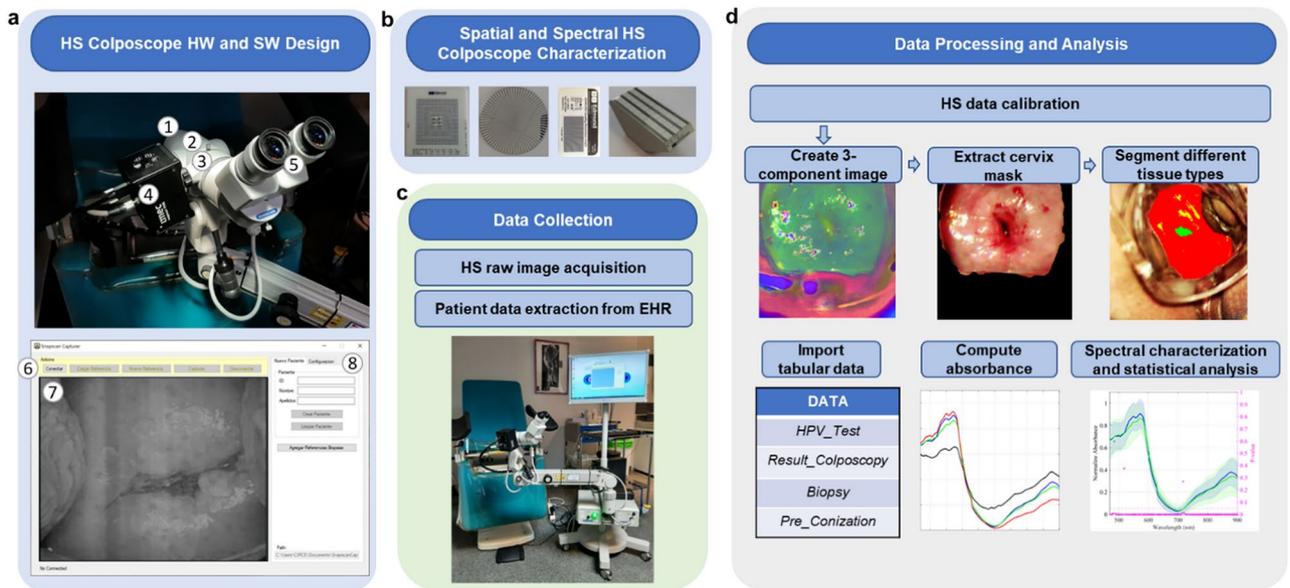
Fig. 6. Methodology followed to develop and study the feasibility of the HS colposcope for gynaecology workflows. (**a**) HS colposcope hardware and software design (1: Front lenses; 2: Main body; 3: Image splitter and bracket; 4: HS camera; 5: Binoculars; 6: Essential system actions of the custom GUI; 7: Central live view for image focus; 8: Textboxes for patient identifiers). (**b**) Spatial and spectral HS colposcope characterisation targets. (**c**) Data collection. (**d**) Data processing and analysis. *HW: Hardware; SW: Software; EHR: Electronic Health Records; HS: Hyperspectral; HPV: Human Papillomavirus.*

current magnification that the HS colposcope obtains over the sample, the empirical FOV is divided by the sensor size.

$$FOV = Ps \cdot N; M = \frac{FOV}{S_s} \tag{1}$$

Subsequently, since the measurement of the pixel size may be limited for characterisation purposes, the maximum frequency reached by the optics was also measured. When objects are imaged too closely together, they can appear merged on the final image. Thus, resolution is better explained using the concept of frequency, measured in line pairs per millimetre ($[Lp/mm]$) which is a pair of black and white squares[33,42]. Lens resolution depends on the ability to distinguish these squares, being the Nyquist frequency the maximum that an optical system can reliably measure[43]. To empirically measure it, a Siemens Star target, characterised by a pattern of radial lines that decrease in thickness towards the centre, was employed (Fig. S1b in the Supplementary Material). The further processing involved calculating the contrast for each one of the radial profiles, which can be defined by the MTF Eq. (2). The objective is to find the maximum spatial frequency that the system can effectively capture while maintaining a certain level of contrast. This step is crucial in assessing the high-frequency performance of the system and its ability to accurately capture fine spatial detail.

