ULPGC
Universidad de
Las Palmas de
Gran Canaria

IUMA
Instituto Universitario de
Microelectrónica Aplicada

CONTRIBUTIONS TO THE STANDARDIZATION AND OPTIMIZATION OF HYPERSPECTRAL MICROSCOPIC SYSTEMS FOR MEDICAL DIAGNOSIS

Doctorado en Tecnologías de
Telecomunicación e Ingeniería Computacional

Laura Quintana Quintana

Las Palmas de Gran Canaria, Diciembre 2025

ULPGC
Universidad de
Las Palmas de
Gran Canaria

IUMA
Instituto Universitario de
Microelectrónica Aplicada

Ph.D. Thesis

Contributions to the Standardization and Optimization of Hyperspectral Microscopic Systems for Medical Diagnosis

by

Laura Quintana Quintana

Acknowledgements

I would like to take a moment to acknowledge all the people who, in one way or another, have influenced my personal and professional growth throughout the course of this PhD thesis. These five years have been filled with learning, challenges, and discoveries, not only in science, but also about myself, and I know I would not have made it this far without the support, patience, inspiration, and kindness of many people along the way. Every conversation, piece of advice, shared coffee, or simple gesture of encouragement has left its mark on this journey.

I want to thank my family, who have been my foundation throughout these years. Gracias, Mamá, por enseñarme a echarle cara a la vida y por los incontables tuppers de potaje que me han mantenido viva durante toda la tesis. Thank you, Dad, for patiently explaining the mechanisms behind so many things, your curiosity and wisdom have kept my mind sharp and always ready to learn. To my brothers, Óscar and Carlos, my constant companions who have supported me every step of the way, always interested in my progress and encouraging my efforts. Special thanks to Óscar, who listened to my pleas and designed the cover, along with several other graphics, of this PhD thesis. I am also super grateful to my grandparents, Paca, Pipo, Sarito y Juan, whose love, stories, and strength will always accompany me. Abuela Paca, te quedó muy poco para verme con el gorrito puesto, pero sé que me estarás viendo allá donde vaya, esta tesis va en parte por ti.

To my lifelong friends: Berto, Gope, Iván, Pablo, Patt, and Paula, thank you for standing by me through every high and low. If I ever write another book, it will be filled with all our countless shared memories and anecdotes. A special thanks goes to Pablo and Berto, who, after living together, endured years of frustration over Reviewer 2's comments. Your patience and support mean the world to me. To Mireia, your unwavering support since the day we met has been invaluable; thank goodness video calls aren't charged by the minute, or we'd have gone broke by now. I'm also deeply grateful to my Ru7a friends. Since meeting you last year, I've had the privilege of connecting with incredible people, curious minds and passionate

hearts, that have kept me inspired, entertained, and mentally grounded throughout the final stretch of this thesis.

I would like to express my genuine thanks to my PhD supervisors, Dr. Gustavo Marrero and Dr. Samuel Ortega. To Gustavo, thank you for believing in me back in the summer of 2019, when I showed up at your office after reading a newspaper article about something called "hyperspectral imaging". I'm grateful for your decided support of the research group over the years, and for connecting us with a broad network of researchers and institutions, which has opened the door to numerous collaborations and opportunities for growth. I also deeply appreciate your efforts to bring disciplines like Biomedical Engineering to our university. There is so much talent on this island, and thanks to people like you, we now have the chance to display it. To Samuel, I cannot thank you enough for finding time, even when there was none, to guide me through these years. Your consistent presence and attention to detail at every stage have helped me develop a more critical and thoughtful approach to my work. Being under your supervision has been a true privilege, and the knowledge, perspective, and encouragement you have shared have had a lasting impact, not only on my academic progress, but also on my growth as a professional and as a person. I also want to extend my sincere thanks to Dr. Himar Fabelo for the continuous support and mentorship he has provided throughout my doctoral journey. Though not formally appointed as my supervisor, he played a key role in my day-to-day work, offering guidance at every stage and taking a genuine interest in my progress. Beyond his technical input, his openness in sharing the realities of academic life has been particularly impactful, helping me understand the inner workings of the research world and shaping my development as a researcher.

Throughout the course of my PhD, I had the privilege of traveling in the name of science, and at every stop, I was welcomed by generous and inspiring individuals who made each experience truly enriching. My journey began in Tortosa, Spain, where I had the opportunity to collaborate with Dr. Carlos López, Esther Sauras, and the entire team of the Oncological Pathology and Bioinformatics Research Group at the Institut d'Investigació Sanitària Pere Virgili. Thank you for embracing me as one of your own and for allowing me to witness firsthand the reality of working in a hospital environment.

I am especially grateful to Esther for her exceptional professionalism and dedication throughout the various projects we shared. From there, I made a brief visit to Madrid, where I was warmly received by the team at the Centro de Investigación en Tecnologías Software y Sistemas Multimedia (CITSEM) at the Universidad Politécnica de Madrid. My sincere thanks to Dr. Eduardo Juárez and the entire group for making me feel at home. I hope I have managed to give acceptable answers to your philosophical coffee-break questions. I truly value the long-standing collaboration between our teams, which continues to foster mutual growth and learning. The next stop brought me to Amsterdam, where I carried out an international research stay. I am grateful to Dr. Behdad Dasht Bozorg for welcoming me into the Image Guided Surgery Research Group at the Netherlands Cancer Institute and for supporting my proposal. Thank you also to Dr. Dirk Sterenborg for the enlightening sessions on light–tissue interactions and their relevance to my thesis. A special mention goes to Mark Witteveen for his curiosity, enthusiasm, and for patiently introducing me to the world of Monte Carlo simulations, we have built a collaboration I truly value. Finally, I would like to thank Dr. Qianqian Fang for organizing and teaching the 5th International MCX Training Workshop at Northeastern University in Boston, USA. Together with Prof. Steve Jaques, you deepened my understanding of light–tissue interactions and their simulation, giving me new tools to apply in my research. And of course, thank you once again, Mireia, for opening the doors of your beautiful home during that stay, it made everything even more special.

I also want to thank all the researchers who contributed in one way or another to this thesis. Dr. María de la Luz Plaza and Rafael Camacho from the Department of Pathological Anatomy, Hospital Universitario de Gran Canaria Doctor Negrín, for their selfless collaboration by providing tissue samples and valuable feedback from the pathologist's perspective. Alfonso Lagares and Carlos Avedaño from SERMAS (Servicio Madrileño de Salud), for the histology crash course and your unwavering support. Furthermore, I would like to thank Dr. Bernardino Clavo from the Research Unit of the University Hospital of Gran Canaria Doctor Negrín, for his continuous willingness to collaborate. Dr. Alejandro Cruz-Guerrero, Dr. Liliana Granados, Juan Nicolás Mendoza-Chavarría, and Prof.

Daniel Ulises Campos-Delgado from the Universidad Autónoma de San Luis Potosí, thank you for traveling to Gran Canaria and sharing your knowledge on spectral unmixing, our collaboration has been invaluable. Also, thanks to Prof. Francesco Leporati and Dr. Emanuele Torti from the University of Pavia for the conferences, conversations, and shared projects.

I am grateful to the IUMA (Research Institute of Applied Microelectronics) at ULPGC for providing a home to develop my research. Thanks to Dr. Eduardo Quevedo and Dr. Francisco Balea for the time shared and the valuable advice throughout this journey. Special thanks to the support staff, Agustín Quintana and Enrique Montesdeoca, for ensuring I had the resources I needed to conduct my experiments. I would also like to express my sincere gratitude to Judith Sánchez, Director of University Outreach at ULPGC, for her unwavering support throughout the organization of Pint of Science Las Palmas de Gran Canaria. This science communication initiative greatly enriched my PhD experience, making it both more enjoyable and meaningful.

Finally, I want to acknowledge my fellow researchers and companions on this journey: Antonio Rodríguez, Abián Hernández, Beatriz Martínez, Carlos Vega, Guillermo Socorro, Javier Santana, María Castro, and Dr. Raquel León. If anyone ever asks me for advice about starting a PhD, my answer is always the same: *"PhDs are not solo missions"*. These incredible people made the road smoother, lighter, and far more enjoyable. Thank you for enduring the maze of Spanish bureaucracy with me, and for the countless coffee breaks that not only helped us clear our minds but often sparked fresh ideas. To Antonio, thank you for your tireless advocacy for workers' rights and for teaching me to stand up for them as well. To Carlos, thank you for the meaningful collaborations, and I truly admire your talent and dedication as a researcher. To my "pupil", Javier, thank you for taking up the torch and always challenging me with thoughtful questions, keeping me moving forward. And to Raquel, a special mention for being my constant supporter, even in the craziest ideas and adventures. Thank you for being my colleague in all the departments we invented together (marketing, communications, IT, events, scientific dissemination, and more) because with your

support, I'd start any project without hesitation. I have learned, and continue to learn, more from you than I could ever express.

To everyone I've mentioned here, and to those who have supported me in ways I may have unintentionally overlooked (*I'm only a human* ♫), thank you, thank you, and thank you.

Abstract

Hyperspectral (HS) microscopy has emerged as a promising modality in biomedical imaging, enabling the capture of high-resolution spatial and spectral information from biological tissues. Its ability to provide non-invasive, label-free, and quantitative analysis of microscopic samples opens new possibilities in digital pathology and computer-aided diagnosis. HS microscopy retains critical biochemical and morphological features in tissue specimens, allows for retrospective reanalysis, and generates reproducible high-dimensional data. These advantages position it as a strong complement to traditional histopathological techniques. However, despite its potential, the adoption of HS microscopy in routine clinical workflows remains limited. Current barriers include significant variability in instrumentation, imaging protocols, and data processing techniques, which challenge reproducibility, scalability, and clinical integration.

Overcoming these limitations requires a holistic and technically rigorous approach, starting with the development and characterization of the imaging instrumentation itself. Reliable and consistent data acquisition depends on thorough evaluation of the sensor's performance, including spectral accuracy, spatial resolution, illumination uniformity, and noise characteristics. Establishing these parameters ensures fidelity and comparability across imaging sessions and platforms, setting the stage for robust downstream analyses. Another critical consideration is the effect of optical focus on image quality. HS data is particularly sensitive to even slight defocusing, which can distort spectral signatures and compromise interpretability. Strategies such as focus quality assessment and optimization routines—incorporating sharpness metrics, autofocus evaluation, or focus stacking—play a vital role in maximizing both spatial and spectral integrity across entire tissue sections.

Equally important is the standardization of sample preparation procedures. Variability in histological processing can introduce inconsistencies that obscure genuine spectral patterns. In particular, tissue thickness has a notable impact on spectral quality and

contrast. The adoption of well-defined sectioning protocols, informed by empirical evidence, is essential for minimizing inter-sample variability and enhancing reproducibility, especially in studies involving multiple centers or longitudinal datasets. Finally, once data are acquired, specific preprocessing steps must be applied to prepare the hyperspectral information for meaningful analysis. These include background subtraction, spectral normalization, denoising, and illumination correction, operations tailored to the unique properties of HS microscopy data. Properly implemented, these workflows mitigate acquisition artifacts and strengthen the reliability of extracted spectral signatures.

Complementing these technical advancements, this Ph.D. thesis presents the development of curated, publicly available HS microscopic databases designed to capture real-world variability in sample preparation and imaging conditions. These benchmark datasets provide a valuable foundation for reproducible algorithm development and foster collaboration between computational researchers and clinical practitioners, addressing a critical bottleneck in the field and accelerating the translation of HS microscopy into clinical practice.

Together, these efforts contribute to the maturation of HS microscopy as a reproducible, scalable, and clinically relevant imaging modality. By emphasizing standardization, reproducibility, and data quality throughout the imaging pipeline, this Ph.D. thesis lays the groundwork for future diagnostic assistance tools based on HS microscopic technology. These developments are expected to facilitate the integration of HS microscopy into both clinical research and practice, supporting more objective and precise diagnostic decision-making in biomedicine.

Contents

Chapter 1. Introduction

1.1 Background

1.1.1 Cancer Diagnosis

Cancer remains one of the most pressing global health challenges worldwide. In 2022, nearly 20 million people were newly diagnosed with cancer and about 9.7 million died from the disease, translating to a lifetime risk of roughly one in five individuals developing cancer and one in nine men or one in twelve women dying from it [1] . Lung cancer is now the most frequently diagnosed malignancy (2.5 million new cases, 12.4 % of the total) and the leading cause of cancer-related death, responsible for 1.8 million deaths (18.7 %), with female breast and colorectal cancers following closely in both incidence and mortality rankings [2]. Among the less frequent but particularly devastating forms is brain and central nervous system (CNS) cancer, which, despite accounting for only about 1.4% of global cancer cases and 2.4% of cancer deaths in 2022, carries disproportionately high morbidity and mortality due to its location and often aggressive nature [3]. Encouragingly, despite population growth and ageing driving up absolute numbers, the global age-standardized cancer mortality rate fell by almost 6 % between 2010 and 2019, and 131 countries recorded declines, underscoring the impact of tobacco-control measures, vaccination, earlier detection, and advances in systemic therapies [4].

The clinical workflow for cancer detection heavily relies on imaging modalities such as ultrasound, computed tomography, x-rays or magnetic resonance imaging [5], which play a crucial role in identifying suspicious lesions. However, definitive diagnosis and treatment planning depend on histopathological examination of tissue obtained through biopsy. While biopsies are a standard component of cancer evaluation, their interpretation can be complex due to the heterogeneity of cancer types. Accurate identification and classification of these types demand specialized expertise [6]. Traditionally, histopathological diagnosis has been carried out using bright-field microscopy of tissue sections stained with hematoxylin and eosin (H&E), a method that has long served as the gold standard for examining tissue architecture [7]. The preparation of these samples is designed to enhance the visualization of cellular and structural details. This intricate, multi-step process involves contributions from various professionals across multiple hospital departments, each playing a crucial role in ensuring the accuracy and quality of the final histological evaluation.

It begins with sample collection (Figure 1a), which takes place in the Operating Room where tissue biopsies are obtained from the patient either during surgery or postmortem [8]. Immediately afterward, the fixation step (Figure 1b) is carried out by immersing the tissue in formalin to prevent decomposition and preserve cellular structures. The samples are then transported to the Pathology Department for reception and registration (Figure 1c), where each is assigned a unique identification number and logged into the laboratory tracking system. Next, a macroscopic examination and sectioning (Figure 1d) is performed by a pathologist, who visually inspects the tissue to assess size, shape, and any visible abnormalities, before cutting representative sections and placing them in labeled cassettes. These cassettes proceed to the embedding stage (Figure 1e), where technicians dehydrate the tissue with alcohol, clear it with xylene, and infiltrate it with molten paraffin wax. During block preparation (Figure 1f), the paraffin-infiltrated samples are poured into molds and correctly positioned before cooling and solidifying into blocks ready for sectioning. In the microtome sectioning step (Figure 1g), thin slices of tissue, typically 2–10 μm thick, are cut and transferred to glass slides. Finally, in the staining and mounting process (Figure 1h), technicians apply dyes to

highlight cellular structures (e.g., H&E). This involves wax removal, rehydration, staining, and dehydration, followed by mounting the sample with a coverslip using synthetic resin to preserve it for microscopic analysis [6], [9], [10].

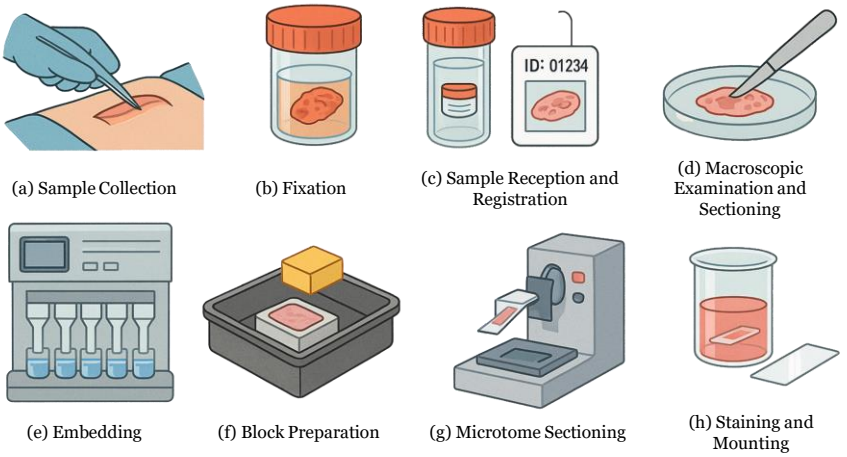


Figure 1. Histology sample preparation: (a) sample collection, (b) fixation, (c) sample reception and registration, (d) macroscopic examination and sectioning, (e) embedding, (f) block preparation, (g) microtome sectioning and (f) staining and mounting.

Once the sample is prepared, it enables pathologists to examine, under observation in the microscope, cellular and subcellular features critical for accurate disease diagnosis [9]. However, this technique has several limitations. The use of exogenous stains can obscure subtle morphological details, and variations in staining protocols, reagent quality, and technician practices contribute to inconsistencies in image quality and interpretation. Furthermore, visual assessment by expert pathologists remains inherently subjective, potentially leading to inter-observer variability, particularly in diagnostically challenging cases [11]. The digitization of pathology slides through whole-slide imaging (WSI) has introduced new possibilities for improving diagnostic accuracy and workflow efficiency. WSIs are high-resolution digital representations of traditional glass slides, enabling easier storage, sharing, and remote consultation [12], [13]. This digital transformation has also laid the groundwork for computational pathology, a field that combines traditional image analysis techniques with modern machine learning (ML) and deep learning

(DL) methods [14], [15]. These computational approaches not only facilitate automated and reproducible assessments of histological features but also enable the integration of multimodal data, such as histopathological images, clinical records, and genomic information, offering a more comprehensive understanding of disease processes and aiding in prognostic and therapeutic decision-making [16], [17], [18]. However, despite these advancements, bright-field microscopy remains limited to the visible spectrum and is thus insensitive to biochemical features outside this range. To address this limitation, alternative imaging techniques are being explored to capture richer, label-free information from biological tissues. These emerging modalities aim to enhance objectivity, reproducibility, and diagnostic precision in histopathological analysis.

1.1.2 Hyperspectral Imaging

Our perception of the world is intrinsically linked to the way we detect and interpret light. Traditionally, human vision, and by extension, most conventional optical imaging systems, has been restricted to the visible portion of the electromagnetic spectrum, approximately 400 to 750 nanometers (nm). While this spectral window allows the observation of color and structural features, it constitutes only a narrow segment of the electromagnetic spectrum. As illustrated in Figure 2, the full spectrum ranges from high-energy gamma rays (<0.01 nm) and X-rays (0.01–10 nm), through ultraviolet (UV: ~10–400 nm), visible light (VIS: ~400–750 nm), and infrared (IR: ~750 nm–25 μm), to microwaves and radio waves (>1 mm).

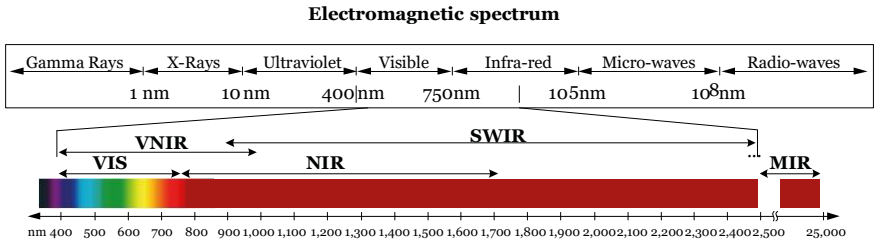


Figure 2. Overview of the electromagnetic spectrum, from gamma rays to radio waves, indicating the spectral regions relevant to visible and NIR imaging techniques.

Each spectral region provides distinct physical and chemical insights: gamma rays and X-rays are extensively employed in diagnostic imaging and material inspection due to their deep tissue penetration, whereas UV light enables the detection of surface-level fluorescence and photochemical activity. However, high-energy radiation can be ionizing and harmful to biological tissues, while lower-energy long-wavelength regions may lack sufficient energy for fine-detail imaging.

A significant advancement within the visible and near-infrared region has been the development of multispectral (MS) and hyperspectral (HS) imaging (MSI and HSI) techniques [19], [20]. These modalities acquire radiance across tens to hundreds of narrows, contiguous spectral bands, extending well beyond the conventional RGB channels and human vision capabilities. MSI and HSI commonly cover the visible (VIS: ~400 – 750 nm) and near-infrared (NIR: ~750–1400 nm), short-wave infrared (SWIR: ~1400–2500 nm), and mid-infrared (MIR: >2500 nm) regions (Figure 2). By capturing detailed spectral information for each pixel, these techniques enable precise material identification and classification based on characteristic patterns of light absorption and reflection. This technique has already made significant contributions to areas such as remote sensing [21], agriculture [22], and geology [23], and is now gaining increasing relevance in biomedical applications [24], [25], particularly in the field of pathology [26].