$$MTF = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{2}$$

Once resolution and magnification have been defined, it is also crucial to assess whether the final image is affected by any form of distortion due to the system's optical design. Distortion refers to a monochromatic optical abnormality that explains alterations in the magnification of an image across the FOV while maintaining a constant WD[34]. This distortion effect changes at different wavelengths due to their different WDs. It is usually quantified as a percentage of the field height, where small distortions (e.g., in the range of ± 2 to 3%) often go unnoticed by the human eye. Distortion can be either negative (barrel distortion), which causes points in the FOV to appear excessively close to the centre; or positive (pincushion distortion), which causes points to appear too far from the centre. To measure the distortion caused by the HS colposcope, a specific target (Fig. S1c in the Supplementary Material), composed of patterns of dots, was captured. The distortion score is calculated following Eq. (3), where AD refers to the actual distance and PD to the predicted distance between dots.

$$D(\%) = \frac{AD - PD}{PD} \times 100 \tag{3}$$

Spectral characterisation is a critical part of the acquisition system characterisation. The aim of this procedure is to gauge the system's capability to capture and accurately replicate the spectral signatures of known calibration

targets[44]. The SG-3333 spectral calibration target (SphereOptics GmbH, Germany) was used as a reference (Fig. S1d in the Supplementary Material). The Zenith Polymer is composed of rare earth oxides (Holmium, Erbium, Dysprosium). The captured spectral signature was qualitatively compared against the certified data from NIST (National Institute of Standards and Technology), which covers the 250–2450 nm range nm in steps of 0.1 nm. Moreover, the SCM was calculated to quantitatively validate the spectral quality of the HS colposcope system[45]. The SCM is calculated following Eq. (4), where s_i and s_j are two spectral signatures and n the number of spectral bands. The correlation may exhibit either a positive or negative relationship, thus accounting to some extent for variations in brightness and shape among spectra.

$$SCM = \frac{n \sum_1^n s_i s_j - \sum_1^n s_i \sum_1^n s_j}{\sqrt{\left(n \sum_1^n s_i^2 - \left(\sum_1^n s_i\right)^2\right)\left(n \sum_1^n s_j^2 - \left(\sum_1^n s_j\right)^2\right)}} \tag{4}$$

The assessment of the DOF is another crucial aspect of the characterisation process, as it allows for the determination of the area in which an object appears sharp in the captured images[46]. This measure has profound implications for the practicality and usability of the HS colposcope, especially in clinical settings where accurate depth perception is required to differentiate between different grades of CIN[6]. In this study, the DOF was determined using the DOF 5–15 target from Edmund Optics (New Jersey, USA), which is composed by two sets of scales. The scales consist of horizontal and vertical lines with a frequency of $5lp/mm$ and $15lp/mm$, captured at 45º (Fig. S1e in the Supplementary Material). The MTF was calculated from a profile of the captured images of this target. The MTF50[47] determined the range in which lines appear clear and sharp, considered as the DOF of the imaging system. The calibration target establishes a standard for focus tolerance that is typically characterised by theoretical methods.

The evaluation of the acquisition time of the imaging system is crucial, as the longer the acquisition time, the greater the possibilities of image blurring and discomfort for the patient and gynaecologist. The integration time of each frame of the sensor was established to have a fixed maximum digital value of 950 (above the 1023 maximum by the 10-bits resolution) when capturing the white reference, with a configuration suitable for our system (integration time delay 5 and gain 1.7 dB). The acquisition time was measured from the time the capture button was pressed on the software until the system stopped saving data on disk, which is a 7200 rpm mechanical hard drive.

Data collection

The HS cervix database was collected during gynaecology screening (Fig. 6c). The medical doctor conducted the regular consultation, collected the patient's demographic and clinical information and captured the cervix images using the HS colposcope. The capturing processes were performed in accordance with the current guidelines and regulations, which was approved by the ethical committee of the *Complejo Hospitalario Universitario Insular Materno Infantil* with reference 2022–081-1. All the participants involved in this study and/or their legal guardians were informed about the research and voluntarily signed an informed consent form allowing their participation and the anonymous publication of the results. In addition, all research methodologies were performed in accordance with the current guidelines and regulations.

Demographic and clinical information for each patient and capture into the database was collected from the EHR (Electronic Health Records) and stored in a separate tabular file. These annotations contained valuable information such as patient ID, date, cytology results, HPV test, colposcopy evaluation and other clinical information. This tabular database was designed to be linked to the main imaging database through a randomized identifier number unique to each capture. To analyse the spectral information, it is important to ensure whether each cervix is normal or not. For this test, the gold standard for normal cervix is a negative HPV test and a normal colposcopic assessment. Each of the 124 HS images that make up the current database was processed to extract the spectral signatures of the different tissue types (Fig. 6d).

HS data calibration

The captured data was calibrated by using flat field calibration, which compensates the response based on a white reference and a dark reference sample, obtaining a value of relative reflectance for each pixel and band. The acquisition and calibration process are explained in detail in the Methods section of the Supplementary Material.

PCA-based cervix segmentation approach

A custom approach based on data transformation through the PCA algorithm (Fig. 7) was proposed for the reduction of the calibrated HS image from 158 bands to a 3-channel image. This image was the input for the proposed cervix area segmentation approach. These are generated employing the three first components of the PCA domain, each component being independently regularised to a range between 0 and 1 based on an adapted Z-score function. Each channel is quantified to 8 bits to generate a 24-bit depth image based on RGB components. In the Methods section of the Supplementary Material, the 3-channel image generation is described in detail.

Taking a look into the 3-channel images generated by the PCA, the cervix area keeps a constant synthetic colour between different patients. This characteristic was used to segment the cervix from the rest of the images. At first, the standard cervical corrected PCA value was captured from the cervical area of 5 manually chosen images. Then, for each capture, it was computed the SAM map against the mean corrected PCA value that was initially generated. The SAM method (Eq. (5)) determines the spectral similarity between a test spectral signature (t) and a reference spectral signature (r), where C is the spectral band, by calculating the angle ($\propto$). The smaller the angle, the greater the similarity of the spectral signatures. For this application, the pixels with a
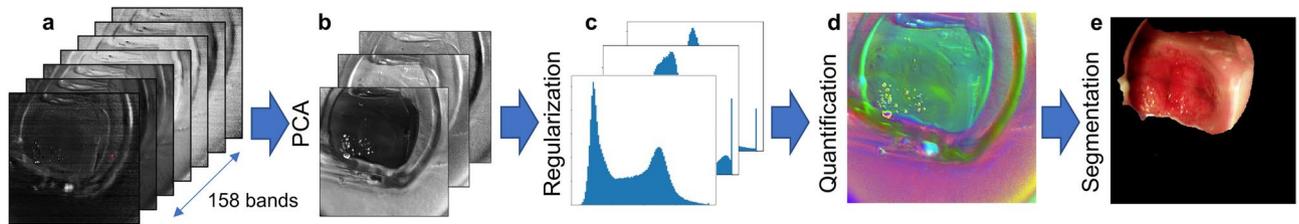
Fig. 7. Generation process of the 3-channel image. (**a**) Calibrated HS image, (**b**) 3 most relevant PCA components, (**c**) channels in 1 to 0 range, (**d**) 3-channel image, and (**e**) cervix area segmentation.

SAM angle lower than 0.2 rad are selected as cervix area, where the threshold was empirically chosen as a trade-off between accuracy of delineation and robustness in cervixes affected with high-grade lesions. The selected cervix pixels are then converted into a logical mask, which is affined by removing the clusters with an area smaller than an equivalent square of 17×17 pixels, to avoid wrong detections. Then the area of the clusters is increased by 10 pixels around to get some possible missing pixels. Finally, the largest cluster of pixels is selected, and its entire area is closed to avoid gaps in the middle of the cervix.

$$\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{C} t_i r_i}{\sqrt{\sum_{i=1}^{C} t_i^2}\sqrt{\sum_{i=1}^{C} r_i^2}}\right) \tag{5}$$