MSI/HSI generates a three-dimensional data structure known as a spectral cube, where two dimensions capture the spatial information of the scene for a single spectral band. The third dimension contains spectral information for each pixel, forming what is known as a *spectral signature*, a unique radiance pattern that describes how a specific material reflects or absorbs light across different wavelengths. Unlike standard imaging techniques that are limited to three broad color channels, MSI/HSI can capture hundreds of contiguous narrow spectral bands, enabling the identification of materials based on their unique spectral signatures. The key difference between MSI and HSI lies in their spectral resolution (i.e., the ability of an instrument to distinguish between different wavelengths of light) and band continuity (i.e., no gaps or breaks in the range of wavelengths being captured): MSI captures a

relatively small number of discrete, often non-contiguous spectral bands (typically ranging from 3 to tens), while HSI acquires a higher number of tightly spaced, contiguous spectral bands (typically ranging from tens to hundreds), providing a much finer and continuous spectral profile. Figure 3 compares these imaging modalities and visually illustrates the fundamental concepts of MSI and HSI.

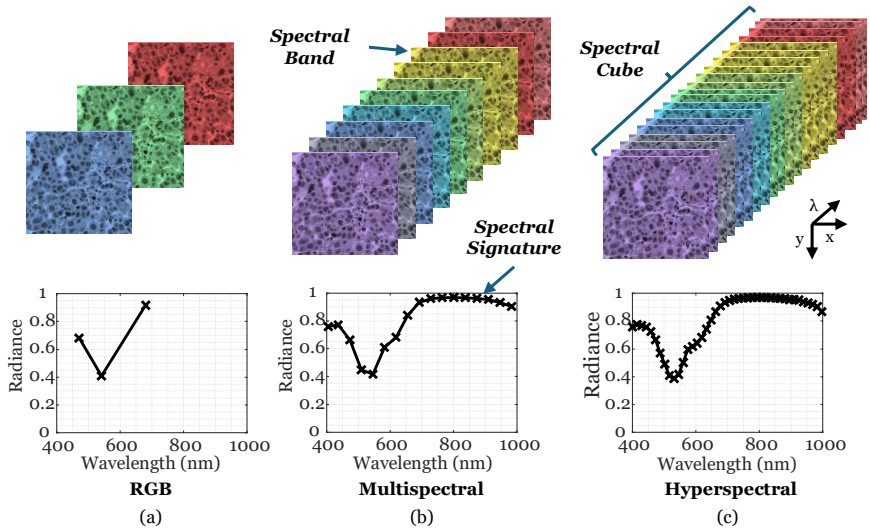


Figure 3. Illustration of different imaging modalities: (a) RGB, (b) MSI, and (c) HSI imaging. Each modality is represented as a spectral data cube composed of multiple bands. The mean spectral signature of each cube is shown below, demonstrating the progressive increase in spectral resolution from RGB to HSI.

MSI/HSI instrumentation is formed by several elements: an illumination sources, optical systems (e.g., lenses, filters), dispersive elements (e.g., gratings, prims, tunable filters), and detectors [25]. The spectral response of an HS acquisition system is determined by the combined characteristics of its individual components. The illumination source defines the emitted spectral range. In environmental applications, sunlight provides a broad and continuous spectrum that allows the full spectral range of a sensor to be utilized. In contrast, laboratory settings require carefully selected artificial light sources tailored to the spectral needs of the application. Common choices include xenon arc lamps (250–1000 nm), broadband LED arrays (400–1000 nm), and halogen or tungsten-halogen lamps (350–2500 nm) [27], [28]. Another critical

factor is the semiconductor material of each detector, whose specific quantum efficiency defines the detectable spectral range of a HS camera (Figure 4a). Silicon-based Charge-coupled devices (CCD) are widely utilized for capturing signals in the UV, VIS, and visible to near-infrared (VNIR) regions (sensitivities from 200 to 1050 nm). Detectors based on indium arsenide (InAs) or gallium arsenide (GaAs) can detect from 900 to 1700 nm, while indium gallium arsenide (InGaAs) extends this sensitivity further into the SWIR spectrum (2500 nm). For broader spectral coverage, mercury cadmium telluride (MCT) is frequently used due to its high quantum efficiency, enabling the detection from the NIR region (starting near 800 nm) to the MIR domain (approximately 2500 nm to 25,000 nm) [29]. The illumination source, specifically its intensity, and the characteristics of the sensor, such as exposure time and radiometric resolution, are also key factors that determine the dynamic range of a HS camera. Dynamic range represents the range between the strongest detectable signal (saturation) and the weakest measurable signal (noise floor) across the spectral range of interest. A wider dynamic range improves the ability to accurately detect and quantify subtle, low-intensity signals within an image, a feature commonly known as intrascene performance [30]. Furthermore, HSI systems can be classified according to the data acquisition methods, which depend on the type of dispersive elements used (Figure 4b):

- Spatial scanning technology: this technology typically uses an optical component that diffracts incoming light to fixed positions on the sensor, where spectral information is collected [31]. There are two primary types: whiskbroom systems, which acquire one spatial line at a time, and pushbroom systems, which collect an entire spatial dimension in a single acquisition. These systems are limited in spatial resolution due to sensor constraints and scanning mechanics. Furthermore, they cannot be easily adapted for applications that lack spatial motion, such as laparoscopy [31]. Acquisition time is also affected by a trade-off between scanner speed and camera frame rate.
- Spectral scanning technology: these devices filter incoming radiation to isolate specific wavelengths, capturing full spatial images at each wavelength step. Examples include Liquid Crystal Tunable Filters (LCTF) and Acousto-Optic Tunable

Filters (AOTF), whose transmission can be electronically controlled [31]. Although spectral resolution is limited by the filter's bandwidth, this method allows the use of conventional monochrome sensors and achieves high spatial resolution. Additionally, acquisition times can be shortened by selecting only the most relevant spectral channels for a given application.

- Spatial-spectral scanning technology: this approach captures diagonal slices of the HS cube by using sensors with filters arranged in stepwise lines. By acquiring images in rapid succession, these systems can provide detailed spectral information while maintaining high spatial resolution. However, the spectral detail is generally inferior to that of spatial scanning systems, and the spatial detail is not as high as in spectral scanning systems [32].
- Snapshot technology: these cameras are designed to overcome the main limitation of other HS systems: the need for spatial or spectral scanning. They capture the entire HS cube in a single frame, enabling real-time acquisition. Nevertheless, this advantage comes at the expense of both spatial and spectral resolution, which are generally lower compared to other acquisition methods (i.e., current snapshot systems are mostly limited to multispectral imaging) [33].

Finally, the multiple system components also influence spatial resolution, defined as its ability to resolve two adjacent structures as distinct [34]. Spatial resolution of a HSI system is shaped by its optical arrangement (such as magnification and lens-object distance), the dispersive elements (i.e., spatial scanning), and the sensor characteristics (including pixel size and count). For example, in HSI systems that rely on spatial scanning, the accuracy and consistency of the mechanical stage determine the spatial resolution along the movement axis. Moreover, CCD detectors generally provide fine spatial detail due to their high pixel density, while InGaAs and MCT detectors tend to have fewer pixels. This reduction in pixel density, combined with increased diffraction and decreased scattering at longer wavelengths, limits spatial resolution at wavelengths beyond 1000 nm [35].

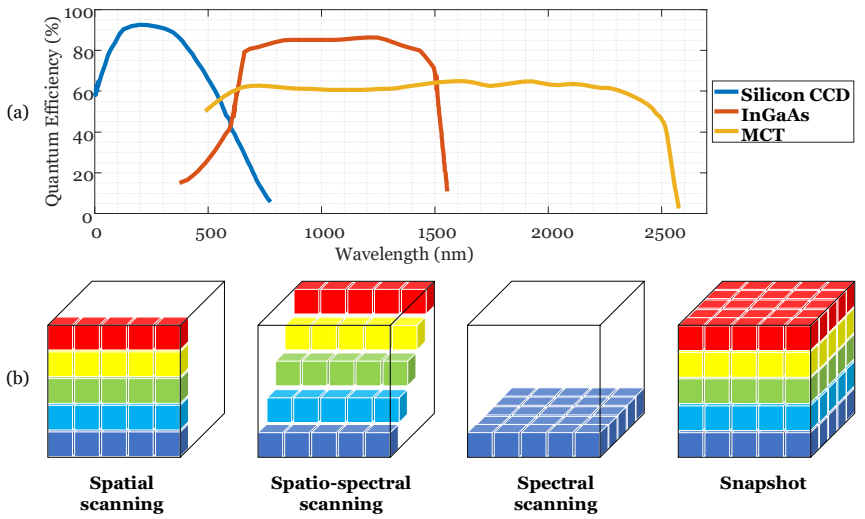


Figure 4. MSI/HSI technology classification based on: (a) quantum efficiency for different sensor materials (i.e., silicon CCD, InGaAS and MCT) (extracted from [36]) and (b) acquisition mode (spatial scanning, spatio-spectral scanning, spectral scanning and snapshot)

1.1.3 Light-Tissue Interaction

A comprehensive understanding of light-tissue interactions is crucial for the development and effective application of MSI and HSI technologies. Light comprises photons, each carrying a discrete quantum of energy $h\nu$, and traveling at the speed of light c. When photons enter biological tissue, their trajectories and energy distributions are governed by the tissue optical properties: absorption (μ_a), which quantify the likelihood of photon absorption by chromophores, and scattering (μ_s), which describes redirection due to microstructural inhomogeneities (Figure 5a) [25], [37]. In biological tissues, which are typically highly scattering, these optical properties result in complex light propagation patterns that can be effectively described by diffusion theory. This theory models the transport of photons as a stochastic process involving numerous scattering events before photons are absorbed within the tissue [38].

Absorption refers specifically to the conversion of photon energy into molecular energy, predominantly through interactions with chromophores, which are molecules that selectively absorb light at

specific wavelengths [37]. Key biological chromophores include oxygenated hemoglobin (HbO_2) [39], deoxygenated hemoglobin (Hb) [39], water [40], lipids [41], and melanin [42]. Each chromophore contributes uniquely to the overall tissue absorption spectrum (absorption coefficients illustrated in Figure 5c). For example, absorption in the IR region is largely influenced by water, which constitutes approximately 75% of tissue volume. The red to NIR spectral range, approximately from 650 to 1300 nm, is characterized by minimal absorption which facilitates deeper light penetration into tissue. This range is known as the *diagnostic window*, because the enhanced in-vivo photon penetration enables improved visualization of internal structures and more effective detection of pathological conditions [43]. While whole blood strongly absorbs light in the diagnostic window, its relatively low volume fraction within tissue moderates its effect on the average absorption coefficient. Nevertheless, photons that encounter blood vessels undergo significant localized absorption due to the high hemoglobin concentration, causing recognizable attenuation in those areas. Similarly, melanosomes are strong absorbers but occupy a relatively small fraction of the epidermis, resulting in a moderate overall contribution to tissue absorption.

Scattering arises from microscopic variations in the refractive index n within tissue at cellular and subcellular scales. These refractive index fluctuations cause photons to deviate from their original paths without energy loss, a phenomenon termed elastic scattering. The nature of scattering depends strongly on the size of the scatterers relative to the incident light wavelength. When scatterers are much smaller than the wavelength, Rayleigh scattering predominates, producing nearly isotropic scattering primarily in the UV and visible spectral regions. When scatterers are comparable to or larger than the wavelength, Mie scattering dominates, leading to predominantly forward-directed scattering typical of the visible to infrared spectral ranges [37]. Specific tissue structures significantly contribute to scattering behavior. Mitochondria, approximately 1 μm in size, contain folded lipid membranes with refractive index mismatches that generate strong scattering, particularly at longer wavelengths (~700–1300 nm range) due to their internal membrane folds [44]. Collagen fibers, typically 2–3 μm in diameter and composed of smaller fibrils approximately 0.3 μm in diameter, are

major sources of Mie scattering, especially influencing the infrared spectrum. At the ultrastructural level, the periodic banding pattern of collagen fibrils introduces Rayleigh scattering components that dominate in the visible and UV regions [45], [46].

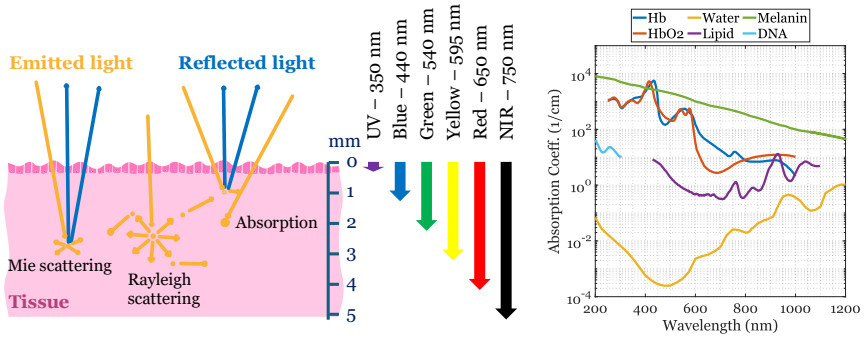


Figure 5. (a) Schematic illustration of the fundamental interactions between light and biological tissue, including absorption and scattering processes. (b) Light penetration depth as a function of wavelength, highlighting variations across ultraviolet (UV), visible, and near-infrared (NIR) ranges. (c) Absorption coefficients of key human tissue chromophores, extracted from [47], [48], including oxygenated and deoxygenated hemoglobin, water, lipids, and melanin, illustrating their spectral contributions to tissue absorption.

The combined effects of tissue optical properties, absorption and scattering, mostly determine the characteristic transport of light within biological tissue, typically quantified by the transport mean free path l_s, as defined in Equation (1). The transport mean free path represents the average distance a photon travels before undergoing significant directional change due to scattering or being absorbed. Because absorption and scattering coefficients are wavelength-dependent, the transport mean free path significantly determines the depth of light penetration (Figure 5b) [49]. Shorter wavelengths, such as UV and blue light (320 – 440 nm range), exhibit limited penetration depths, generally confined to 200–1000 μm, primarily due to strong absorption and scattering. In contrast, wavelengths within the mid-visible spectrum (440 –595 nm range) penetrate more deeply, typically reaching depths of 1 to 3 mm. Notably, longer wavelengths in the red to NIR range (700 – 1400 nm range), the diagnostic window, achieve the greatest penetration, often extending

to depths of 1 to 3 cm, thus enabling effective imaging of subsurface tissue structures [37].

$$l_s = \frac{1}{\mu_a} + \frac{1}{\mu_s} \quad (1)$$

The optical properties of tissue fundamentally determine the diffuse reflectance spectra captured by MSI and MSI systems, which have vital information about the biochemical composition of the tissue. Given that different biomolecules (e.g., hemoglobin, water, fat, melanin) exhibit unique, wavelength-dependent absorption and scattering profiles, spectrally resolved imaging techniques leverage these variations to noninvasively extract detailed biochemical and morphological characteristics [37], [50]. This capability enables accurate differentiation between tissue types, improving diagnostic precision and tissue characterization. For example, Bekina et al. [51] used MSI with four targeted wavelengths to analyze skin lesions, focusing on absorption by hemoglobin (545 nm) and melanin (660 nm). By computing a simple diagnostic index based on their intensity ratio, they found that malignant lesions, due to higher blood content, showed significantly higher values than healthy skin, enabling physiology-driven tissue differentiation. Building on this approach, Rey-Barroso et al. [52] developed a VNIR MSI technique that leveraged the complementary spectral signatures of melanin and hemoglobin to classify melanomas versus benign nevi. By incorporating a more refined spectral analysis across a broader range of wavelengths, their system achieved 79 % sensitivity and 85 % specificity. Moreover, Fei and Lu [53] developed a label-free HSI method that uses the distinct absorption patterns of oxy- and deoxy-hemoglobin, reflecting differences in blood supply and oxygenation, to accurately distinguish cancerous from healthy tissue in head and neck surgery. Their ML approach achieved around 90–94% accuracy on patient specimens, outperforming other imaging techniques and demonstrating strong potential for real-time tumor margin detection. Similarly, Kho et al. [54] leveraged the distinct absorption features of important tissue molecules (i.e., oxy- and deoxy-hemoglobin, water, and lipids) in HSI data to accurately classify breast tissue. Using over 22,000 spectra from lumpectomy specimens, their model achieved more than 90% accuracy in distinguishing malignant from healthy tissue and was able to rapidly (i.e., 10 minutes) identify tumor margins during surgery.

Other research groups have used the optical properties of tissues for enhancing cancer diagnosis by combining experimental MSI/HSI data with Monte Carlo (MC) light transport simulations. This model stochastically traces millions of photon paths through layered tissue models to predict how chromophore concentrations and scattering morphology shape measured spectra. Palmer et al. [55] applied an MC-based inverse model to UV–VIS breast-tissue reflectance, retrieving total/oxygenated hemoglobin, β-carotene and mean reduced-scattering coefficients; a support-vector classifier using only these extracted parameters reached 82 % sensitivity and 92 % specificity for malignancy detection. Also, Zhu et al. [56] coupled MC models of diffuse reflectance and fluorescence to analyze ex-vivo breast samples, showing that jointly fitted absorption, scattering and fluorescence parameters improve cancer/benign discrimination compared with fluorescence alone. These studies highlight how MSI/HSI can noninvasively capture the biochemical and morphological heterogeneity of tissues by leveraging their distinct optical properties, thereby improving diagnostic accuracy across various pathologies. Furthermore, by generating physics-based synthetic spectra that represent diverse chromophore compositions and tissue structures, MC simulations have become invaluable both as a reliable “ground truth” for training diagnostic algorithms and as a powerful tool to interpret in vivo MSI/HSI data when direct measurements of optical properties are challenging.

1.1.4 HS Microscopic Imaging

The foundations of modern microscopic analysis were laid in 1673, when Anton van Leeuwenhoek began developing simple microscopes with single lenses that offered improved magnification and resolution, enabling the first observations of cellular structures [6]. More recently, MSI/HSI have emerged as transformative technologies in histopathology by leveraging the intrinsic spectral properties of biological tissues. Unlike conventional bright-field microscopy, which is limited to the visible spectrum and often overlooks critical biochemical information, MSI/HSI acquire data across a broad range of wavelengths, offering richer, label-free

molecular insights. This expanded spectral information improves the objectivity, reproducibility, and diagnostic accuracy of histopathological assessments. In cancer diagnostics, HSI has demonstrated considerable potential by enabling the differentiation of healthy and diseased tissues based on their distinct spectral signatures. Notably, it has been successfully applied to identify pancreatic tumors with varying prognostic profiles [57], quantify Ki67 expression as a prognostic marker in lymphomas [58], and evaluate tumor–immune cell interactions in lung cancer patients undergoing immunotherapy [59]. Additionally, high diagnostic accuracy has been reported across multiple cancer types, including breast cancer [60], melanoma [61], and head and neck tumors [62]. Collectively, these findings highlight MSI/HSI as powerful, non-destructive diagnostic tools that complement and enhance traditional histopathological methods.

Although there is no standardized configuration for MS or HS microscopy, HS microscopy is usually achieved by integrating a MS or HS camera with a conventional bright-field microscope [35], [63], [64], allowing high-resolution acquisition of spectral signatures from fine tissue structures such as mammalian cells (5–25 μm) and red blood cells (6–8 μm) [7]. This integration is typically achieved through a standard C-mount, allowing compatibility with a wide range of commercial microscope models. Figure 6 illustrates a diagram of a generic HS microscope and examples of such configuration from two institutions: System A, belonging to University of Las Palmas de Gran Canaria (ULPGC) employs a pushbroom HS camera, and System B belonging to Polytechnic University of Madrid (UPM) is based on a linescan wedge sensor [65].

The light source is a bright-field halogen lamp (Figure 6f), and the configuration can operate in either transmittance or reflectance mode. Köhler illumination is used to ensure uniform radiance across the sample. In transmittance mode, light travels from the lamp through the specimen placed on the stage (Figure 6d), and continues through an objective lens optimized for infrared (IR) observation (Figure 6c). This mode benefits from low magnification, which increases the field of view (FOV), and the amount of light captured. In reflectance mode, the light is directed from above, focused

through the objective lens onto the sample surface, and reflected back through the same lens to reach the HS camera. This setup is particularly suitable for opaque or thick specimens where transmitted light cannot penetrate the sample.

In both illumination modes, HS cameras are used to acquire spectral data (Figure 6a). Since both System A and System B require spatial scanning to build the HS cube, each capturing one spatial line at a time, custom software is typically developed to synchronize the scanning stage movement with the camera's frame rate. This coordination ensures accurate and consistent data acquisition. To support the imaging workflow, the systems also incorporate binoculars (Figure 6b) for real-time visual inspection and a joystick (Figure 6e) that allows the operator to navigate across the sample and select the region of interest. These additional components make the system versatile and user-friendly, facilitating precise sample positioning before HS data capture.

For a HS/MS microscope to operate effectively, it must meet specific quality parameters. The dynamic range should be optimized in relation to the signal-to-noise ratio (SNR) to ensure reliable spectral information. If the system relies on any form of scanning, its accuracy and repeatability must be thoroughly evaluated. Additionally, both spatial and spectral resolutions should be adequate for the intended application. These parameters are essential to ensure the system delivers consistent, high-quality MS/HS data suitable for robust analysis.

The HS data acquisition process usually begins with the collection of a white reference (WR) and a dark reference (DR), both of which are essential for accurate flat-field correction. The DR is captured by turning off the illumination entirely, allowing the sensor to record the system's noise floor and electronic background signal. The WR, on the other hand, depends on the chosen illumination mode. In transmittance mode, a clear region of the microscope slide without any tissue is used. In reflectance mode, a diffuse reflectance standard, typically a matte, Lambertian surface that evenly reflects light in all directions is employed [66]. These standards are designed to approximate 100% reflectance across the spectrum, providing a consistent and reliable baseline for calibrating the system's spectral response [67].

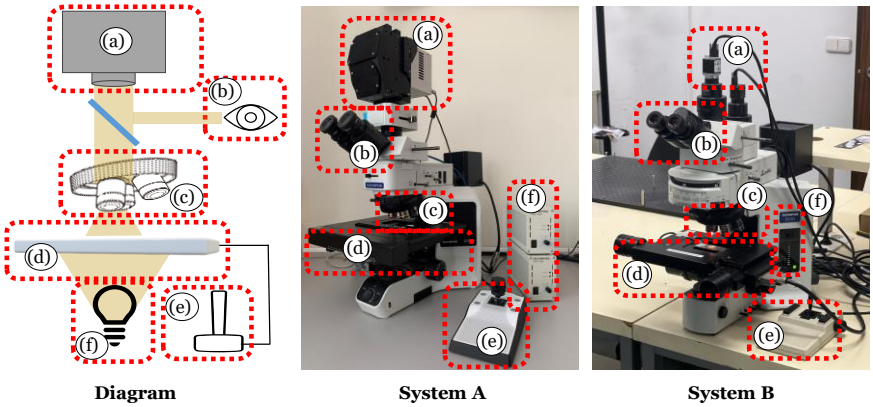


Figure 6. Diagram of HS microscopic systems (left) and example of two HS microscopic systems (System A in the middle and System B right): (a) HS camera, (b) binoculars, (c) objective lenses, (d) specimen stage, (e) joystick for spatial scanning and (f) halogen light source.

Once both reference images have been collected, the next step is to locate and select a region of interest (ROI) on the sample. This ROI is chosen based on the specific goals of the analysis (e.g., focusing on tissue areas annotated as tumor or healthy). The HS scan of the selected ROI is then performed, and the resulting raw data cube (R) is stored for subsequent analysis.

The first and most critical signal processing step is flat-field correction, which compensates for inconsistencies introduced by sensor characteristics, lighting conditions, and optical imperfections [68]. This correction transforms the raw digital values (R) captured by the sensor into physically meaningful reflectance [69], [70], [71] or radiance values [72]. The standard formula used to compute a calibrated HS image (CI) from the raw data is [35]:

$$CI(\%) = \frac{R - DR}{WR - DR} \cdot 100\ \% \qquad (2)$$

Subsequent processing may vary depending on the specific use case [73], [74]. One common step is spectral dimensionality reduction, which addresses the redundancy across adjacent spectral bands and helps reduce both noise and computational complexity [73], [75]. For example, averaging every three neighboring bands could reduce a dataset size, improve efficiency, while retaining essential spectral information. Normalization is also often applied to facilitate consistent data analysis, especially in classification tasks

where relative differences in spectral shape are more important than absolute intensity values. Common normalization algorithms can rescale the data to a fixed range (e.g., 0 to 1) or standardize it to have zero mean and unit variance [76]. In most analyses, it is also important to exclude non-informative background pixels, such as those representing the slide or cover slip, which exhibit minimal or no absorbance. This can be done by thresholding on spectral intensity or using automated background masking techniques [77]. Finally, depending on the research objective, the processed MS/HS data may be used in traditional image analysis pipelines (e.g., for tissue segmentation) or fed into ML/DL models for automated classification [25], [26].

Recent advances in HS microscopy have demonstrated its potential to significantly enhance cancer diagnostics in histopathology. For instance, Ma et al. [78] implemented a compact HSI microscopic system for the automatic detection of squamous cell carcinoma (SCC). They combined principal component analysis (PCA)-based nuclei segmentation with spectra-based support vector machines (SVM) and patch-based convolutional neural networks (CNN). The DL model outperformed traditional ML classifiers (accuracy of 82.4% and an area under the curve (AUC) of 93.7%), demonstrating its enhanced diagnostic capability. Further work by the same group [79] evaluated DL models on 77 histologic slides from 51 patients, showing that a spatial-spectral vision transformer (SST) trained on HSI achieved 79% accuracy, 74.4% specificity, and 78.9% sensitivity—outperforming models trained on RGB images. Notably, the DinoV2 model trained on HSI showed a 10% increase in sensitivity compared to its RGB counterpart, emphasizing the diagnostic value of HSI in combination with advanced DL architectures. Similarly, Hu et al. [80] constructed a spectral-spatial database from gastric cancer pathological samples and developed a CNN that effectively extracted deep spectral-spatial features within the 410–910 nm wavelength range, attaining a classification accuracy of 97.57%, with sensitivity and specificity values of 97.19% and 97.96%, respectively. The study illustrated the practical value of combining DL with microscopic spectral data to support more accurate and automated gastric cancer diagnostics. Ortega et al. [81] utilized a VNIR range HSI microscope coupled with a nine-layer CNN to discriminate between normal and tumor breast cancer cells,

achieving AUC values exceeding 89%, thus further confirming the potential of HSI combined with artificial intelligence (AI) for automated cancer cell identification. In the context of brain cancer, Ortega et al. [82] applied HSI and supervised classifiers (i.e., SVM, artificial neural networks, and random forests) to distinguish tumor from normal brain tissue in glioblastoma samples, achieving over 80% accuracy while addressing challenges related to inter-patient variability and limited training data. Finally, Pertzborn et al. [83] demonstrated the feasibility of using HSI coupled with a supervised three-dimensional CNN for intraoperative tumor margin assessment in oral squamous cell carcinoma. Their approach yielded 76% accuracy, 89% specificity, and 48% sensitivity on unstained tissue sections, supporting the use of HSI as a complementary tool for real-time surgical guidance. Collectively, these studies illustrate the substantial advances and clinical potential of integrating HSI with microscopy to enhance biomedical imaging, improve diagnostic accuracy, and support more informed therapeutic assessment.

1.2 Knowledge gap

Beyond its diagnostic potential, MSI/HSI microscopy offers several practical advantages: it preserves the native biochemical composition of samples, supports digital storage and post-acquisition analysis, and generates objective, reproducible data suitable for integration into ML/DL workflows. However, despite these strengths, clinical adoption remains limited. The technology is still largely confined to experimental settings, and several technical, methodological, and regulatory challenges must be addressed before MSI/HSI can be routinely implemented in pathology laboratories [26].

Lack of standardization of HSI microscopic instrumentation

One of the principal challenges is the lack of standardization across HSI systems. The heterogeneity of custom-built HSI systems, combined with the limited availability and validation of commercial prototypes, has become a major bottleneck in the development of reliable diagnostic tools. Each research group may use a different

combination of cameras, optics, light sources, and acquisition methods. Without standardization, this diversity leads to inconsistencies in data quality and hinders collaborative efforts [26], [84]. While efforts toward establishing universal standards for HSI microscopy are in progress, they remain at an early stage. For example, the ongoing development of the "Standard for Characterization and Calibration of Ultraviolet through Shortwave Infrared (250 nm to 2500 nm) Hyperspectral Imaging Devices" [85] does not address the specific requirements of HSI microscopic systems. Also, existing protocols frequently overlook critical optical, spatial, and spectral constraints unique to microscopy-based imaging [35], [64], [86]. Key performance metrics in HSI systems span a diverse set of characteristics that collectively define system capability and determine its suitability for specific clinical or research applications. Among these, spectral sensitivity, the ability of the sensor to detect light as a function of its signal frequency, is crucial for accurate tissue classification [87]. Spatial resolution, influenced by the optical system's numerical aperture, pixel pitch, and magnification, governs the ability to discern microscopic structural details within the sample [34]. Equally important is the SNR, which impacts the detectability of subtle spectral variations, especially in weakly absorbing or low-contrast biological tissues [88]. Radiometric accuracy [68] and dynamic range [30] are critical for capturing accurate spectral intensities across a broad range of reflectance levels. Another important concern is repeatability, the ability of the system to consistently produce the same results under comparable external conditions [89]. Altogether, imaging the same biological sample with different HSI systems can produce inconsistent results, which undermines the development of robust algorithms and delays clinical validation. To mitigate this variability, key characterization parameters must be carefully optimized and balanced in alignment with the specific clinical or research application. This requires a systematic approach to hardware characterization, spectral calibration, and overall system validation. While initial efforts have been made to define guidelines and best practices, a universally accepted and comprehensive framework has yet to be established.

Deficit of Image Quality Strategies in HS Microscopy

Beyond system configuration, HS microscopic image acquisition protocols play a crucial role in data quality. One often-overlooked variable is sample focus. In microscopy, particularly at high magnifications, maintaining sharp focus is vital. Improper focus causes blurring, loss of spatial detail, and distortion of spectral signatures due to spatial averaging of heterogeneous tissue components. While this challenge has been largely addressed in the context of WSI, where automated focus mapping and dynamic refocusing techniques are commonly employed, in HSI microscopy, this issue is exacerbated by the narrow depth of field (DOF) inherent to high-magnification optics [90], where even slight deviations in sample flatness or mounting can shift regions of interest out of focus. Furthermore, HSI systems, particularly those based on scanning architectures, often require several seconds to minutes for image acquisition, during which mechanical drift can further degrade focus [35]. These effects lead to spatial and spectral inconsistencies across the FOV, ultimately undermining the accuracy of tissue algorithms [91]. These distortions are challenging to correct post-acquisition and can significantly compromise spectral integrity, particularly in high-precision applications such as tumor margin assessment or subcellular classification. To mitigate these issues, robust focus control protocols must be incorporated into the HS acquisition framework. One effective solution to blurring caused by imaging non-flat samples is the use of z-stacking techniques, which capture multiple images at varying focal planes and combine them into a single sharply focused composite image [92]. For this process to be fully automated, reliable image quality assessment (IQA) methods must first evaluate the focus quality of each image [93]. Although the human eye remains the most precise evaluator of image quality, being the ultimate end-user, subjective assessments are costly and inefficient [94]. Therefore, objective and efficient IQA algorithms are essential to automatically predict image quality, enabling accurate selection of focused images across the entire spectral range and ensuring consistent spatial and spectral integrity.

Need for Comprehensive Data Processing in Medical HS Microscopy

Moreover, given the variability introduced during acquisition, processing becomes an essential step once the HS data is captured, helping to correct artifacts and standardize input for further ML/DL classification (e.g. inhomogeneities in illumination and noise in acquisition systems). These processing steps are essential for improving information extraction, enabling more accurate tissue classification by, for example, compensating for absorption effects described by Lambert-Beer's law or enhancing the performance of ML/DL models. Common processing techniques include noise filtering to reduce random signal variations, illumination correction to compensate for uneven lighting, baseline offset removal to eliminate background shifts, normalization to scale spectra for comparison, and spectral derivatives to find characteristic peaks on the spectra. These operations help reduce variability introduced by acquisition conditions, hardware differences, or environmental factors [73], [74], [95]. However, in the context of medical HSI, particularly microscopy, there is currently no universally accepted pre-processing pipeline [76], [96]. Researchers often rely on empirical methods or adapt workflows from adjacent fields such as chemometrics [97], [98] or remote sensing [75], which may not be fully suitable for histological data. A comprehensive pre-processing pipeline should not only enhance spectral clarity and reduce noise but also ensure spatial sharpness and consistency throughout the dataset. This integration is essential for the development of reliable, reproducible, and diagnostically useful HSI systems.

Absence of Sample Preparation Protocols Tailored for HS Microscopy

Furthermore, a critical factor that directly affects HS data quality, and by extension the diagnostic performance of HSI microscopy systems, is sample preparation, with tissue thickness playing a central role in modulating light-tissue interactions. While considerable effort has been devoted to advancing sensor technologies and imaging modalities in recent years [73], [99], [100], [101], sample preparation protocols have largely remained aligned with conventional histology practices. Biopsies are typically embedded in paraffin or frozen, then sectioned into 2–10 μm slices

using a microtome or cryostat [7], a range optimized for bright-field microscopy rather than spectroscopic imaging. In HSI microscopy, however, this standard thickness may not be optimal. The technique depends on capturing rich spectral information derived from how light interacts with tissue structures. If sections are too thin, photons traverse the sample with minimal scattering or absorption, resulting in low spectral contrast and a poor SNR ratio. Conversely, overly thick sections introduce excessive scattering and absorption, which can distort spectral signatures and hinder accurate material discrimination [38]. These effects are particularly detrimental in high-precision applications such as tumor characterization or subcellular analysis, where even subtle spectral variations carry diagnostic weight. Moreover, light behavior at microscopic scales differs from that in bulk tissue, meaning that the optical path length, and therefore the resulting spectra, can vary non-linearly with thickness. Such variability undermines spectral consistency across samples, complicating downstream tasks like classification and segmentation. To mitigate these issues, HSI-specific sample preparation protocols may be required. Adjustments to sectioning thickness, mounting media, and dehydration protocols can significantly influence the spectral quality and reproducibility of acquired data. Thus, establishing standardized, application-specific sample preparation workflows for HSI microscopy is required for ensuring the robustness and clinical reliability of these systems.

Limited Availability of Public HS Microscopic Datasets

A significant limitation in the current landscape of HS imaging research is the lack of publicly available HS microscopic databases. This scarcity presents a barrier to progress for the research community, particularly for those working on the development and benchmarking of advanced HS image processing techniques. Without access to standardized datasets, it becomes difficult to compare the performance of methods such as HS image classification, spectral unmixing, and data compression in a consistent and reproducible manner. Public databases would provide a shared foundation for researchers to train and validate algorithms under comparable conditions, fostering collaboration and accelerating technological advancement. Beyond the algorithmic development community, such resources would also be

invaluable to researchers in computational pathology and practicing pathologists. Exposure to real HS microscopic data would allow them to assess the potential of this technology for clinical applications, such as improving diagnostic precision, supporting decision-making, or revealing spectral features not visible with traditional imaging modalities. The availability of such datasets could help bridge the gap between technological innovation and clinical translation, ultimately contributing to the adoption of HS imaging in routine pathological workflows.

1.3 Hypothesis

HS microscopy holds significant promise as a diagnostic modality, particularly in the early detection of cancer, by enabling non-invasive, label-free analysis of tissue morphology and biochemical composition. Its capacity to capture rich spatial and spectral information at the microscopic scale makes it a compelling tool for histopathological assessment and computer-aided diagnosis. Although HS microscopy shows great potential, its adoption in clinical settings is still limited. Several challenges impede its effective, reproducible, and integrated use in routine biomedical applications. This Ph.D. thesis is guided by two central hypotheses aimed at advancing HS microscopy toward clinical applicability:

H1. HSI can be effectively adapted to the microscopic scale to enable non-invasive, label-free, and quantitative analysis of biological tissues, providing spectral signatures that reflect subtle biochemical and structural variations and are diagnostically relevant.

H2. The development and clinical validation of a HSI-based microscopic diagnostic tool, driven by data reproducibility and the need for robust inter-laboratory comparability, can be significantly enhanced through the standardization of HS microscopic system configurations, acquisition protocols, and tissue preparation parameters combined with advanced data processing techniques.

1.4 Objectives

The main objective of this Ph.D. thesis is to develop and validate standardized methods for system configuration, acquisition, preprocessing, and sample preparation in HS microscopic imaging to improve data reproducibility, enable inter-system comparability, and enhance diagnostic reliability. This objective can be divided into:

O1. To analyze and evaluate **HS acquisition systems** for microscopic applications:
   a. Identify the critical parameters influencing system performance in biological sample imaging.
   b. Determine optimal system configurations for the capture of histological samples.

O2. To obtain the necessary knowledge to investigate and optimize **pre-acquisition parameters** for improving **HS microscopic imaging quality**:
   a. Analyze factors affecting image focus, illumination, and signal quality prior to data capture.
   b. Develop strategies and protocols for consistent and reproducible image acquisition.

O3. To develop and implement **image processing techniques** for HS microscopic data after acquisition:
   a. Design methods for noise reduction, artifact correction, and spectral enhancement.
   b. Extract diagnostically relevant information from HS images.

O4. To investigate the impact of **biological sample preparation** on HS microscopic data:
   a. Evaluate how variables such as tissue thickness affect image spatial and spectral characteristics.
   b. Propose guidelines for sample preparation that improve image quality and diagnostic reliability.

O5. To acquire **HS images of histological samples** under varying preparation and acquisition conditions:
   a. Capture data across different sample types.
   b. Create representative datasets of practical laboratory conditions to support algorithm development and validation.

1.5 Conceptual Framework and Methodology

A comprehensive conceptual framework has been developed to define the methodology to follow for the creation of a robust HS microscopic diagnostic tool, aligned with the previously outlined hypothesis and objectives (Figure 7). This framework is composed of six interrelated and essential stages that collectively address the complexities inherent in HS microscopy. The first stage, system instrumentation and characterization, focuses on establishing reliable and well-calibrated hardware capable of consistent and reproducible data acquisition (Figure 7a). Next, focus assessment and optimization refines optical parameters and imaging conditions to ensure image quality (Figure 7b). The third stage, sample processing optimization, aims to refine protocols for handling biological specimens in HS microscopy imaging (Figure 7c).

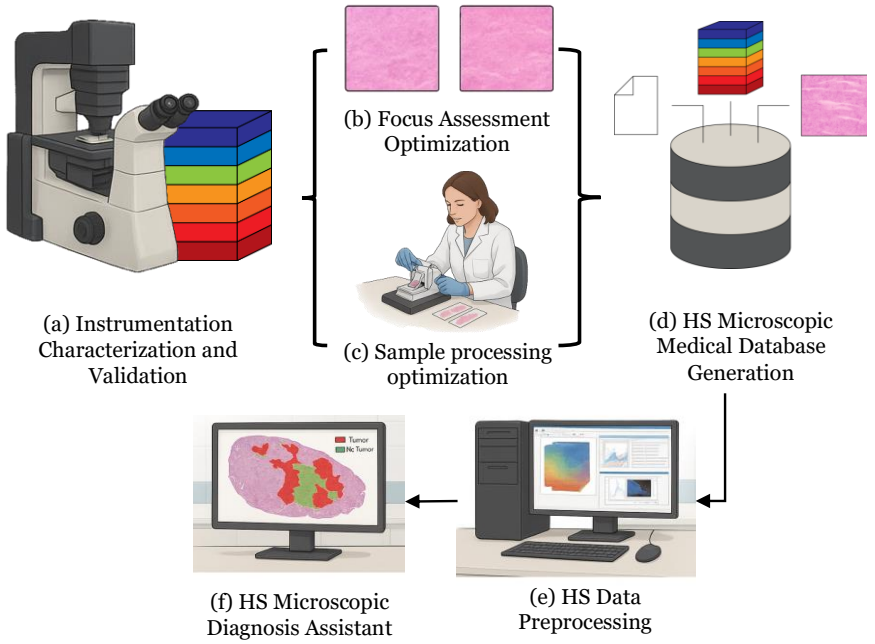


Figure 7. HSI Microscopic diagnostic tool development framework: (a) instrumentation, characterization, and validation, (b) focus assessment optimization, (c) sample processing optimization, (d) HS microscopic medical database generation, (e) HS data pre-processing and (e) automatic diagnosis assistant.

Following, the fourth stage, HS microscopic database generation, emphasizes the creation of publicly available HS microscopic datasets for algorithms design and development (Figure 7d). The fifth stage involves HSI-specific processing workflows, which are critical for cleaning, normalizing, and preparing spectral data for further analysis (Figure 7e). Finally, the framework culminates in the development of an automatic diagnosis assistant, where advanced algorithms extract meaningful insights from HS microscopic data to support accurate decision-making (Figure 7f). By systematically addressing each of these stages, this framework aims to bridge the gap between experimental research and practical clinical application, ultimately advancing the utility of HS microscopy in biomedical diagnostics.

1.6 Document organization

This document has been structured in 6 chapters, following the consecutive steps necessary to carry out the conceptual framework to develop a HS microscopic diagnosis assistant tool (Figure 7). Each chapter includes its corresponding research outputs, categorized as journal publications (J) and conference presentations (C). Additionally, any related research mobility (M) activities, such as research stays, are also indicated where applicable. A brief explanation of each chapter is presented next.

Chapter 1: **Introduction.** In this chapter, the motivation behind the Ph.D. thesis is explored, along with some background to help understand its context. The main goals of the work as well as a conceptual framework and methodology are outlined. And, finally, an overview of the organization of the manuscript is provided.