SAM-based approach for cervical tissue discrimination

The segmentation process categorises the captured cervical pixels into four different classes: *exocervix*, *endocervix*, *outliers*, and *cervical cancer*, each embodying unique regions and attributes of cervical tissue. The following threshold were empirically selected to evaluate the interpatient variability. A pixel is considered an outlier if the SAM between its spectra and the mean cervix spectral signature is higher than 0.25 rad. These may include other tissues, imaging artifacts, or noise. Segmentation of outliers is important as it allows for a cleaner and more accurate analysis of the cervix tissue. For the segmentation of the endocervix, it is important to understand that it is a different tissue, with a variation in the spectral signature and positioned in the centre of the cervix. The segmentation is performed by multiplying the previously computed SAM map ($\alpha_{x,y}$) by a correction map described in Eq. (6). This correction map calculates the distance between the centre of mass of the cervix area in the image ($X_{CervixCentre}, Y_{CervixCentre}$) and the distance to the current pixel evaluated (x, y) while it is escalated over the larger distance possible from each coordinate of the mask (X_n, Y_n) to its cervix centre. The coefficient resulting from this calculation has a value of 1 for the position of the centre of mass of the cervix mask. As the distance to the centre increases, the coefficient decreases. Pixels with a corrected SAM angle between 0.10 and 0.18 rad are then selected as the endocervix.

$$\alpha_{\text{corrected}_{x, y}} = \alpha_{x,y} * \frac{1 - \sqrt{(x - X_{CervixCentre})^2 + (y - Y_{CervixCentre})^2}}{\frac{1}{3}\max_n \sqrt{(X_n - X_{CervixCentre})^2 + (Y_n - Y_{CervixCentre})^2}} \tag{6}$$

The pixels of the exocervix are selected from those that remain from the other groups. The database only contains 6 patients with invasive cancer, which is insufficient to justify the effort of automating tissue segmentation. For these patients, the other tissue classes are not segmented as most of the cervix area is affected by the invasive cervical cancerous tissue. These patients were manually annotated by an expert gynaecologist based in the whitening effect of the acetic acid. Figure S3 of the Supplementary Material shows the acetowhite images, which were the reference for the gynaecologist, and the generated annotations. In the annotations it is noticeable that only some areas of the cancerous tissue were marked, avoiding regions with blood and specular reflections. The evaluation of the exocervix, endocervix, and outlier pixels is limited to patients who meet the gold standard criteria for being healthy. This is to prevent any contamination from pixels that may be affected by precancerous lesions.

Statistical analysis

The objective of this statistical analysis was to quantitatively evaluate and compare the spectral signatures obtained from different types of cervical tissues: the exocervix, the endocervix and cervical cancer. The statistical analysis was performed on the reflective coefficients (R), calibrated data captured from the sensor without applying normalization, and absorption coefficients (A), computed following Eq. (7) derived from Beer-Lambert Law[48]. Additionally, the statistical analysis was performed using Lilliefors test, Wilcoxon Rank-Sum, and Student's T-test, which are detailed in the Methods section of the Supplementary Material.

$$A = -\log_{10}(R) \tag{7}$$

Data availability

The datasets generated during the characterization for the current study are available from the corresponding author, under reasonable request, through https://hsidatabase.iuma.ulpgc.es/.