Chapter 2: **HS Microscopic System.** This chapter focuses on the evaluation and design of HS microscopic systems for medical histology. It includes an in-depth analysis of instrumentation performance and outlines a roadmap for the development and

characterization of tailored HSI setups aimed at enhancing histological assessment.

Chapter 3: **Image Quality Optimization.** This chapter addresses the enhancement of HS microscopic image quality, with a focus on blur-specific no-reference assessment methods for focus quantification. It also explores tailored image quality metrics and presents processing strategies to improve the reliability of medical HS data analysis.

Chapter 4: **Light - Histology Sample Interaction.** This chapter investigates how sample preparation, specifically tissue thickness, influences data obtained from HS microscopy. It examines whether variations in thickness affect image quality and spectral information, providing insights into optimizing histological sample preparation for more reliable HSI analysis.

Chapter 5: **HS microscopic Medical Databases.** This chapter presents the development of publicly available HS microscopic databases, addressing the current lack of standardized datasets. It highlights its importance for benchmarking image processing methods and supporting reproducible research in biomedical HS microscopic imaging.

Chapter 6: **Conclusions & Future Lines.**

Bibliography: List of references employed in the manuscript.

Annex A: Resumen en Español.

Chapter 2. HS Microscopic System

HSI is a powerful imaging technique capable of capturing both spatial and spectral information, enabling the non-invasive characterization of materials, tissues, and other elements. While HSI is widely adopted at the macroscopic level, its application in the microscopic domain remains limited due to several technical challenges. One of the main limitations is the absence of standardized methodologies for characterizing HS microscopic systems, which are essential to ensure proper data acquisition and experimental reproducibility across different research institutions.

This chapter contributes to Objective **O1** by focusing on the analysis and assessment of HS microscopic acquisition systems tailored for biological sample imaging. Specifically, it aims to identify the essential parameters that affect system performance, such as spatial and spectral resolution, noise behavior, and geometric fidelity, and to define optimal configurations for effectively imaging histological specimens. To achieve these goals, a comprehensive characterization roadmap is proposed. This roadmap brings together key parameters identified in the current state of the art and includes a detailed list of materials required for system evaluation. The methodology assesses a wide range of performance metrics, including dynamic range, noise quantification, pixel size, spatial frequency response, spatial scanning accuracy, spatial repeatability, flat-field correction, tone transfer, and spectral sensitivity. These metrics are integrated into a unified, standardized, and replicable framework that has been validated through a round-robin test using independent HS microscopic systems from two institutions, demonstrating its generalizability. Consequently, the roadmap not only promotes consistency and reliability in the development and validation of HS microscopic systems but also

provides clear guidance for future research to publicly report quality metrics, promoting consistency and reliability in the development and validation of HSI microscopic systems.

Preliminary results were presented in the conference paper (**C1)** titled **"Instrumentation Evaluation for Hyperspectral Microscopy Targeting Enhanced Medical Histology"** at the *XXXVI Conference on Design of Circuits and Integrated Systems (DCIS)*, held in Portugal in 2021. This work introduced the main challenges associated with HS microscopic instrumentation. The full methodology was later developed and tested through a collaborative study with the Research Center on Software Technologies and Multimedia Systems at Universidad Politécnica de Madrid, during the Ph.D. Candidate's research stay (**M1**). The proposed roadmap was validated on two representative HS microscopic systems, one located at the University of Las Palmas de Gran Canaria and the other at Universidad Politécnica de Madrid. Both systems integrate an HSI sensor into a bright-field microscope, a configuration commonly used in biomedical imaging.

The results of this work were published in the journal article (**J1**) titled **"Roadmap for the Design and Characterization of Microscopic Hyperspectral Systems"**, in *IEEE Transactions on Instrumentation and Measurement*. Overall, the proposed characterization roadmap provides practical and detailed guidelines for the scientific community aiming to develop and evaluate reliable, efficient, and accurate HS microscopic systems.

J1. Roadmap for the Design and Characterization of Microscopic Hyperspectral Systems

Title:	Roadmap for the Characterization and Validation of Hyperspectral Microscopic Systems		
Authors:	**Laura Quintana-Quintana**, Gonzalo Rosa, Javier Santana-Nunez, Miguel Chavarrías, Samuel Ortega, Jaime Sancho, Himar Fabelo, Eduardo Juárez, and Gustavo M. Callico		
Journal:	Transactions on Instrumentation and Measurement		
Publisher:	IEEE		
Date: 2025	**Vol.**: 74	**Pages**: 1-13	
doi:	https://doi.org/10.1109/TIM.2025.3575989		
JIF Quartile:	Q1 in Instruments & Instrumentation	**IF:**	5.9
Open Access, National Collaboration, Corresponding Author			

Roadmap for the Characterization and Validation of Hyperspectral Microscopic Systems

Laura Quintana-Quintana, Gonzalo Rosa Olmeda, Javier Santana-Nunez, Miguel Chavarrías, Samuel Ortega, Jaime Sancho, Himar Fabelo, Eduardo Juárez, *Member, IEEE*, and Gustavo M. Callico, *Senior Member, IEEE*

***Abstract*—Hyperspectral imaging (HSI) is a powerful image technique that allows capturing spatial and spectral information, being able to characterize materials, tissues, and elements in a noninvasive manner. HSI technology is well established at the macroscopic level, but there are still technical challenges to overcome before it can be applied to the microscopic world, such as the lack of standardized characterization methodologies to HS microscopic systems that allow the correct data acquisition as well as ensure the repeatability of the experiments. In this work, we propose a comprehensive roadmap for characterizing and validating such systems, integrating essential parameters highlighted in the current state of the art. Furthermore, we provide a list of the materials needed for their characterization and testing of the methodology on two different HS microscopic systems chosen as representative of common configurations in the field, where an HS camera is integrated into a bright-field microscope. Our proposed roadmap assesses the following parameters: dynamic range (DR), noise quantification, pixel size, spatial frequency response (SFR), spatial scanning accuracy, spatial repeatability, flat-field correction, tone transfer, and spectral sensitivity. We address the challenge of unifying these parameters into a unified and standardized roadmap. All data used to characterize both systems have been captured by the authors. In summary, this comprehensive analysis provides a guideline for the scientific community to develop and characterize HS microscopic systems to ensure reliability, efficiency, and accuracy.**

***Index Terms*—Hyperspectral (HS) imaging (HSI), instrumentation, microscopy, system characterization.**

Received 2 December 2024; revised 25 March 2025; accepted 20 May 2025. Date of publication 4 June 2025; date of current version 24 June 2025. This work was supported in part by the Spanish Government and European Union (FEDER Funds) through the TALENT (HypErsPEctRal Imaging for Artificial Intelligence Applications) under Project PID2020-116417RB-C41 and Project PID2020-116417RB-C42 AEI/10.13039/501100011033; in part by the European Commission through the HORIZON-HLTH-2023-TOOL-05-05, European Project STRATUM (3-D Decision Support Tool for Brain Tumor Surgery), under Grant 101137416; in part the Spanish Ministry of Science and Innovation and European Union (FEDER Funds) through the Open Ai-driven Stack para plataformas hpec mejoradas en Sistemas Integrado (OASIS) Project under Grant PID2023-148285OB-C43 and Grant PID2023-148285OBC44; and in part by the Norwegian Research Council through the IKTPLUSS-IKT og digital innovasjon Project under Grant 332901. The work of Laura Quintana-Quintana was supported in part by the Consejería de Universidades, Ciencia e Innovación y Cultura; in part by the Fondo Social Europeo Plus; and in part by the Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI) of the Consejería de Economía, Conocimiento y Empleo, European Social Fund (FSE) [POC 2014-2020, Eje 3 Tema Prioritario 74 (85%)]. The work of Javier Santana-Nunez was supported by the Investigo Program of Canary Islands Employment Service (Fondos del Plan de Recuperación, Transformación y Resiliencia—Next Generation European Union (EU)), under Grant 32/39/2022-0923131539. The Associate Editor coordinating the review process was Dr. Kamel Haddadi. *(Corresponding author: Laura Quintana-Quintana.)*

Laura Quintana-Quintana and Gustavo M. Callico are with the Research Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain (e-mail: lquintana@iuma.ulpgc.es).

Gonzalo Rosa Olmeda, Miguel Chavarrías, Jaime Sancho, and Eduardo Juárez are with the Research Center on Software Technologies and Multimedia Systems, Universidad Politécnica de Madrid, 28040 Madrid, Spain (e-mail: gonzalo.rosa.olmeda@upm.es).

Javier Santana-Nunez is with the Fundación Canaria Instituto de Investigación Sanitaria de Canarias, 35019 Las Palmas de Gran Canaria, Spain, and also with the Research Unit, Hospital Universitario de Gran Canaria Dr. Negrin, 35019 Las Palmas de Gran Canaria, Spain.

Samuel Ortega is with the Norwegian Institute of Food, Fisheries and Aquaculture Research, 9291 Tromsø, Norway, also with the Researcher Department of Mathematics and Statistics, UiT The Arctic University of Norway, 9019 Tromsø, Norway, and also with the Research Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain.

Himar Fabelo is with the Research Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain.

This article has supplementary downloadable material available at https://doi.org/10.1109/TIM.2025.3575989, provided by the authors.

Digital Object Identifier 10.1109/TIM.2025.3575989

I. Introduction

SPECTRAL analysis techniques allow to gather information noninvasively from the medium by analyzing the response of light interacting with different materials or tissues. This interaction can be produced in terms of transmission, reflection, absorption, and scattering [1]. For this purpose, multispectral or hyperspectral (HS) imaging (HSI) systems are used to record both spatial information and spectral information of a particular scene. Multispectral systems capture a limited number of discrete bands, in the order of tens, while HS systems are able to capture a much larger number of bands, allowing for a more comprehensive spectral analysis [2].

While HSI technology is well established at the macroscopic level [3], [4], [5], the latest integration of HSI with optical microscopy has opened up new opportunities for the analysis of samples at a microscopic scale [6]. This has led to applications such as quality assessment of food products [7], identification of different bacteria and contaminants [8], classification of microplastics [9], [10], or extracting information from microalgae [11]. Fields like cultural forensics or medical imaging [12] have also benefited from microscopic HSI tools to study cultural heritage samples [13], detection of different types of cancers [14], [15], or the identification of different types of blood cells [16].

Currently, the most prominent approach to extract useful information from HS images is the use of machine learning and deep learning techniques [17]. The potential of such techniques relies on training numerous datasets with high amount of HS images, ensuring minimal variation within each class to maintain consistent feature representation and improve classification accuracy [18], [19]. However, the absence of

commercial HS microscopic systems and the extensive use of custom and nonstandardized instrumentation developed by researchers for their standalone application make consistent data acquisition a challenge. Pillay et al. [20] analyze eight different macroscopic commercial HSI systems, and a notorious amount of variability was found among their captures. The data showed spectral, geometric, and colorimetric inaccuracies, along with common residual errors, and substantial differences in noise levels. This variability is likely to be even greater when working with ad hoc systems at the microscopic scale. Thus, despite its potential, the characterization and validation of HS microscopic systems present significant challenges. The lack of standardized procedures for characterizing these systems inhibits their portability across different applications and research groups, resulting in inconsistent results and limited reproducibility [21].

To address the challenge of characterizing HS microscopic systems, several researchers have proposed ad hoc solutions. Ortega et al. [22] developed a custom 3-D printed mechanical system for a precise push-broom HS microscopic system and characterized its robustness by measuring some of its spatial features. Similarly, Stergar et al. [23] introduce a method for spectral characterization by simply measuring the full-width at half-maximum (FWHM) response of their HS microscopic system. Meanwhile, Paterova et al. [24] mixed both spatial and spectral characterization by measuring their system resolution on both dimensions. However, the characterization procedures developed cannot be applied interchangeably to ad hoc HS microscopic systems, underlining the need for standardized methods that can provide consistent results across different setups [21]. These contributions demonstrate a growing interest in characterizing HS microscopic systems, but they also highlight the need for a comprehensive framework that addresses the full range of parameters involved.

To the best of our knowledge, this work is the first to provide a systematic approach to characterizing and validating HS microscopic systems. The contribution of this article includes: 1) a complete list of parameters to be measured; 2) their theoretical background; 3) metrics to evaluate their quality; 4) materials needed for measurement; and 5) descriptions of HS image processing methods required to derive those parameters. While individual methods for specific aspects of characterization are established in the state of the art, their integration into a unified and standardized roadmap, as the one we propose, remains an open challenge. Our structured methodology consolidates these characterization aspects into a single, widely applicable approach, essential to ensure repeatability and comparability across different systems. Our roadmap is proposed within the context of standardization efforts in the field of HSI, as exemplified by the ongoing development of a "Standard for Characterization and Calibration of Ultraviolet through Shortwave Infrared (250–2500 nm) HSI Devices" [25], which does not include guidelines for HS microscopic systems. Additionally, the proposed roadmap is validated through a round-robin test, i.e., an interlaboratory test. Two systems were previously developed by integrating an HS camera into a bright-field microscope, following the common approach for constructing HS microscopic systems [6], [7], [11], [13], [15], [21], [22], [23], [24]. For each system, a different spectral acquisition principle is used as a reference: spatial scanning and linescan wedge HS cameras. In this way, the roadmap is validated and allows the results to be then extrapolated to other types of HS technologies, such as spectral scanning or snapshot-based HS cameras [12].

II. Theoretical Background

This section briefly describes each one of the parameters to be used for the HS microscopic system characterization and validation, providing the necessary background for carrying out the subsequent work.

A. Dynamic Range

The dynamic range (DR) characterizes the variation of the maximum (saturation) and minimum (noise levels) across the spectrum of interest. From the characterization perspective, an exposure time is selected to optimize the DR of the sensor. Furthermore, in HSI capture systems, the DR varies across the spectrum due to differences in the quantum efficiency of the sensor with respect to the wavelengths. The DR for a specific monochromatic image, depicted in (1), is the ratio between the highest intensity captured by the sensor without saturation (N_{sat}) and the lowest intensity captured over the background noise (N_{noise}), i.e., values captured by the sensor when no light interacts with the sensor [26]. In this context, intensity refers to digital values produced by the sensor's analog-to-digital (A/D) converters, which are dimensionless and expressed as digital numbers (DNs). Although these values do not directly represent any physically meaningful magnitude, they are proportional to the detected light intensity based on the sensor's response characteristics

$$\text{DR}\,(\text{dB}) = 20\log\left(\frac{N_{\text{sat}}}{N_{\text{noise}}}\right). \tag{1}$$

The maximum DR that a system can capture depends on the number of bits (n) of the analog-to-digital converter (ADC) of each camera (2). In brief, the higher the DR, the better the ability to quantitatively measure the dimmer intensities within an image; this feature is also known as intrascene performance. Typically, on reflectance-based acquisition systems, a chart compliant with the ISO-21550 standard [27] is used to determine the DR of the system. In the case of transmittance-based systems, an empty sample holder is used

$$\text{DR}_{\text{max}}(\text{dB}) = 20\log\left(2^{n}\right). \tag{2}$$

B. Noise Quantification

All acquisition systems are affected by noise due to the electronics used to convert physical magnitudes into digital values. In the case of digital cameras, whether red, green and blue (RGB) or HS, this noise is due to the movement of electrons in the photoreceptors of the sensors when no photons are interacting with the sensor, which is called dark current (DC) [28]. Quantifying the DC of a system allows the signal-to-noise ratio (SNR) to be calculated. This metric shows how much useful information is captured by an imaging system, where an SNR of 0 dB means that the system cannot discriminate signal from noise. Following the methodology

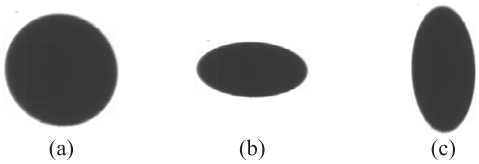


Fig. 1. Dot target captured at different speeds. (a) Optimal speed, (b) too fast, and (c) too slow speed.

proposed by Shaikh et al. [29], the SNR is obtained by capturing an image of a reference target at different exposure times and then capturing an image with the optics completely covered at the same exposure times, to extract the DC. Then, the mean of the HS image is divided by the standard deviation of the DC as follows:

$$\mathrm{SNR\,(dB)} = 20\log\left(\frac{\mathrm{mean\,(HS)}}{\mathrm{std\,(DC)}}\right). \tag{3}$$

As can be observed in both Section II-A and in this Section II-B, only the noise generated by the DC has been considered as the sole noise source. This is because it is the nonimaging error that has the greatest influence on the characterization of the systems, compared to quantum efficiency, gain, or vignetting [30]. In addition, the current state-of-the-art proposes different ways of digitally denoising images, but all of them lack an in-depth analysis of these noise sources [31], [32], [33].

C. Spatial Scanning Accuracy

When integrating a spatial scanning HS camera into a microscopic system, it is crucial for the spatial scanning to be adequate to obtain properly formed HS images. The theoretical platform speed can be established as the ratio between the space captured by one frame and the time required to perform the capture, also known as the frame period. In practice, the frame period may slightly drift from the desired value, and the frames might not be captured synchronized with the platform's stops. Thus, an additional analysis is needed to enhance and validate an accurate system.

The methodology to quantitatively evaluate the spatial scanning accuracy was described by Ortega et al. [22]. In that work, the authors detected camera misalignments and suboptimal movement speeds by assessing the round shape (eccentricity) of a captured circle target. A circle exhibits a flawless round shape when captured at the optimal speed [Fig. 1(a)] but presents an elliptical appearance when the speed is too high [Fig. 1(b)] or too low [Fig. 1(c)]. To quantify how much the ellipse deviates from being a perfect circle, the eccentricity of the circle is calculated at each magnification of the HS microscopic systems. First, 2-D principal component analysis (PCA) is computed over a binary image, generating two eigenvalues ($\phi_{\min}$ and $\phi_{\max}$) that correspond to the directions of the ellipse's longest and shortest axes. From these values, eccentricity can be computed using (4). A perfect circle yields eccentricity values close to zero [22], [34]

$$e = \sqrt{1 - \frac{(\phi_{\min})^2}{(\phi_{\max})^2}}. \tag{4}$$

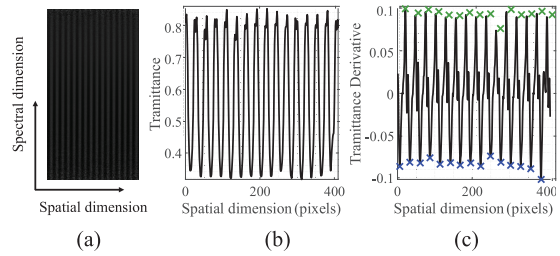


Fig. 2. (a) One-line HS image of the micrometer ruler using System A at 20×. (b) Profile resulting from the spectral average of the one-line HS image. (c) Spatial resolution empirical calculation using the first derivative where green and blue crosses indicate maxima and minima, respectively.

D. Spatial Repeatability

Spatial repeatability can be measured as a metric of a system's ability to produce consistent results under comparable external conditions [35], i.e., the capacity to obtain two equal HS images given the same illumination conditions. Following the methodology developed by Peleg et al. [36], the relative difference (RD) percentage determines the spatial repeatability of an HS system in a particular wavelength. It can be calculated following (5), where "x_i" and "y_i" are the homologous pixel values of the two images to be compared, "i" is the pixel number, and "P" is the total number of pixels in each image [37]. For the RD calculation, images of the micrometer ruler or grid are obtained in a short time span, minimizing the environmental influence over them

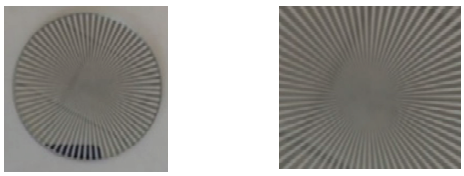
$$\mathrm{RD}_{\lambda}\,(\%) = \frac{\sum_{i=1}^{P} |x_i - y_i|}{\left(\frac{1}{P}\sum_{i=1}^{P} x_i + \frac{1}{P}\sum_{i=1}^{P} y_i\right)/2} \cdot 100\%. \tag{5}$$

E. Spatial Resolution

Spatial resolution refers to the ability of an imaging modality to differentiate two adjacent structures as being distinct from one another, a crucial characteristic of imaging systems [38]. The bigger the magnification power, the higher the spatial resolution thus enhancing the ability to distinguish smaller objects. The spatial resolution of the HS microscopic systems can be evaluated theoretically and empirically. The theoretical computation of the spatial resolution (S_R), as shown in (6), considers variables such as pixel size of the sensor (P_S) and magnification (M):

$$S_R = \frac{P_S}{M}. \tag{6}$$

The empirical determination of the pixel size, following the methodology designed by Ortega et al. [22], is performed over a one-line image of a certified micrometer ruler [Fig. 2(a)]. The profile at each wavelength [Fig. 2(b)] is derived to find the local maxima and minima [Fig. 2(c)]. The average spatial distance between lines of the target (0.01 mm in the case of Fig. 2) is divided by the mean distance between local maxima and minima (in pixels), thus empirically determining the size of a single pixel.



(a) (b)

Fig. 3. (a) Monochromatic image of an HS image of the harmonic modulated Siemens star target and (b) its center showing the aliasing region, where MTF10 is reached [38].

F. Spatial Frequency Response

Spatial resolution is calculated to determine the smallest size of an object that can be detected by the acquisition system at different magnification powers since aliasing may affect resolution. The spatial frequency response (SFR) of a digital imaging system describes its ability to maintain high contrast along certain spatial frequencies. SFR varies with wavelength, influencing spatial resolution across the spectrum, due to diffraction limits (light diffraction is more pronounced at longer wavelengths), sensor sensitivity (sensors may exhibit lower sensitivity at the extremes of the spectrum), and chromatic aberration (optical lenses can cause different wavelengths to focus on slightly different planes). In HSI, an instrument with a high SFR across all wavelengths captured preserves fine spatial details, enhancing spectral accuracy for distinguishing similar materials. This reduces artifacts and supports detailed multiscale analysis [39].

Loebich et al. [38] described a methodology to measure the SFR of a digital camera capturing a Siemens star [Fig. 3(a)]. These are made of sinusoidal oscillations in a polar coordinate system (given by several cycles) so that the spatial frequency decreases for concentric circles of larger radius (5) [38], [39], [40], [41], [42], [43], [44].

The range of spatial frequencies being captured over a monochromatic image is given by the size of the Siemens star, where the Nyquist frequency is the maximum frequency, the sensor is able to capture [42]. For every other spatial frequency, the modulation transfer function (MTF) is measured as a general term of the SFR (6). $I_{\max}$ and $I_{\min}$ refer to, respectively, the maximum and minimum digital values obtained at that specific radius profile [42]. The resolution limit is determined using the r10 criterion [47], which defines the resolution limit as the frequency at which MTF achieves a contrast value of 10%, referred to as MTF10 [38] [Fig. 3(b)]. Emphasis was placed on studying the SFR over wavelengths since diffraction-limited spatial resolution is inversely proportional to wavelength, a relationship first explained by Abbe [46]

$$f(\text{lp/mm}) = \frac{\text{No of cylces}}{2\pi r} \tag{7}$$

$$\text{MTF} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \tag{8}$$

G. Flat-Field Correction

Flat-field correction is the process employed to correct the variations in the measured radiance values in an HS image caused by the sensor, environmental conditions, and other factors [47]. This step is crucial before continuing with the analysis of the HS data. It involves converting raw DNs, obtained directly from the sensor, into actual radiation intensity [48] or true reflectance values [49], [50], [51]. The process of computing a calibrated HS frame (CI) from a raw HS frame (R) is well established in [22] (9). It is needed a DC capture, explained in Section II-B, and a white reference (WR), the obtention of which depends on the illumination mode. In the transmittance mode, an area of the slide with no sample is usually employed, while in the reflectance mode, a diffuse reflectance standard is used [52]. These standards are frequently constructed with a matte Lambertian reflecting surface, which means that the reflected light has nearly equal intensity in all directions. The total reflection integrated in all directions should be close to 100% [53]

$$\text{CI}(\%) = \frac{\text{R} - \text{DC}}{\text{WR} - \text{DC}} \cdot 100\%. \tag{9}$$

H. Tone Transfer

The calibration process makes an HS image independent of the environment light. However, the so-called tone transfer, the relationship between the optical input and digital signal output, is usually nonlinear [54]. The tone transfer metric examines the significance of the tone transfer between the captured scene and the resulting HS image, allowing to characterize the equilibrium needed among the various spectral bands [55]. Understanding tone transfer is crucial for ensuring accurate data interpretation and improving the reliability of HSI applications across diverse fields, as it directly affects how well the sensor converts light into usable spectral data.

This metric is calculated by capturing the optoelectronic conversion function (OECF) target, which provides a reference on how a sensor converts the illumination it receives into DNs, following ISO-14524 [56]. The mean reflectance of each gray patch at the monochromatic image is calculated and compared to the reference provided by Edmund Optics (Barrington, NJ, USA).

I. Spectral Sensitivity

It is essential to ensure the reliability of HS data by accurately characterizing the spectral response of an HSI system [57]. Spectral sensitivity is the ability of the sensor to detect light as a function of its signal frequency. Typically, the sensitivity of HS camera channels fluctuates through the different wavelengths due to the spectral responsivities of the HS sensor and the nonuniform output from diffractive or filtering elements. Characterizing these irregularities makes it possible to estimate their impact on the captured spectral range [57].

Spectral sensitivity has been analyzed in the state of the art by assessing captured spectra against a ground truth spectrum obtained with a spectrometer [58] or evaluating a captured wavelength calibration standard against a known spectrum [59], [60]. Usually, the latter is preferred for its simplicity. The National Institute of Standards and Technology (NIST) traceable wavelength calibration standards can be employed

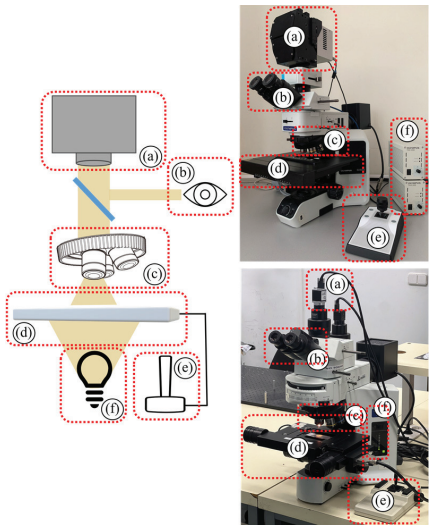


Fig. 4. Diagram of HS microscopic systems (left) and HS microscopic systems employed in this work: System A (top right) and System B (bottom right). (a) HS camera, (b) binoculars, (c) objective lenses, (d) specimen stage, (e) joystick for spatial scanning, and (f) halogen light source.

since they combine rare earth oxides producing very specific absorption peaks, suitable for calibration purposes.

Once the wavelength calibration standard has been captured, the spectral correlation measure (SCM), designed by van der Meero and Bakker [61], quantitatively assesses the spectral quality of the HS image with respect to the reference provided by the manufacturer [62]. SCM is calculated following (10), where n is the number of spectral bands being compared, r is the reference spectral signature given by the manufacturer, and t is the test spectra captured with each HS microscopic system and magnification. The resulting correlation value represents, to some degree, variations in brightness and shape within the spectra

$$\mathrm{SCM} = \frac{n\sum_{i=1}^{n} t_i r_i - \sum_{i=1}^{n} t_i \sum_{i=1}^{n} r_i}{\sqrt{\left(n\sum_{i=1}^{n} t_i^2 - \left(\sum_{i=1}^{n} t_i\right)^2\right)\left(n\sum_{i=1}^{n} r_i^2 - \left(\sum_{i=1}^{n} r_i\right)^2\right)}}. \tag{10}$$

III. Materials and Methods

A. HS Microscopic Systems

To test the proposed roadmap methodology, two HS microscopic systems were set up by coupling HS cameras to two different commercial microscopes via a standard C-Mount. The optical path of the systems starts at the bright-field halogen lamp [Fig. 4(f)] for reflected (top) or transmitted (bottom) illumination. In the transmittance mode, the light travels through the specimen on the stage [Fig. 4(d)] through the selected objective lens [Fig. 4(c)]. In this mode, low magnification enlarges the field of view, increasing the amount of incoming light captured.

In the reflectance mode, the light source is positioned and concentrated above the sample by the objective lens, where it reflects from the sample surface back to the lens. Finally, transmittance/reflectance light is captured by their respective HS camera [Fig. 4(a)]. In this work, System A employs a push-broom HS camera, while System B uses a linescan wedge sensor [63], [64]. Thus, both HS microscopic systems need spatial scanning to generate an HS cube. Custom software was developed to synchronize the movement of the scanning platform with respect to the HS camera frame rate (see Section II-E). Binoculars [Fig. 4(b)] and joystick [Fig. 4(e)] help the operator to visually investigate the sample and move around it to position the field of view in the region of interest to be captured. Table I details the characteristics of the optical and electronic parts of each HS camera and commercial microscope.

TABLE I
HS Microscopic Systems Employed in This Work

		System A	*System B*
Microscope	Model	OLYMPUS BX-53 (Olympus, Tokyo, Japan)	OLYMPUS BX-51 (Olympus, Tokyo, Japan)
	Lenses	LMPLN-IR (5×, 10×), LCPLN-IR (20×, 50×)	UPlanSApo (4×), UPlanApo (10×), PlanApo (40×)
	Halogen lamp	100W TH4 (Olympus, Tokyo, Japan)	100W 7724 (Philips, Amsterdam, Netherlands)
	Wavelength range	400-1800 nm	400-1800 nm
	Light path	Transmittance and reflectance	
	Stage	Scanning stage SCAN 130 × 85 (Märzhäuser, Wetzlar, Germany)	Stage X-Y PRior H101BXDK (Prior Scientific Instruments, Fulbourn, Cambridge, United Kingdom)
HS Camera	Model	Hyperspec® VNIR A-Series (HeadWall Photonics, Fitchburg, MA, USA)	MQ022HG-IM-LS150-NIR (Ximea, Münster, Germany)
	HS technology	Pushbroom	Linescan wedge
	Sensor	CCD (Charge-Coupled Device)	CMOS (Complementary metal-oxide-semiconductor)
	Spectral range	400-1000 nm	470-900 nm
	No. of bands	826	150
	Slit image FWHM	2.5 nm	15.0 nm
	Spatial size of one wavelength	1×1004 pixels	2048×5 pixels
	Pixel size of the sensor	7.4 µm	5.5 µm
	ADC	12 bits	8 bits
	HS cube size	1.6 Mb/line (1×1004×826 bands)	2.5 Gb/cube (2048×1088×150 bands)

FWHM: full width at half maximum. ADC: analog-to-digital converter.

B. Characterization Targets

Here, we display the characterization targets that are needed to measure the parameters under investigation (Fig. 5). For reproducibility, Table II details the manufacturer and specifications of each target employed in the experiments.

C. Proposed Methodology

The methodology, proposed as a roadmap to design and characterize an HS microscopic system, involves assessing the

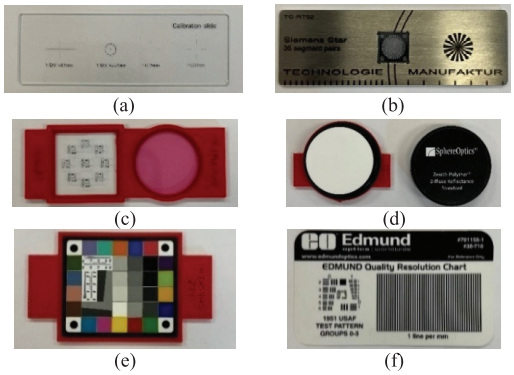


Fig. 5. Characterization targets. (a) Universal calibration slide, (b) Siemens star target, (c) transmittance United States Air Force (USAF) (left) and transmittance wavelength calibration standard (right), (d) reflectance wavelength calibration standard, (e) rez checker matte, and (f) reflectance quality resolution chart.

TABLE II
CHARACTERIZATION TARGETS EMPLOYED IN THIS WORK

Target name	Manufacturer	Specifications
Universal calibration slide (Fig. 2 (a))	Graticules Optics (Tonbridge, UK)	- Coarse grid array: 5 mm/0.5 mm square array + central 2 mm/0.25 mm square, line width 20 µm). - Fine grid array: 5 mm/0.1 mm square array + central 2 mm/0.05 mm square, line width 8 µm) -Dot array: Ø 0.25 mm dot, 0.5 mm spacing, 11×11 grid.
Siemens star target (Fig. 2b))	Edmund Optics (Barrington, NJ, USA)	- 36 sectors. - Spectral range: 200 – 2000 nm - Out diameter: 8 mm. - Tolerance: 100 nm/cm. - OD > [7, 6, 4.5, 3.6] at [400, 550, 750, 1000] nm.
Transmittance wavelength calibration standard (Fig. 2 (c))	Avian Technologies (New London, USA)	- Designation for NIST SRM-2065 standard [65]. -Spectral range: 400-2200 nm. - Focused using the transmittance USAF target.
Reflectance wavelength calibration standard (Fig. 2 (d))	Zenith Polymer® (LabSphere, North Sutton, NH, USA) [52]	- NIST wavelength calibration standard. - Reflectance: 99%, 80%, 50%, and 10%. - Spectral range: 250-2450 nm. - Diameter: 30 mm.
Rez checker (Fig. 2 (e))	Edmund Optics (Barrington, NJ, USA)	- OECF chart: 12 spectrally neutral gray patches.
Reflectance quality resolution chart (Fig. 2 (f))	Edmund Optics (Barrington, NJ, USA)	- Size: 3 ½"× 2". - USAF Pattern: Group 0, Element 1, to Group 3, Element 6. - Ronchi Ruling pattern: 1 line/mm.

OD: optical density, NIST: National Institute of Standards and Technology; USAF: United States Air Force; OECF: features an opto-electronic conversion function.

previously described parameters in a specific order (Fig. 6). First, the exposure time should be configured to maximize the DR and the SNR. Then, spatial resolution and SFR must be acknowledged to ensure that the HS image has the spatial resolution needed for a specific application. If the HS camera needs spatial scanning (e.g., push-broom or line scan), its

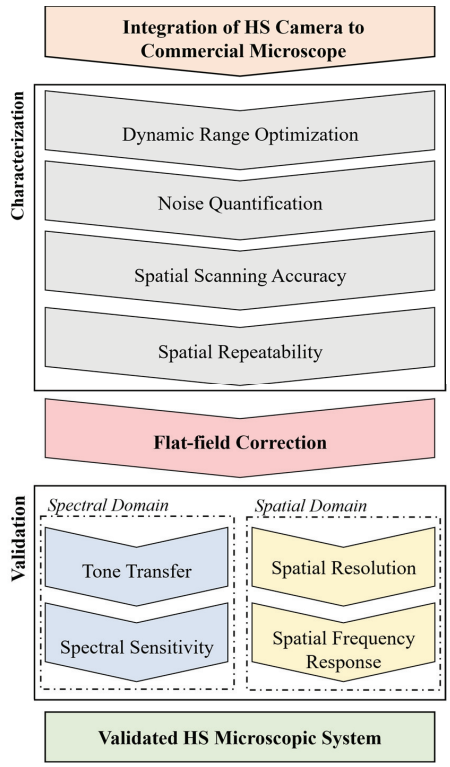


Fig. 6. Proposed roadmap for the characterization and validation of HS microscopic systems.

accuracy must be validated through eccentricity and spatial repeatability parameters. Furthermore, flat-field correction must be performed over the raw HS images; however, the tone transfer between real sample and calibrated images may not be linear. Finally, a spectral sensitivity test would describe the spectral accuracy of the HS microscopic system.

In Section IV, a brief description of each parameter to be used for the HS microscopic system design and characterization will be performed, providing the necessary background for carrying out the subsequent work.

IV. EXPERIMENTAL RESULTS

The results obtained through the previously outlined methodology are presented here. Although both systems operate at various magnifications, in order to improve clarity, only the 10× magnification results are shown in this section for comparison purposes. Data for all other magnifications are provided in the Supplementary Material and are available for comparison with other systems from the state of the art.

A. Characterization of an HS Microscopic System

1) Dynamic Range: The methodology proposed by Shaikh et al. [29] was followed to obtain the DR in both systems. In transmittance, an empty zone of the calibration target [Fig. 2(a)] is captured, increasing the exposure time

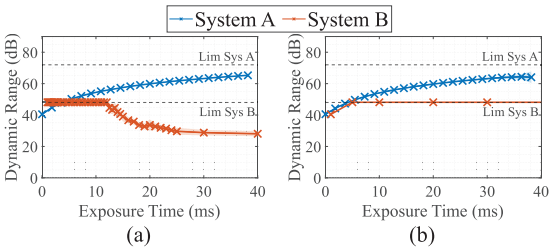


Fig. 7. DR mean and standard deviation for Systems A and B in (a) transmittance and (b) reflectance. The theoretical maximum of DR is 72 and 48 dB for Systems A and B, respectively.

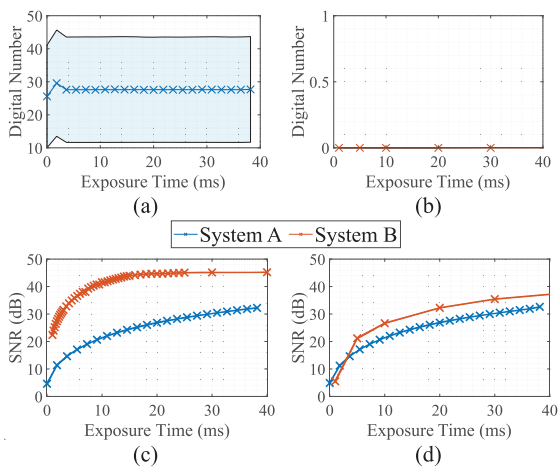


Fig. 8. DC mean and standard deviation values of (a) System A and (b) System B. The maximum DNs are 4095 and 255 for Systems A and B, respectively. SNR for Systems A and B in (c) transmittance and (d) reflectance.

to sweep from the minimum to the maximum exposure time allowed by the camera. The same procedure is followed in reflectance, but in this case, the whitest section of the ISO-21550 standard [Fig. 2(e)] is captured. The brightest pixel and darkest pixel are obtained for each capture to calculate the DR (Table SI of the Supplementary Material). In the transmittance mode, the exposure time values in System A range from 0 to 38 ms in steps of 2 ms and in System B range from 0 to 50 ms, being this range wider and following a nonlinear increment (i.e., $1.1, 1.2, 1.4, 1.6, \ldots, 30, 40$, and 50 ms). In the reflectance mode, the same values as in transmittance mode were employed for System A, and for System B, a range is between 0 and 500 ms following a nonlinear increment (i.e., $1, 5, 10, 20, \ldots, 400, 450$, and 500 ms). These differences in the range of exposure times and the intervals are due to the characteristics of both cameras (see Fig. S1 in Supplementary Material). From the captured values, the DR values are calculated, following (1), for both HS systems and for each magnification (Fig. 7).

Theoretically, the DR of System A can be calculated from the number of bits of the ADC, see Table I, using (2), resulting in 72 dB. From Fig. 7(a), it can be observed that the sensor does not saturate at 10× magnification (it does not at any magnification either, see Fig. S2), due to the low intensity of the emitted light and the relatively short maximum exposure time (40 ms). Thus, its maximum efficiency point would be the highest possible exposure time. Similarly, the maximum DR of System B was calculated to be 48 dB and was tested to quickly saturate at 10× magnification [Fig. 7(b)]. However, since in transmittance, the darkest pixel intensity gradually increases, although the brightest pixel keeps saturating, the DR of System B decreases after reaching its maximum efficiency point, with a loss of 20 dB over the entire remaining range of exposure times.

Exposure times were selected at the point of maximum efficiency: 40 ms for transmittance in System A and 9 ms in System B, with 40 ms for reflectance in both systems. However, as previously noted, System A does not exhibit a distinct maximum efficiency point, so the highest feasible exposure time was chosen for this system. It should be noted that when capturing samples in the transmittance mode, light is sent through the sample, undergoes a few scattering events, and then reaches the objective at the other end of the sample. Thus, the higher the magnification, the thinner the objective opening, and so less light reaches the sensor [Fig. S2(a) and (c)]. However, in the case of samples captured in the reflectance mode, light is sent from above and only rays with a reflectance angle of 180° are collected by the sensor. In this scenario, lower magnifications, with higher field of view, cause more dispersion and not as much light comes back to the sensor [Fig. S2(b) and (d)]. Therefore, for a given exposure time, smaller magnifications provide higher DR in transmittance than in reflectance.

2) Noise Quantification: To quantify the noise of the HS microscopic systems, the DC was captured over the aforementioned exposure times detailed in Section IV-A1 for each system. The DC signal is captured by completely blocking light from entering the HS camera, avoiding the interaction of the light with the sensor. The HS cameras of Systems A and B work with different numbers of bits (Table I), and thus, the maximum DNs are 4095 and 255, respectively. Considering these values, the results show low and constant DC values for both systems [Fig. 8(a) and (b)]. Comparing the DC standard deviation of System A with respect to System B, the HS sensor of System A has a higher standard deviation, showing a variation of ~30 DNs. This may be due to the systematic offset of 20 DNs that the manufacturer of System A HS sensor applies to avoid the expected noise [66].

Afterward, the SNR is calculated following (3), over a capture of the lighter step of the OECF target [Fig. 2(e)]. In concordance with the previously obtained values, the SNR of System A [Fig. 8(c)] follows a logarithmically increasing behavior from 5 to 35 dB. In terms of System B [Fig. 8(d)], it quickly reaches peak values of 110 and 86 dB for transmittance and reflectance modes, respectively. The behavior is similar among all the magnifications studied (Fig. S3).