References

1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
3. World Health Organization. *Cervical Cancer Spain 2021 Country Profile.* (2021).
4. Herrera, A. *et al. Estimaciones de La Incidencia de Cáncer En Canarias 2018.* (2018).
5. Walboomers, J. M. et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* **189**, 12–19 (1999).
6. A G Ostör. Natural history of cervical intraepithelial neoplasia: A critical review. In *International Journal of Gynecological Pathology* (1993).
7. Hillemanns, P., Soergel, P., Hertel, H. & Jentschke, M. Epidemiology and early detection of cervical cancer. *Oncol. Res. Treat.* **39**, 501–506 (2016).
8. Prakash, M., Patterson, S. & Kapembwa, M. S. Evaluation of the cervical cytobrush sampling technique for the preparation of CD45+ mononuclear cells from the human cervix. *J. Immunol. Methods* **258**, 37–46 (2001).
9. Chase, D. M., Kalouyan, M. & DiSaia, P. J. Colposcopy to evaluate abnormal cervical cytology in 2008. *Am. J. Obstet. Gynecol.* **200**, 472–480 (2009).
10. Wentzensen, N. et al. A prospective study of risk-based colposcopy demonstrates improved detection of cervical precancers. *Am. J. Obstet. Gynecol.* **218**(604), e1-604.e8 (2018).
11. Lycke, K. D. et al. Agreement on lesion presence and location at colposcopy. *J. Low Genit. Tract Dis.* **28**, 37–42 (2024).
12. Mustafa, R. A. et al. Systematic reviews and meta-analyses of the accuracy of HPV tests, visual inspection with acetic acid, cytology, and colposcopy. *Int. J. Gynecol. Obstet.* **132**, 259–265 (2016).
13. Brown, B. H. & Tidy, J. A. The diagnostic accuracy of colposcopy—A review of research methodology and impact on the outcomes of quality assurance. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **240**, 182–186 (2019).
14. Guido, R., Schiffman, M., Solomon, D. & Burke, L. Postcolposcopy management strategies for women referred with low-grade squamous intraepithelial lesions or human papillomavirus DNA-positive atypical squamous cells of undetermined significance: A two-year prospective study. *Am. J. Obstet Gynecol.* **188**, 1401–1405 (2003).
15. Lange, H. & Ferris, D. G. Computer-aided-diagnosis (CAD) for colposcopy. In (eds. Fitzpatrick, J. M. & Reinhardt, J. M.) 71 (2005). https://doi.org/10.1117/12.594658.
16. Acosta-Mesa, H. G. *et al.* Cervical cancer detection using colposcopic images: a temporal approach. In *Sixth Mexican International Conference on Computer Science (ENC'05)* 158–164 (IEEE). https://doi.org/10.1109/ENC.2005.14.
17. Hou, X. et al. Artificial intelligence in cervical cancer screening and diagnosis. *Front. Oncol* https://doi.org/10.3389/fonc.2022.851367 (2022).
18. Peng, G., Dong, H., Liang, T., Li, L. & Liu, J. Diagnosis of cervical precancerous lesions based on multimodal feature changes. *Comput. Biol. Med.* **130**, 104209 (2021).
19. Yu, Y., Ma, J., Zhao, W., Li, Z. & Ding, S. MSCI: A multistate dataset for colposcopy image classification of cervical cancer screening. *Int. J. Med. Inform.* **146**, 104352 (2021).
20. Pesapane, F., Volonté, C., Codari, M. & Sardanelli, F. Artificial intelligence as a medical device in radiology: Ethical and regulatory issues in Europe and the United States. *Insights Imaging* **9**, 745–753 (2018).
21. Petersen, E. et al. Responsible and regulatory conform machine learning for medicine: A survey of challenges and solutions. *IEEE Access* **10**, 58375–58418 (2022).
22. European Parliament. *Regulation (EU) 2024/1689 of the European parliament and of the council.* http://data.europa.eu/eli/reg/2024/1689/oj (2024).
23. Halicek, M., Fabelo, H., Ortega, S., Callico, G. M. & Fei, B. In-vivo and ex-vivo tissue analysis through hyperspectral imaging techniques: Revealing the invisible features of cancer. *Cancers (Basel)* **11**, 756 (2019).
24. Wang, P., Wang, S., Zhang, Y. & Duan, X. multispectral image under tissue classification algorithm in screening of cervical cancer. *J. Healthc. Eng.* **2022**, 1–9 (2022).
25. Karthika, J., Anantharaju, A., Koodi, D., Pandya, H. J. & Pal, U. M. Label-free assessment of the transformation zone using multispectral diffuse optical imaging toward early detection of cervical cancer. *J. Biophotonics* https://doi.org/10.1002/jbio.202400114 (2024).
26. Kamruzzaman, M. & Sun, D. W. Introduction to hyperspectral imaging technology. In *Computer Vision Technology for Food Quality Evaluation: Second Edition* 111–139 (2016). https://doi.org/10.1016/B978-0-12-802232-0.00005-0
27. Wang, A. M. J., Nammalvar, V. & Drezek, R. A. Targeting spectral signatures of progressively dysplastic stratified epithelia using angularly variable fiber geometry in reflectance Monte Carlo simulations. *J. Biomed. Opt.* **12**, 044012 (2007).
28. Mirabal, Y. N. et al. Reflectance spectroscopy for in vivo detection of cervical precancer. *J. Biomed. Opt.* **7**, 587 (2002).
29. Wang, C. et al. Multi-scale hyperspectral imaging of cervical neoplasia. *Arch. Gynecol. Obstet.* **293**, 1309–1317 (2016).
30. Vega, C. *et al.* development of a hyperspectral colposcope for early detection and assessment of cervical dysplasia. In *2022 25th Euromicro Conference on Digital System Design (DSD)* 863–870 (IEEE, 2022). https://doi.org/10.1109/DSD57027.2022.00121.
31. Ortega, S. et al. Hyperspectral push-broom microscope development and characterization. *IEEE Access* **7**, 122473–122491 (2019).
32. Birch, G. C. & Griffin, J. C. Sinusoidal Siemens star spatial frequency response measurement errors due to misidentified target centers. *Opt. Eng.* **54**, 074104 (2015).
33. Loebich, C., Wueller, D., Klingen, B. & Jaeger, A. Digital camera resolution measurement using sinusoidal Siemens stars. in (eds. Martin, R. A., DiCarlo, J. M. & Sampat, N.) 65020N (2007). https://doi.org/10.1117/12.703817.
34. Chandler, D. M. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Electron. Imaging* **19**, 011006 (2010).
35. Lycke, K. D. et al. Agreement on lesion presence and location at colposcopy. *J. Low Genit. Tract Dis.* **28**, 37–42 (2024).
36. Oliveira, L. M. C. & Tuchin, V. V. *The Optical Clearing Method* (Springer, 2019). https://doi.org/10.1007/978-3-030-33055-2.
37. Cohen, P. A., Jhingran, A., Oaknin, A. & Denny, L. Cervical cancer. *The Lancet* **393**, 169–182 (2019).
38. Waller, J. et al. The clinical utility of imaging methods used to measure hypoxia in cervical cancer. *Br. J. Radiol.* https://doi.org/10.1259/bjr.20190640 (2020).
39. Kim, C. K., Park, S. Y., Park, B. K., Park, W. & Huh, S. J. Blood oxygenation level-dependent MR imaging as a predictor of therapeutic response to concurrent chemoradiotherapy in cervical cancer: A preliminary experience. *Eur. Radiol.* **24**, 1514–1520 (2014).