3) Spatial Scanning Accuracy: Following the methodology outlined in Section II-E, the alignment of the HS system must be tested to ensure the proper configuration of the scanning parameters. Eccentricity was calculated for each

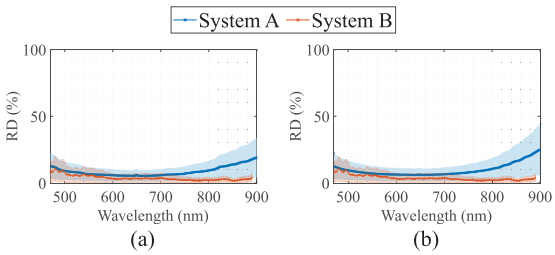


Fig. 9. Spatial repeatability mean and standard deviation for Systems A and B in (a) transmittance and (b) reflectance.

HS microscopic system over all spectral bands, computing the mean and standard deviation values using each available magnification. The obtained values are close to zero (0.037 ± 0.041 for System A and 0.027 ± 0.019 for System B), having System A 27% better eccentricity than System B at $10\times$ magnification. Since both systems produce a mean eccentricity, over all magnifications, of 0.04 and 0.05 for Systems A and B, respectively, it can be established that the platform scanning speed is adequately characterized for the optimal exposure times of each magnification and system configuration (determined in Section IV-A1).

4) Spatial Repeatability: As indicated in Section II-D, the spatial repeatability of the systems has been obtained as the RD of two images of the micrometer ruler. Spatial repeatability results show that transmittance and reflectance have similar mean and standard deviation RD values (Fig. 9). These results are consistent with the SNR values obtained in Section IV-A2 and with the results obtained by Fabelo et al. [37].

When evaluating repeatability across wavelengths, the central spectral bands consistently exhibit better RD performance than those at the extremes due to the quantum efficiency of the sensors. This trend is more evident in System A, which uses a grating to disperse white light, resulting in reduced sensitivity at the spectral limits, which affects the overall performance and reliability of the data in those regions. In contrast, System B, with its filter-based sensor, demonstrates greater consistency across the entire wavelength range. This pattern is persistent when measuring spatial repeatability with other magnification lenses (Fig. S4).

B. Flat-Field Correction

Following the methodology described in Section II-G, an HS image of a transmittance wavelength calibration standard [Fig. 5(c)], composed of 100 lines, was captured in both systems and calibrated using (9). The calibration process is repeated for each raw frame, and finally, to obtain an HS image, a spatial stitch is performed for each frame according to the requirements of the HS technology employed. In addition, only in System B, due to the manufacturer's requirements, it is necessary to apply a spectral correction matrix to sort the bands and obtain the spectral information correctly.

Fig. 10 shows how raw values are different for each system. The radiance range depends on the number of bits of the ADC of the sensor (see Table I for the actual values).

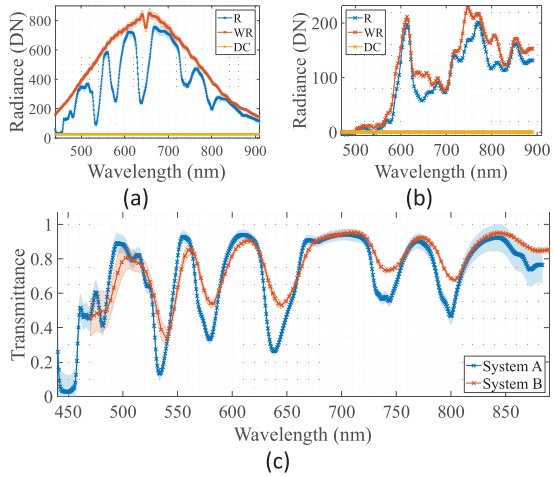


Fig. 10. Mean and standard deviation of the spectral signatures extracted from the raw image (R), WR, and DC of (a) System A and (b) System B. The maximum DNs are 4095 and 255 for Systems A and B, respectively. (c) Calibrated data of both systems. Following the recommendation from a previous study [37], noisy bands have been removed in System A.

However, after calibration, spectral signatures are normalized between 0 and 1, and different HS microscopic system configurations obtain similar spectral signatures.

To assess the accuracy of the flat-field correction, Systems A and B calibrated spectra were compared between them. First, they were interpolated to get the same spectral bands and then correlated using (10), obtaining a value close to one (SCM = 0.87 ± 0.02). It is worth noticing that System A, using a push-broom camera, requires a straightforward process to conform an HS cube by stitching spatial lines sequentially. On the other hand, System B utilizes linescan wedge technology, involving both spatial scanning and spectral scanning. Thus, achieving the final HS image with System B requires careful manipulation of the lines to account for both spatial information and spectral information. The comparison of both systems spectra with respect to the reference will be shown in Section IV-C4.

C. Validation of an HS Microscopic System

1) Spatial Resolution: To assess the spatial resolution of the HS systems, a micrometer ruler [Fig. 2(a)] was imaged 100 times. At $10\times$ magnification, the empirical spatial resolution (0.739 ± 0.001 μm for System A and 0.558 ± 0.002 μm for System B) closely matches the theoretical values (0.74 μm for System A and 0.55 μm for System B). These results yield a mean RD of 0.13% for System A and 1.45% for System B. When comparing both setups, System B, with a pixel size 24.29% smaller than System A, demonstrates superior spatial resolution, consistent with the sensor pixel sizes of each system (Table I).

2) Spatial Frequency Response: To determine the Nyquist resolution, HS images of the Siemens star target were converted to monochromatic images (by averaging all bands) and divided into eight segments following the methodology

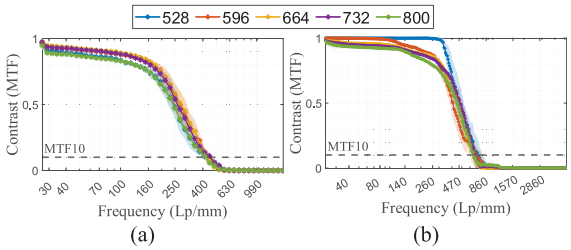


Fig. 11. MTF mean and standard deviation at 10× magnification across different spatial frequencies for (a) System A and (b) System B.

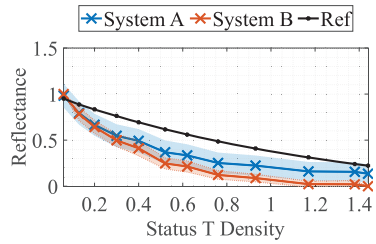


Fig. 12. Mean and standard deviation of the OECF at 10× magnification of Systems A and B. Reference (Ref) is displayed for comparison.

employed in [38]. Then, MTF10 was computed in each segment, following (6), to evaluate the SFR at different regions of the image. At 10× magnification, the high-resolution slide target contains greater Nyquist frequencies (15,485.35 lp/mm for System A and 20,834.83 lp/mm for System B) than those empirically achieved by the microscopes under study (371.02 ± 13.03 lp/mm for System A and 632.29 ± 38.05 lp/mm for System B). It can be determined that System B has a 70% higher limiting frequency than System A, being better at detecting two adjacent objects. These results are consistent with the spatial resolutions of Systems A and B at this magnification.

Furthermore, the SFR was evaluated for each system at different wavelengths (Fig. 11). In this case, each single-band image was also divided into eight segments to check the SFR at different regions of the images. The mean MTF values at each configuration follow the same trend, starting at maximum contrast where all white and dark lines are distinguishable ($I_{max} = 1$, $I_{min} = 0$, and MTF = 1), to the point of no contrast, where black and white lines are mixed into gray ($I_{max} = 0.5$, $I_{min} = 0.5$, and MTF = 0). Consistent results showed that the lower the wavelength, the higher the MTF10 (Fig. S5), following Abbe's approximation [46] of the diffraction limit.

However, no linear relationship was found between the limiting resolutions at different wavelengths in any of the systems. For example, a decrease from 800 to 528 nm, which represents a reduction of 34% in wavelength, produced an increase of the limiting spatial frequency of 9% at System A and 44% at System B. The discussion of the crucial balance between resolution and wavelength was also described by the Rayleigh criterion equation ($R = 0.61\lambda/\text{NA}$) [67], [68]. For a given numerical aperture (NA), lower wavelengths (λ) provide a higher resolution, which is characterized by a lower value of the minimum resolvable distance (R), providing the instrument's capability to discern closely spaced objects. However, some limitations, such as the noise produced by the systems and the uncertainty of the SFR methodology [53], provide nonlinear results, showing the necessity of further investigations in this field.

3) Tone Transfer: Although flat-field correction helps to standardize the HS image with respect to the illumination and sensor employed, several gray tones absorb differently onto the sensor (see Section II-H). The mean and standard deviation reflectance was calculated for each gray step of the OECF target [Fig. 2(e)]. Results show, far from a linear relationship,

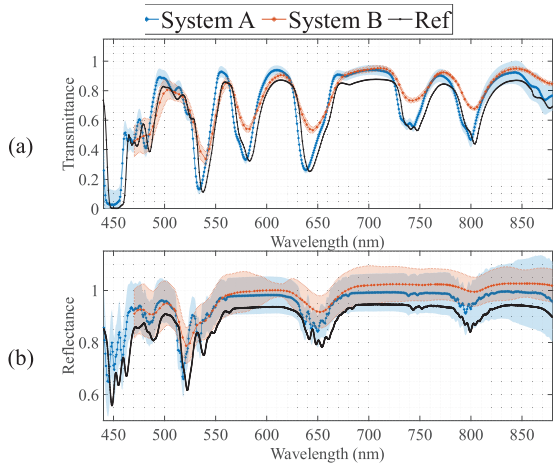


Fig. 13. Mean and standard deviation of Systems A and B signatures from the NIST wavelength calibration standard for (a) transmittance and (b) reflectance.

an exponential decay of the reflectance with respect to the optical density of the sample (Fig. 12). Root-mean-square error (RMSE) was calculated between Systems A and B with respect to the reference, providing close to zero values (0.201 ± 0.063 for System A and 0.275 ± 0.015 for System B). However, while System B deviates more from the reference, reaching a closely zero reflectance value for a status T density of 1.38, System A has greater differences between the obtained mean reflectances (Fig. S6).

4) Spectra Sensitivity: The last parameter being tested is the spectra sensitivity of the HS microscopic systems. Following Section II-I, 100 lines of the transmittance and reflectance NIST wavelength calibration standards [Fig. 2(c) and (d), respectively] were captured at transmittance and reflectance (Fig. 13). To measure the spectra sensitivity of the HS microscopic systems, the obtained spectral signatures were compared, using the SCM metric, to the NIST reference provided by the manufacturer (transmittance: 0.876 ± 0.008 for System A and 0.911 ± 0.024 for System B; reflectance: 0.590 ± 0.074 for System A and 0.771 ± 0.096 for System B).

As presented in Section IV-A1, transmittance captures tend to represent the reference more accurately than reflectance captures and, thus, their smaller standard deviation. At 10×,

System B provides better SCM results than System A, 4% and 21% for transmittance and reflectance, respectively. At other magnifications, System B also presents higher SCM values, ranging from 0.5 to 0.9 (Fig. S7).

V. Discussion

Understanding the performance parameters of a microscopic HS system is essential for ensuring accurate data capture and analysis. To support future developers and users/operators of HS microscopic systems, we offer practical suggestions for implementing the methodology proposed in this article, which should be followed in the given order.

1) Examine the DR of the microscopic HS system to set its optimal exposure time. Transmittance measurements generally offer more DR than reflectance ones.
2) Assess DC and SNR to know when the captures are reliable. Some sensors have a constant DC settled by the manufacturer (e.g., for System A is 20).
3) Evaluate spatial resolution, especially when differentiating small-sized spatial features, such as identifying cells on a histology slide, where achieving high spatial resolution is essential. Use a micrometer ruler, however, there may be a misidentification of spatial frequencies due to aliasing. Thus, assessing SFR is essential, particularly Nyquist resolution, is important. Lower wavelengths yield better SFR, following the essential connection between resolution and wavelength, as outlined by the Rayleigh criterion. To understand the limitations of an HSI system in terms of spatial frequency, measurements should be conducted at longer wavelengths.
4) Determine the appropriate platform velocity for HS cameras requiring spatial scanning, once the optimal exposure time and spatial resolution have already been determined. To check proper alignment, circles should be captured, and their eccentricity checked for each magnification. This step is not necessary for those HS microscopic systems that do not require spatial scanning (i.e., spectral scanning and snapshots HS systems).
5) Identify spatial repeatability using consecutively captured HS cubes to evaluate the robustness of the microscopic HS system. Most probably it experiences a decline at extreme bands, where the SNR is usually lower due to the quantum efficiency of the sensor.
6) Perform flat-field correction once HS cubes can be properly captured. This step serves to standardize HS captures by mitigating the impact of variations in the surrounding light environment. Afterward, the image should be assembled following its HSI technologies (i.e., push-broom or linescan wedge).
7) Analyze tone transfer, as flat-field correction standardizes HS captures but does not address reflectance variation. Typically, the analysis reveals an exponential decay of the reflectance with respect to optical density.
8) Quantify spectral sensitivity, which becomes essential when analyzing spectral properties of the materials under investigation and aiming to identify specific absorption peaks. While HS microscopic systems typically offer strong spatial sensitivities, fluctuations in SNR have the potential to cause specific spectral signatures to deviate more significantly from the established reference standard. Systematically identify and quantify these fluctuations to correct and enhance spectral analysis reliability.

TABLE III
Summary of Best-Performing HS Microscopic System for Each Parameter Studied

Parameter	Metric	*System A*		*System B*	
		T	**R**	**T**	**R**
Dynamic range	Exposure time (ms)	40	40	9	40
	Dynamic range (dB)	65.3 ± 0.1	64 ± 4	48.1± 0.1	48.1 ± 0.1
Noise quantification	SNR (dB)	32.2 ± 0.1	32.6 ± 0.1	103.97 ± 1	93.2 ± 20
Spatial scanning accuracy	Eccentricity	0.04 ± 0.04	N/A	0.03 ± 0.02	N/A
Spatial repeatability	Relative difference (%)	14 ± 8	16 ± 9	4 ± 3	4 ± 2
Spatial resolution	Spatial resolution (μm/pixel)	0.739 ±0.001	N/A	0.558 ± 0.002	N/A
Spatial frequency response	Modulation transfer function (lp/mm)	370 ± 10	N/A	630 ± 40	N/A
Tone transfer	Mean reflectance	N/A	0.20 ± 0.06	N/A	0.28 ± 0.02
Spectral sensitivity	Spectral correlation measure	0.88 ± 0.01	0.59 ± 0.07	0.91 ±0.02	0.77 ± 0.10

*T: Transmittance. R: Reflectance. N/A: Not applicable.

Moreover, Table III provides a summary of the values obtained at 10× for each HS microscopic system, determining which one better satisfies each characterization and validation parameter (Table SI completes these data by showing results at all possible magnifications for each system). This information may serve as a valuable reference for future developers and users/operators of HS microscopic systems, enabling them to compare their results with those presented in this article. It must be noted that spatial scanning, accuracy, spatial resolution, and SFR were not evaluated in the reflectance mode due to the unavailability of specific targets. Similarly, tone transfer in the transmittance mode could not be tested.

System B employs a linescan sensor composed of spectral filters arranged side by side, requiring scanning of the entire sensor over the sample to capture material data. This design excels in spatial parameters, capturing the 2-D scene simultaneously and enhancing spatial resolution and frequency response. Additionally, System B offers a superior DR, making it particularly suitable for applications demanding high spatial precision and DR, such as detailed analysis of microscopic structures. However, the complexity of system assembly can be seen as a drawback.

Conversely, System A uses a push-broom HS camera that diffracts a ray of light from a spatial line to capture all its spectral bands simultaneously, offering a plug-and-play solution. This characteristic simplifies the system setup and operation, making it more accessible for various applications. System A is superior for high spectral resolution tasks, capturing

826 bands compared to the 150 bands of System B. This makes it ideal for applications that require detailed spectral information, such as identifying specific chemical compositions or detecting subtle spectral features.

VI. Conclusion

The lack of a standard methodology to characterize and validate HS microscopic systems limits their transferability between different applications and research teams, leading to variability in results and reduced reproducibility. This lack of standardization also makes the formal validation of HS microscopic systems challenging. However, a methodology based on well-established procedures for characterizing the different optical and spectral properties of imaging systems could solve this issue. The main contribution of this article is to provide a structured roadmap to characterize and validate HS microscopic systems to improve repeatability and enable comparison across various systems. To the best of our knowledge, this is the first time in which a unified methodology has been proposed for this purpose, potentially serving as a foundation for incorporating HS microscopic systems into the "Standard for Characterization and Calibration of Ultraviolet through Shortwave Infrared (250–2500 nm) HSI Devices." Furthermore, the proposed roadmap has been validated through a round-robin test using two different HS microscopic systems to test its generalizability. Future research in the field would employ the proposed roadmap as a technical verification for the development of HS microscopic systems. Reporting the quality metrics of these systems would enhance the publicly available characterization information, which can improve reproducibility and standardize the technical validation of the public microscopic HS image datasets.

References

[1] S. L. Jacques, "Optical properties of biological tissues: A review," *Phys. Med. Biol.*, vol. 58, no. 11, pp. R37–R61, Jun. 2013, doi: 10.1088/0031-9155/58/11/r37.

[2] H. Grahn and P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis*. Hoboken, NJ, USA: Wiley, 2007.

[3] B. Wang et al., "The applications of hyperspectral imaging technology for agricultural products quality analysis: A review," *Food Rev. Int.*, vol. 39, no. 2, pp. 1043–1062, Feb. 2023, doi: 10.1080/87559129.2021.1929297.

[4] S.-E. Qian, "Hyperspectral satellites, evolution, and development history," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7032–7056, 2021, doi: 10.1109/JSTARS.2021.3090256.

[5] J. Yoon, "Hyperspectral imaging for clinical applications," *BioChip J.*, vol. 16, no. 1, pp. 1–12, Mar. 2022, doi: 10.1007/s13206-021-00041-0.

[6] K. S. Banu, M. Lerma, S. U. Ahmed, and J. L. Gardea-Torresdey, "Hyperspectral microscopy-applications of hyperspectral imaging techniques in different fields of science: A review of recent advances," *Appl. Spectrosc. Rev.*, vol. 59, no. 7, pp. 935–958, Aug. 2024, doi: 10.1080/05704928.2023.2270035.

[7] B. Park, T. Shin, J.-S. Cho, J.-H. Lim, and K.-J. Park, "Improving blueberry firmness classification with spectral and textural features of microstructures using hyperspectral microscope imaging and deep learning," *Postharvest Biol. Technol.*, vol. 195, Jan. 2023, Art. no. 112154, doi: 10.1016/j.postharvbio.2022.112154.

[8] A. Soni, Y. Dixit, M. M. Reis, and G. Brightwell, "Hyperspectral imaging and machine learning in food microbiology: Developments and challenges in detection of bacterial, fungal, and viral contaminants," *Comprehensive Rev. Food Sci. Food Saf.*, vol. 21, no. 4, pp. 3717–3745, Jul. 2022, doi: 10.1111/1541-4337.12983.

[9] T. Kitahashi et al., "Development of robust models for rapid classification of microplastic polymer types based on near infrared hyperspectral images," *Anal. Methods*, vol. 13, no. 19, pp. 2215–2222, May 2021, doi: 10.1039/d1ay00110h.

[10] A. Faltynkova, G. Johnsen, and M. Wagner, "Hyperspectral imaging as an emerging tool to analyze microplastics: A systematic review and recommendations for future development," *Microplastics Nanoplastics*, vol. 1, no. 1, pp. 1–19, Dec. 2021, doi: 10.1186/s43591-021-00014-y.

[11] Z. Xu, Y. Jiang, J. Ji, E. Forsberg, Y. Li, and S. He, "Classification, identification, and growth stage estimation of microalgae based on transmission hyperspectral microscopic imaging and machine learning," *Opt. Exp.*, vol. 28, no. 21, p. 30686, Oct. 2020, doi: 10.1364/oe.406036.

[12] S. Ortega, M. Halicek, H. Fabelo, G. M. Callico, and B. Fei, "Hyperspectral and multispectral imaging in digital and computational pathology: A systematic review [invited]," *Biomed. Opt. Exp.*, vol. 11, no. 6, p. 3195, Jun. 2020, doi: 10.1364/boe.386338.

[13] B. Males, L. H. Oakley, O. Cossairt, and M. Walton, "Maximizing the microscope: Instrument design and data processing strategies for hyperspectral imaging of cross-sectional cultural heritage samples," *Proc. SPIE*, vol. 11058, pp. 24–31, Jul. 2019.

[14] E. Aloupogianni et al., "Hyperspectral imaging for tumor segmentation on pigmented skin lesions," *J. Biomed. Opt.*, vol. 27, no. 10, Oct. 2022, Art. no. 106007, doi: 10.1117/1.jbo.27.10.106007.

[15] L. Ma, J. V. Little, A. Y. Chen, L. Myers, B. D. Sumer, and B. Fei, "Automatic detection of head and neck squamous cell carcinoma on histologic slides using hyperspectral microscopic imaging," *J. Biomed. Opt.*, vol. 27, no. 4, Apr. 2022, Art. no. 046501, doi: 10.1117/1.jbo.27.4.046501.

[16] Q. Huang, W. Li, B. Zhang, Q. Li, R. Tao, and N. H. Lovell, "Blood cell classification based on hyperspectral imaging with modulated Gabor and CNN," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 160–170, Jan. 2020, doi: 10.1109/JBHI.2019.2905623.

[17] S. Prasad and J. Chanussot, *Hyperspectral Image Analysis*. Cham, Switzerland: Springer, 2020.

[18] A. Atla, R. Tada, V. Sheng, and N. Singireddy, "Sensitivity of different machine learning algorithms to noise," *J. Comput. Sci. Colleges*, vol. 26, pp. 96–103, Jun. 2011.

[19] A. Saseendran, L. Setia, V. Chhabria, D. Chakraborty, and A. B. Roy, "Impact of noise in dataset on machine learning algorithms," *Mach. Learn. Res.*, vol. 2019, pp. 1–8, Jul. 2019.

[20] R. Pillay, M. Picollo, J. Y. Hardeberg, and S. George, "Evaluation of the data quality from a round-robin test of hyperspectral imaging systems," *Sensors*, vol. 20, no. 14, p. 3812, Jul. 2020, doi: 10.3390/s20143812.

[21] H. Pu, L. Lin, and D.-W. Sun, "Principles of hyperspectral microscope imaging techniques and their applications in food quality and safety detection: A review," *Comprehensive Rev. Food Sci. Food Saf.*, vol. 18, no. 4, pp. 853–866, Jul. 2019, doi: 10.1111/1541-4337.12432.

[22] S. Ortega et al., "Hyperspectral push-broom microscope development and characterization," *IEEE Access*, vol. 7, pp. 122473–122491, 2019, doi: 10.1109/ACCESS.2019.2937729.

[23] J. Stergar, R. Hren, and M. Milanič, "Design and validation of a custom-made hyperspectral microscope imaging system for biomedical applications," *Sensors*, vol. 23, no. 5, p. 2374, Feb. 2023, doi: 10.3390/s23052374.

[24] A. V. Paterova, S. M. Maniam, H. Yang, G. Grenci, and L. A. Krivitsky, "Hyperspectral infrared microscopy with visible light," *Sci. Adv.*, vol. 6, p. 44, Oct. 2020, doi: 10.1126/sciadv.abd0460.

[25] *GRSS/SC—Standards Committee Standard for Characterization and Calibration of Ultraviolet Through Shortwave Infrared (250 Nm To 2500 Nm) Hyperspectral Imaging Devices*, Standard P4001, 2018.

[26] M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic range improvement through multiple exposures," in *Proc. Int. Conf. Image Process.*, Aug. 1999, pp. 159–163.

[27] Photography Electronic Scanners for Photographic Images Dynamic Range Measurements, ISO Standard 21550:2004, 2004.

[28] C. Pan et al., "Analysis of OMPS in-flight CCD dark current degradation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 1966–1969.

[29] M. S. Shaikh, K. Jaferzadeh, and B. Thörnberg, "Extending effective dynamic range of hyperspectral line cameras for short wave infrared imaging," *Sensors*, vol. 22, no. 5, p. 1817, Feb. 2022, doi: 10.3390/s22051817.

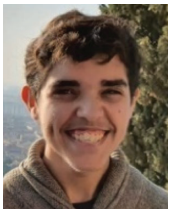
[30] D. H. Foster and K. Amano, "Hyperspectral imaging in color vision research: Tutorial," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 36, no. 4, p. 606, Apr. 2019, doi: 10.1364/josaa.36.000606.

[31] T. Zhang, Y. Fu, and J. Zhang, "Guided hyperspectral image denoising with realistic data," *Int. J. Comput. Vis.*, vol. 130, no. 11, pp. 2885–2901, Nov. 2022, doi: 10.1007/s11263-022-01660-2.
[32] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, "Noise reduction in hyperspectral imagery: Overview and application," *Remote Sens.*, vol. 10, no. 3, p. 482, Mar. 2018, doi: 10.3390/rs10030482.
[33] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 50–63, Sep. 2013, doi: 10.1016/j.isprsjprs.2013.06.001.
[34] D. Cohen, *Precalculus: With Unit Circle Trigonometry*, 4th ed., Boston, MA, USA: Cengage Learning, 2005.
[35] A. S. Lister, "7 validation of HPLC methods in pharmaceutical analysis," in *Handbook of Pharmaceutical Analysis By HPLC*, vol. 6. Amsterdam, The Netherlands: Elsevier, 2005, pp. 191–217.
[36] K. Peleg, G. L. Anderson, and C. Yang, "Repeatability of hyperspectral imaging systems—Quantification and improvement," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 115–139, Jan. 2005, doi: 10.1080/01431160412331291288.
[37] H. Fabelo et al., "In-vivo hyperspectral human brain image database for brain cancer detection," *IEEE Access*, vol. 7, pp. 39098–39116, 2019, doi: 10.1109/ACCESS.2019.2904788.
[38] C. Loebich, D. Wueller, B. Klingen, and A. Jaeger, "Digital camera resolution measurement using sinusoidal Siemens stars," *Proc. SPIE*, vol. 6502, pp. 214–224, Feb. 2007.
[39] U. Artmann, "Quantify aliasing a new approach to make resolution measurement more robust," *Electron. Imag.*, vol. 31, no. 10, pp. 320–326, Jan. 2019, doi: 10.2352/issn.2470-1173.2019.10.iqsp-320.
[40] S. J. Yoon, P. Bajcsy, M. Litorja, and J. J. Filliben, "Evaluation of lateral resolution of light field cameras," *Opt. Eng.*, vol. 57, no. 9, p. 1, Sep. 2018, doi: 10.1117/1.oe.57.9.093101.
[41] G. C. Birch and J. C. Griffin, "Sinusoidal Siemens star spatial frequency response measurement errors due to misidentified target centers," *Opt. Eng.*, vol. 54, no. 7, Jul. 2015, Art. no. 074104, doi: 10.1117/1.oe.54.7.074104.
[42] U. Artmann, "Linearization and normalization in spatial frequency response measurement," *Electron. Imag.*, vol. 28, no. 13, pp. 1–6, Feb. 2016, doi: 10.2352/issn.2470-1173.2016.13.iqsp-011.
[43] Photography Electronic Still-Picture Cameras Resolution Measurements, ISO Standard 12233, 2000.
[44] T. Osborne, V. Ramachandra, K. Atanassov, and S. Goma, "Introducing the cut-out star target to evaluate the resolution performance of 3D structured-light systems," *Proc. SPIE*, vol. 8650, pp. 213–219, Mar. 2013, Art. no. 86500P.
[45] D. Williams, "R10 criteria: A simple single valued MTF performance metric for reporting DSC pixel count," in *Proc. ISO TC42 Conf.*, Sweden, Sweden, Jun. 2004.
[46] E. Abbe, "Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung," *Archiv für Mikroskopische Anatomie*, vol. 9, no. 1, pp. 413–468, Dec. 1873, doi: 10.1007/bf02956173.
[47] B. Boldrini, W. Kessler, K. Rebner, and R. W. Kessler, "Hyperspectral imaging: A review of best practice, performance and pitfalls for in-line and on-line applications," *J. Near Infr. Spectrosc.*, vol. 20, no. 5, pp. 483–508, Oct. 2012, doi: 10.1255/jnirs.1003.
[48] G. Yang et al., "The DOM generation and precise radiometric calibration of a UAV-mounted miniature snapshot hyperspectral imager," *Remote Sens.*, vol. 9, no. 7, p. 642, Jun. 2017, doi: 10.3390/rs9070642.
[49] P. Geladi, J. Burger, and T. Lestander, "Hyperspectral imaging: Calibration problems and solutions," *Chemometric Intell. Lab. Syst.*, vol. 72, no. 2, pp. 209–217, Jul. 2004, doi: 10.1016/j.chemolab.2004.01.023.
[50] D. Nouri, Y. Lucas, and S. Treuillet, "Calibration and test of a hyperspectral imaging prototype for intra-operative surgical assistance," *Proc. SPIE*, vol. 8676, pp. 229–237, Mar. 2013.
[51] A. Noviyanto and W. H. Abdulla, "Segmentation and calibration of hyperspectral imaging for honey analysis," *Comput. Electron. Agricult.*, vol. 159, pp. 129–139, Apr. 2019, doi: 10.1016/j.compag.2019.02.006.
[52] (2021). *Encapsulated Gray Scale Standards—Avian Technologies*. Accessed: Jun. 3, 2021. [Online]. Available: https://aviantechnologies.com/product/encapsulated-gray-scale-standards/
[53] (2021). *Reflectance Wavelength Calibration Standards—Avian Technologies*. Accessed: Jun. 2, 2021. [Online]. Available: https://aviantechnologies.com/product/reflectance-wavelength-calibration-standards/
[54] P. D. Burns, "Tone-transfer (OECF) characteristics and spatial frequency response measurements for digital cameras and scanners," *Proc. SPIE*, vol. 5668, pp. 123–128, Jan. 2005.
[55] W. F. Hsu, K. W. Chuang, and Y. C. Hsu, "Comparisons of the camera OECF, the ISO speed, and the SFR of digital still-picture cameras," *Prco. SPIE*, vol. 4080, pp. 104–111, 2000.
[56] Photography-Electronic Still Picture Cameras-Methods for Measuring Opto-Electronic Conversion Functions (OECFs), ISO Standard 14524, 2009.
[57] O. Pekkala, T. Pulli, A. Kokka, and E. Ikonen, "Setup for characterising the spectral responsivity of Fabry–Pérot-interferometer-based hyperspectral cameras," *Metrologia*, vol. 56, no. 6, Dec. 2019, Art. no. 065005, doi: 10.1088/1681-7575/ab3fd1.
[58] L. Gao, L. Cao, Y. Zhong, and Z. Jia, "Field-based high-quality emissivity spectra measurement using a Fourier transform thermal infrared hyperspectral imager," *Remote Sens.*, vol. 13, no. 21, p. 4453, Nov. 2021, doi: 10.3390/rs13214453.
[59] R. Leon et al., "VNIR–NIR hyperspectral imaging fusion targeting intraoperative brain cancer detection," *Sci. Rep.*, vol. 11, no. 1, Oct. 2021, Art. no. 19696, doi: 10.1038/s41598-021-99220-0.
[60] S. George et al., "A study of spectral imaging acquisition and processing for cultural heritage," in *Digital Techniques for Documenting and Preserving Cultural Heritage*. Amsterdam, The Netherlands: Arc Humanities Press, 2017, doi: 10.1515/9781942401353-012.
[61] F. van der Meero and W. Bakker, "Cross correlogram spectral matching: Application to surface mineralogical mapping by using AVIRIS data from cuprite, Nevada," *Remote Sens. Environ.*, vol. 61, no. 3, pp. 371–382, Sep. 1997.
[62] F. van der Meer, "The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 8, no. 1, pp. 3–17, Jan. 2006, doi: 10.1016/j.jag.2005.06.001.
[63] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, Jan. 2014, Art. no. 010901, doi: 10.1117/1.jbo.19.1.010901.
[64] B. Fei, "Hyperspectral imaging in medical applications," in *Data Handling in Science and Technology*, vol. 32. Amsterdam, The Netherlands: Elsevier, 2020, pp. 523–565.
[65] S. J. Choquette, D. L. Duewer, L. M. Hanssen, and E. A. Early, "Standard reference material 2036 near-infrared reflection wavelength standard," *Appl. Spectrosc.*, vol. 59, no. 4, pp. 496–504, Apr. 2005, doi: 10.1366/0003702053641414.
[66] *ADIMEC-1000m Operating and Technical Manual*, Adimec Adv. Image Syst. B.V., Eindhoven, The Netherlands.
[67] S. Ram, E. S. Ward, and R. J. Ober, "Beyond Rayleigh's criterion: A resolution measure with application to single-molecule microscopy," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 12, pp. 4457–4462, Mar. 2006, doi: 10.1073/pnas.0508047103.
[68] A. Stacey and C. Pask, "Spatial-frequency response of a photoreceptor and its wavelength dependence i coherent sources," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 11, no. 4, p. 1193, Apr. 1994, doi: 10.1364/josaa.11.001193.

Laura Quintana-Quintana received the bachelor's degree in biomedical engineering from the University Carlos III of Madrid, Madrid, Spain, in June 2019, and the master's degree in applied electronics and telecommunications research from the University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain, in 2020, where she is currently pursuing the Ph.D. degree in bioengineering with the Integrated Systems Design Division, Research Institute for Applied Microelectronics.

Gonzalo Rosa Olmeda received the bachelor's degree in electronics of communications engineering and the master's degree in Internet of Things (IoT) from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2020 and 2021, respectively, where he is currently pursuing the Ph.D. degree with the Electronic and Microelectronic Design Group (GDEM), Software Technologies and Multimedia Systems for Sustainability (CITSEM) Research Center.

His research interests are related to microscopic hyperspectral imaging applied to histopathology applications and digitalization 4.0 of aquaculture facilities.



Javier Santana-Nunez received the bachelor's degree in industrial electronics and automation engineering from the University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain, in 2023, where he is currently pursuing the M.Sc. degree in intelligent systems and numerical applications applied to engineering. He developed his bachelor thesis at the Research Institute for Applied Microelectronics (IUMA), ULPGC, focusing on designing and evaluating hyperspectral image registration techniques.

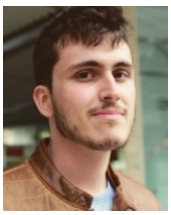
He is currently working at IUMA as a Researcher and starting his Ph.D. thesis. His research interests are related to microscopic hyperspectral imaging applied to biological sample analysis, especially for histopathology applications.

Miguel Chavarrías received the Ph.D. degree from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2017.

He is currently an Associate Professor at UPM, where he has been with the Electronics and Microelectronics Design Group (GDEM) since 2011 and the Software Technologies and Multimedia Systems for Sustainability (CITSEM) Research Center since 2012. He has authored more than 30 journal articles and conference papers. He has participated in more than 15 research projects in competitive calls and more than ten contracts with the industry. His research interests include edge computing, hyperspectral imaging, computer vision, and machine learning.

Samuel Ortega received the bachelor's degree in telecommunication engineering and the M.Sc. and Ph.D. degrees in telecommunication technologies from the University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain, in 2015, 2016, and 2021, respectively.

In 2021, he started to work as a Researcher at the Norwegian Institute of Food, Fisheries and Aquaculture Research, Tromsø, Norway. His research interests are the use of hyperspectral imaging for medical and food quality applications.



Jaime Sancho received the M.Sc. and Ph.D. degrees (cum laude) in systems and services engineering for the information society from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2018 and 2023, respectively.

He is currently a Teacher and a Research Assistant (Ayudante) at the Telematics and Electronics Engineering Department (UPM). He develops his research career at the Research Center in Software Technologies and Multimedia Systems for Sustainability (CITSEM), UPM. His main research interests include biomedical real-time systems, computer vision applied to immersive video, and GPU computing.

Dr. Sancho is a member of CITSEM.

Himar Fabelo received the master's degree in telecommunication engineering and the Ph.D. degree in telecommunication technologies from the University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain, in 2014 and 2019, respectively.

He has conducted his research activity at the Institute for Applied Microelectronics, ULPGC. Currently, he works in the STRATUM EU project as an Assistant Coordinator and a PI at the Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Las Palmas de Gran Canaria. He has published more than 50 journal articles and 60 international conference papers in the field of medical technology, particularly in the application of artificial intelligence and hyperspectral imaging for cancer diagnosis in various medical applications.

Dr. Fabelo received the Juan de La Cierva Formación Post-Doctoral Grant at FIISC in 2022.

Eduardo Juárez (Member, IEEE) received the Ph.D. degree from the École Polythecnique Fédéral de Lausanne (EPFL), Lausanne, Switzerland, in 2003.

From 1994 to 1997, he worked as a Researcher at the Digital Architecture Group, Universidad Politécnica de Madrid (UPM), Madrid, Spain, and was a Visiting Researcher at ENST, Brest, France, and the University of Pennsylvania, Philadelphia, PA, USA. From 1998 to 2000, he worked as an Assistant at the Integrated Systems Laboratory (LSI), EPFL. From 2000 to 2003, he worked as a Senior Systems Engineer at the Design Centre, Transwitch Corporation, Geneva, Switzerland, while continuing his research toward the Ph.D. degree at EPFL. In December 2004, he joined UPM as a Post-Doctoral Researcher, where he has been an Associate Professor since 2007. He has co-authored one book and authored or co-authored more than 100 articles and contributions to technical conferences. He has participated in more than 15 competitive research projects and 20 noncompetitive industrial projects. His research activity is mainly focused on hyperspectral imaging for health applications, real-time depth estimation and refinement, and heterogeneous high-performance computing.

Gustavo M. Callico (Senior Member, IEEE) received the M.S. degree in telecommunications engineering and the Ph.D. and European Doctorate degrees from the University of Las Palmas de Gran Canaria (ULPGC), Las Palmas de Gran Canaria, Spain, in 1995 and 2003, respectively.

Since 2022, he has been a Full Professor at ULPGC and has developed his research activities at the Institute of Applied Microelectronics (IUMA). He has co-authored seven book chapters and authored or co-authored more than 200 articles (83 JCR articles) and contributions to technical conferences (134 conferences). He has participated in 27 competitive research projects (six as PI) and holds two international patents. His current research areas include hyperspectral systems for cancer detection, artificial intelligence algorithms, real-time super-resolution algorithms, synthesis-based design for SOCs and circuits for multimedia processing, and video coding standards.

Chapter 3. Image Quality Optimization

Once HS microscopic systems have been properly developed and characterized, attention must shift toward the quality and fidelity of the data they produce. In this context, two critical areas emerge: optimizing data quality during acquisition and through subsequent processing. During image capture, focus is typically adjusted manually by the operator using only visual inspection. This approach often results in non-uniform image sharpness, especially given the low-light conditions inherent to HSI microscopic and the heterogeneous nature of biological specimens. Following acquisition, dedicated processing techniques are essential to mitigate noise and artifacts, thereby improving data quality and ensuring robust, accurate spectral analysis.

This chapter addresses objective **O2**, which aims to analyze factors influencing image focus, illumination, and signal quality prior to data capture, and to develop strategies and protocols for consistent, reproducible image acquisition. Here, an in-depth evaluation of the limitations of manual focusing is performed, and state-of-the-art methods are examined to achieve more uniform and reliable sharpness across the FOV. This includes assessing focus quality using no-reference image quality assessment (NR-IQA) algorithms adapted from RGB to HS data, evaluated over a generated database, *HIDFA* (Hyperspectral Imaging Dataset for Focus Assessment), containing 125 reference frames captured at 11 working distances (1,375 HSI cubes in total). Results indicated that while some NR-IQA algorithms showed fluctuating responses across different spectral bands, others provided consistent scores regardless of wavelength variations. This consistency is particularly important in microscopic HSI, where all spectral bands correspond to the same focal plane of the microscope. Therefore, the most robust

algorithms were identified as those with stable performance across the spectrum. These findings lay the groundwork for future applications in focus correction, such as automated selection of sharpest regions in z-stacking procedures. These findings underscore both the relevance and complexity of adapting NR-IQA methods to HS imaging.

Preliminary findings of this work were presented in the conference paper (**C4**) titled **"Blur-Specific No-Reference Image Quality Assessment for Microscopic Hyperspectral Image Focus Quantification"** presented in the *11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)* held in Amsterdam in 2021. It introduced the challenges related to image sharpness in HSI microscopic. A more detailed investigation was carried out in the journal article (**J2**) titled **"Blur-specific Image Quality Assessment of Microscopic Hyperspectral Images"** and published in *Optics Express*. Together with this work was published a public database (HIDFA) for focus quantification so other researchers can benefit from it.

The chapter also addresses objective **O3**, which involves testing several image processing techniques to improve HS microscopic data quality following acquisition. Specifically, it focuses on assessing effective methods for noise reduction, artifact correction, and spectral enhancement, thereby improving spectral fidelity and supporting accurate analysis. A series of processing strategies, including smoothing filters, standard normal variate (SNV) normalization, and spectral derivative computations, were systematically benchmarked using root mean square error (RMSE) against a known wavelength reference standard. Processing efficiency was also considered to balance quality improvement with practical applicability. Results demonstrated that individual methods such as moving average smoothing, SNV, and first spectral derivatives significantly enhanced data quality, with combined application producing even greater improvements. Relevant contributions were presented in the conference paper (C1) titled **"Assessing Processing Strategies on Data from Medical Hyperspectral Acquisition Systems"** presented in the *27th*

Euromicro Conference on Digital System Design (DSD) held in Paris in 2024.