40. SNAPSCAN VNIR hyperspectral camera | imec. https://www.imechyperspectral.com/en/cameras/snapscan-vnir.
41. Lasch, P. & Naumann, D. Spatial resolution in infrared microspectroscopic imaging of tissues. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1758**, 814–829 (2006).
42. Artmann, U. Quantify Aliasing a new approach to make resolution measurement more robust. *Electron. Imaging* **31**, 320-1–320-6 (2019).
43. Artmann, U. Linearization and normalization in spatial frequency response measurement. *Electron. Imaging* **28**, 1–6 (2016).
44. Pekkala, O., Pulli, T., Kokka, A. & Ikonen, E. Setup for characterising the spectral responsivity of Fabry–Pérot-interferometer-based hyperspectral cameras. *Metrologia* **56**, 065005 (2019).
45. van der Meer, F. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *Int. J. Appl. Earth Obser. Geoinf.* **8**, 3–17 (2006).
46. Kosara, R., Miksch, S. & Hauser, H. *Semantic Depth of Field* 97–104 (2001).
47. Nasse, H. H. How to Read MTF Curves. (2008).
48. Oshina, I. & Spigulis, J. Beer-Lambert law for optical tissue diagnostics: Current state of the art and the main limitations. *J. Biomed. Opt.* **2**, 6. https://doi.org/10.1117/1.JBO.26.10.100901 (2021).

Acknowledgements

Author contributions

C.V., L.Q-Q., R.L., and H.F. conceptualized the study. C.V. developed the acquisition system, the software, and validated it. C.V., N.M., L.Q-Q., and R.L. performed the data curation. C.V., R.L., and L.Q-Q. wrote the manuscript. J.R., A.M., H.F. and G.M.C. supervised the entire study, provided resources, and acquired funding. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-84422-z.

Correspondence and requests for materials should be addressed to C.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.