Altogether, this chapter provides a comprehensive framework for enhancing data fidelity in HS microscopic imaging. By addressing both acquisition-time and post-acquisition challenges, it contributes to the development of robust methodologies that promote more accurate, consistent, and interpretable results in biomedical imaging and beyond.

J2. Blur-Specific Image Quality Assessment of Microscopic Hyperspectral Images

Title:	Blur-Specific Image Quality Assessment of Microscopic Hyperspectral Images		
Authors:	**Laura Quintana-Quintana**, Samuel Ortega, Himar Fabelo, Francisco J. Balea-Fernández, and Gustavo M. Callico		
Journal:	Optics Express		
Publisher:	Optica Publishing Group		
Date: 2023	**Vol.**: 31	**Pages**: 12261-12279	
doi:	https://doi.org/10.1364/OE.476949		
JIF Quartile:	Q2 in Optics	**IF:**	3.2
Open Access, Open Data			

Research Article Vol. 31, No. 8 / 10 Apr 2023 / Optics Express 12261
Optics EXPRESS

Blur-specific image quality assessment of microscopic hyperspectral images

Laura Quintana-Quintana,[1] Samuel Ortega,[2,1,*] Himar Fabelo,[1,3] Francisco J. Balea-Fernández,[1,4] and Gustavo M. Callico[1]

[1]Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[2]Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima), Tromsø, Norway
[3]Fundación Instituto de Investigación Sanitaria de Canarias (FIISC), Las Palmas de Gran Canaria, Spain
[4]Department of Psychology, Sociology and Social Work, University of Las Palmas de Gran Canaria, Spain
*sortega@iuma.ulpgc.es

Abstract: Hyperspectral (HS) imaging (HSI) expands the number of channels captured within the electromagnetic spectrum with respect to regular imaging. Thus, microscopic HSI can improve cancer diagnosis by automatic classification of cells. However, homogeneous focus is difficult to achieve in such images, being the aim of this work to automatically quantify their focus for further image correction. A HS image database for focus assessment was captured. Subjective scores of image focus were obtained from 24 subjects and then correlated to state-of-the-art methods. Maximum Local Variation, Fast Image Sharpness block-based Method and Local Phase Coherence algorithms provided the best correlation results. With respect to execution time, LPC was the fastest.

1. Introduction

Traditionally, the study of histology slides in medicine is considered as the gold standard for the clinical diagnosis of cancer and performing differential diagnosis with tissue alterations (e.g., infection or inflammation) [1]. In classical histology analysis, pathologists, using standard bright-field (BF) microscopy, visually examine the sample (regularities of cell shapes, distributions, etc.). Then, they study different parameters to determine if the tissue is cancerous and if positive, evaluate its malignancy level. This procedure is broadly used in hospitals, including prostate [2] and, cancer diagnosis [3].

Nowadays, the histopathology research trend is to digitize histology slides into whole RGB (red-green-blue) slides for further computational image analysis. This approach can enhance diagnosis accuracy, making it more objective in a shorter time [4]. Conventional RGB imaging differs from hyperspectral (HS) imaging (HSI) in the big number of spectral channels (also called bands or wavelengths) which HSI can acquire in the electromagnetic spectrum (within and beyond the visual range). This technology enables the highly precise differentiation of the materials captured, and so, it can help pathologists to identify and diagnose tissue histology samples with higher precision compared to traditional RGB microscopy [5].

In previous works of this research group, HS histological images were captured for tumor cell classification [6] or to synthesize HS images from standard RGB images of normal and cancer cells using conditional generative adversarial networks (GANs) [7]. However, images may not have high quality due to several distortions, being the focus one of the most relevant parameters in HS microscopic images [8]. BF microscopy is a good technique for capturing thin materials (one layer cell cultures or thin tissue portions) on glass slides but has limitations imaging thick samples [9,10].

#476949 https://doi.org/10.1364/OE.476949
Journal © 2023 Received 4 Oct 2022; revised 26 Dec 2022; accepted 30 Jan 2023; published 29 Mar 2023

Optics EXPRESS

Z-stacking techniques are a solution to blurred images produced by imaging non-flat samples. This approach uses images captured at several working distances (having different parts of the image focused) and stacks them together into one clearly focused image [11]. For this process to be automatic, image quality assessment (IQA) must firstly be evaluated [12]. Until now, the most precise quality evaluation of an image is made by the human eye since it is the end-user of the multimedia devices. However, obtaining subjective rating by users is a high costly and inefficient task [13]. Thus, effective, and objective IQA methods are needed for automatically predict the quality of images before further processing methods are applied.

In this work, we aim to quantify the focus of microscopic images captured at different working distances. First, a background framework will be presented to deeply explain the different IQA methods for blur quantification. Then, the materials and methods employed in the experiments will be described. A microscope with HS capabilities will be used to capture a hyperspectral image dataset for focus analysis (HIDFA) [14]. Focus of the images will be evaluated using the human visual system (HVS) and state-of-the-art algorithms [13]. Concluding, both type of evaluations will be correlated to analyze and discuss the performance of the studied algorithms applied to different configurations of the HS data, such as synthetic RGB (generated from the HS images) or monochromatic images.

2. Background framework

2.1. Limitations in HS microscopic data capturing

Blurring in HS microscopic images is highly influenced by the depth of field (*DOF*) [15]. is the length between the closest and the farthest points where an object can be captured in acceptably sharp focus (Fig. 1(a)). Eq. (1) shows the parameters involved to compute the *DOF*: focal length (f), which is the distance from the center of the magnifying lens to the focal point of the sensor; working distance (*WD*), which is the distance between the magnifying lens and the surface of the specimen; f-number (N), which is the ratio of the system's focal length to the diameter of the lens aperture; and diameter of circle of confusion (c), which is the light ring formed by the light beams when not focusing on the same sport. Since c and N are intrinsic to the optical system, the blurring (Fig. 1(c)) occurs when there are changes in *WD* and/or f (Fig. 1(b) and (d)).

$$DOF = \frac{2 \cdot WD^2 \cdot N \cdot c}{f^2} \tag{1}$$

First, changes in the *WD* may occur because of the irregular nature of microscopic samples. Since the samples are not flat, different areas of the image can be focused on several *WDs* (Fig. 1(b)). Figure 1(c) shows the final image when capturing at 20× magnification a non-flat rat intestine sample using a custom designed HS microscope. Second, variations in f could happen when several wavelengths are captured. A typical optical issue, chromatic aberration, takes place when a lens cannot guide all wavelengths' rays to the same focal plane (Fig. 1(d)).

2.2. Image quality assessment (IQA)

IQA is defined as a process that systematically evaluates the quality of an image based on human quality judgments (sharpness, graininess, tone scale, and color rendition) [16]. Focused captures are crucial in most of his applications (e.g., cancer diagnosis in histological samples [17]), therefore, IQA becomes a key technique that must be investigated. IQA algorithms can be divided into:

- **Full-Reference IQA (FR-IQA)** calculates the image quality index by comparing it with a ground truth capture (e.g., image versus its compressed version). The Peak Signal to Noise Ratio (PSNR) is a technique of this type [18].

Optics EXPRESS

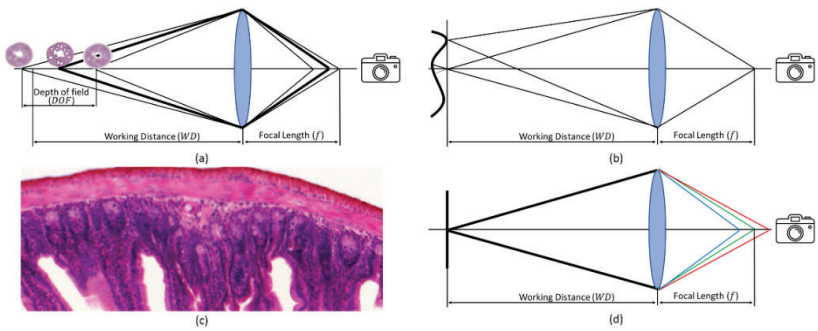


Fig. 1. Graphical representation of (a) DOF, (b) WD shift and, (d) chromatic aberration problems. (c) Example of an unfocused rat intestine (20×).

- **Reduced-Reference IQA (RR-IQA)** extracts features from an image and its reference for further calculation of the image quality (e.g., extraction of image features for easier recompositing after compression). The free-energy-based distortion metric is an algorithm of this type [19].
- **No-reference IQA (NR-IQA)** does not need any reference image for providing the quality of an image (e.g., focus of blurred image). The perceptual sharpness index (PSI) algorithm is classified in this type [20].

In this work, the interest is to enhance the quality of microscopic HS images. Sometimes, specimens are not perfectly flat, and the high gradient of the sample is bigger than the DOF offered by the imaging system. The aim of this work is to quantify the blurriness of the captured image, for a later selection of the most focused image regions. However, due to the lack of ground truth of blur images, NR-IQA is the best option, but it comes with its challenges. State-of-the-art studies indicate that the most relevant strategy for judging image quality when no reference is available, is HVS [13]. Thus, effectiveness of NR-IQA methods can be calculated by correlating their results (called Objective Scores (OS)) to the values given to the images by the HVS (called Subjective Scores (SS)). NR-IQA algorithms applied to RGB captures are further classified into learning-free (edge-free and edge-based), and learning-based (wavelet and Fourier like) methods. A general summary of the different techniques can be found in Fig. 2 [13]. Next Sections 2.4 and 2.5, will show a more detailed explanation of the state-of-the-art methods to finally select the ones employed in this work at Fig. 7.

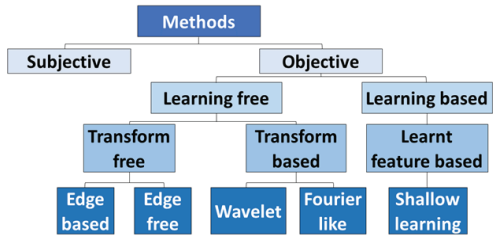


Fig. 2. Summary of the main NR-IQA methods found in the literature.

Optics EXPRESS

2.3. Database design

Before calculating image quality scores, a database (DB) showing different levels of blurriness must be created. The state-of-the-art shows that available DBs for NR-IQA are mainly based on RGB images, where blurriness is obtained synthetically by applying a filter over a reference image. These DBs employ several distortions, such as gaussian noise, chromatic aberrations, or image denoising to the reference images using several intensity levels.

Some of the most famous RGB DBs used by numerous researchers are the TID2013 [21], LIVE [22] and CSIQ [23]. Only a few multi/HS image DBs have been found, being the most relevant one the SIDQ DB [24]. A summary of the state-of-the-art in DBs for NR-IQA is shown in Table 1.

Table 1. Summary of the Characteristics of the NR-IQA State-of-the-Art DBs

	No. reference images	Spectral range (nm)	Spatial range	Types of distortion	Levels of distortion	Total No. of images
TID2013 [21]	25	RGB	Macro	24	5	3000
LIVE [22]	29	RGB	Macro	5	5	779
CSIQ [23]	30	RGB	Macro	6	4 to 5	~810
SIDQ [24]	9	410 - 1000 160 bands	Macro	5	1	45

2.4. Subjective scores (SSs)

Traditional methodologies for *SS* calculation are based on simple stimuli. Nowadays they are more complex, also registering user strategies and confidence in an opinion [25]. Some of the new parameters under study are measurement of consistency of the opinion (reliability), of correct feature (validity) and of specific empirical manipulation (sensitivity) [26]. Before planning a test to obtain subjective ratings, the panel of subjects must first be decided following the considerations shown in Table 2. Afterwards, a testing method should also be chosen. The main characteristics of the different methods can be summarized as follows in Table 3 [25]. Finally, a summary of the main characteristics of the most famous NR-IQA DBs is shown in Table 4.

Table 2. Characteristics of Subject's Panel for NR-IQA

Parameter	Range
No. of evaluators	There are several ITU (International Telecommunication Union) recommendations (Rec.), such as ITU-T Rec. P.911 [27] or ITU-R Rec. BT.500-11 [28], which describe subjective assessment methods for evaluating one-way overall audiovisual quality at multimedia applications. ITU-T Rec. P.911 states: "The possible number of subjects in a viewing and listening test [. . .] is from 6 to 40. Four is the absolute minimum for statistical reasons, while there is rarely any point in going beyond 40". Moreover, ITU-R Rec. BT.500-11 advocates at least 15 subjects.
Representative evaluation panel for the target application	Subject demographic data (e.g., sex, age, culture, and education level) are important and can affect the *SS* [29].
Expertise of the evaluation panel	Expert users are usually in better agreement than regular users. This way, a smaller number of people is needed for the experiment. However, their critical opinions may bring them to lower ratings, creating a bias in the measurements [30,31].
Testing environment	ITU recommendations remark that the experiment must be carried out under controlled lab conditions, although it may not perfectly reconstruct the target application conditions [32].

Optics EXPRESS

Table 3. Characteristics of Subjective Tests for NR-IQA

Parameter	Range
No. of images compared	It can be single, double, or multi-stimulus. Multiple stimuli may be presented simultaneously or sequentially.
No. of times an image is presented	It could be from one to multiple times.
Presence of a reference image	It can be shown or hidden, depending on whether the subjects know which one are them.
Reference image rating	Users may value the quality of the image alone, both the image and the reference, or the contrast between both.
Interactivity	One or more subjects may be rating the images at the same time.
Rating collection	Ratings may be collected per image (time-discretely) or per time interval (continuously). Recommendations given in [33] indicates that the total time a user takes to fill a subjective test should not exceed 30 minutes.
Range of test images	Usually, rating scales have five discrete levels (e.g., from "bad" to "excellent"), but language dependent scales are non-linear due to different understanding of the words. Thus, bigger ranges are also employed (e.g., 7, 9 or 11 points). Continuous scales could be employed too. However, this option may incorporate noise to the ratings due to limitation of humans to distinguish infinite levels of a certain quantity. Moreover, different users can have quite several opinions about the quality range of the images under study, so high and low anchor captures could be shown for better scale understanding [34].

Table 4. Subjective Test Features for Dataset Labeling

		TID2013 [21]	LIVE [22]	CSIQ [23]	SIDQ [24]
Subject's Panel	**No. Observers**	971	161	25	14
	Expertise in evaluation panel	Students, tutors, and researchers	Male college students	NF	22-66 years old
	Testing environment	Controlled lab setup or Internet, flexible viewing distances	Controlled lab setup, flexible viewing distances	Controlled lab setup, fixed viewing distance	Dark room
Subjective Test	**No. of images compared**	3	1	866	3
	No. of times an image is presented	9	1	1	NF
	Presence of a reference image	Yes	No	No	Yes
	Reference image rating	No	No	No	No
	Interactivity	Indifferent	Indifferent	Indifferent	Indifferent
	Rating collection	~17 mins	NF	NF	NF
	Range of test images	Pair-wise sorting with respect to the original (0-9)	Excellent, Good, Fair, Poor, Bad (1-5)	Linear displacement of the images across four calibrated LCD monitors (0-1)	Pair-wise sorting with respect to the original. Ties allowed. (0.5 to majority hit rate)

Optics EXPRESS

Before the testing, subjects are typically given training to familiarize them with the interface. Once the test has taken place, studying the results can help removing unreliable subject responses. However, we must be careful since it may also eliminate just valid opinions. After screening, the *SSs* are transformed to mean opinion scores (MOSs) [25], which inevitably represents the view of a majority. Over the years, MOS have become the most popular ratings for media quality. But we must consider that the arithmetic averaging of opinions assumes homogeneity among subjects. Thus, statistics should be applied over MOS values [25].

2.5. Objective scores (OSs)

Although we have seen that subjective quality tests are essential for image quality analysis, they involve a lot of disadvantages (e.g., they are time-consuming to perform, biased by the user, etc.). Thus, objective quality scores (*OSs*) should be employed. The following sections explain the objective models, how they work and some applications in which they are used.

2.5.1. Edge based

Cumulative probability of blur detection (CPBD) [35] is based on the study of human blur perception for varying contrast values. Edge pixels are firstly counted, if no edges are detected, the area is classified as smooth block and, edge block otherwise. For the last ones, contrast and edge width are computed. Then, the probability of finding blur at each edge is calculated and the normalized histogram of these values give the final CPBD value.

In [36], a NR-IQA metric is presented based on edge model (EMBM). The edge model can estimate width and contrast at the same time for each pixel. Thus, probability of belonging to a salient edge pixel can be determined, simulating the blur assessment performed by the HVS. The final value is obtained by adding the probability of the different pixels.

Perceptual sharpness index (PSI) [20] is calculated based on the statistical analysis of local edge gradients. It selects the biggest edges in the image through an adaptive edge selection procedure. Then, the edge widths of the selected edges are computed and the ones above the Just Noticeable Blur (JNB) width are subtracted. These steps are repeated in a block-wise way to create the local sharpness map, where the highest qth percentile average of the local sharp scores provide the final PSI score.

2.5.2. Edge free

ARISM (autoregressive-based image sharpness metric) is a blind edge free method for focus quantification. It calculates the sharpness score by examining the variance of the locally evaluated autoregressive values in a pointwise way. ARISM also takes into account the color information to assess the sharpness of the image. In [37], the authors propose ARISMc, after extending ARISM to the YIQ space.

Maximum local variation (MLV) is defined by Bahrami et al. [38] as the higher intensity contrast of a pixel with respect to its 8-neighbors. Sharpness is defined by high variations, thus the pixels' MLVs are subjected to a weighting strategy. Heavier weights are assigned to greater MLVs, making the tail end of the MLV distribution thicker, and becoming more discriminative for different blur degrees. At the end, the metric value is computed from the standard deviation of the weighted MLV distribution.

2.5.3. Wavelet

The fast image sharpness (FISH) method estimates both local and global image sharpness [39]. First, it decomposes the image in three different discrete wavelet transformations (Cohen-daubechies-fauraue 9/7 filters [6]). Next, the log-energies of each sub-band are calculated for all the transform levels. Finally, the sharpness score is calculated by a weighted average of these log-energies.

Optics EXPRESS

Hassen et al. [40] followed the idea that local phase coherence (LPC) structures are just shown in sharp edges. That way, they proposed the LPC-based sharpness index (LPC-SI). The algorithms first pass an image through 3-scale 8-orientation log-Gabor filters and then, repeat for each orientation and spatial location. They do not use block-based computation, but an efficient algorithm that largely simplifies the LPC calculation, making it easily usable in multiple applications.

2.5.4. Fourier like

Blind image blur evaluation (BIBLE) method [41] starts by calculating the gradient of an image to evaluate its shape. Afterwards, the gradient map is split into blocks and the Tchebichef moments are computed. The sum of squared non-DC moment values defines the energy of a block. At the end, the variance-normalized moment energy (simulating the HVS) provides the BIBLE index.

Vu et al. [42] proposed a spectral and spatial sharpness ($S3$) measure. The spectral measure $S1(x)$ is calculated by the reduction of high-frequency components in unfocused captures followed by a rectification using the sigmoid function to follow the HVS. The spatial measure $S2(x)$ is obtained based on the local total variation. The sharpness map $S3$ is calculated by obtaining the geometric average of $S1(x)$ and $S2(x)$ of each independent block. Finally, to consider the HVS, the final blur index is calculated as the mean of the largest 1% scores of $S3$.

2.5.5. Shallow learning

SPARISH (NR Sparse representation-based Image Sharpness) is based on sparse representations, representing signals with as few as possible significant coefficients [43]. Since a focused capture is defined by the localization of edges, a dictionary (made of edge patterns) can be used to measure the blurriness level. The sharpness score is defined as the variance-normalized energy (relative strength of blur in an image) of a group of chosen high-variance blocks. This algorithm is beneficial for real-world applications since it is not sensitive to training images and so, can evaluate different images using a universal dictionary.

3. Materials and methods

3.1. Instrumentation and sample description

The instrumentation employed in this study consists of an HS camera coupled to a conventional BF microscope (Fig. 3). The BF microscope is an Olympus BX-53 (Olympus, Tokyo, Japan). The HS camera is a Hyperspec VNIR A-Series from HeadWall Photonics (Fitchburg, MA, USA), which is based on an imaging spectrometer coupled to a CCD (Charge-Coupled Device) sensor, the Adimec-1000 m (Adimec, Eindhoven, Netherlands). It has four magnification lenses: 5×, 10×, 20× and 50×. These numbers refer to the magnification achieved by the system to capture the sample (e.g., a 4× objective lens is magnifying the sample four times its size). This HS system works in the visual and near-infrared (VNIR) spectral range from 400 to 1000 nm with a spectral resolution of 2.8 nm, sampling 826 spectral channels and 1004 spatial pixels per line. The push-broom camera performs spatial scanning to acquire an HS cube with a mechanical stage (SCAN, Märzhäuser, Germany) attached to the BF microscope, which provides accurate movement (±3 µm accuracy) of the specimens in the 3 axes directions: $[x, y, z]$. The objective lenses are from the LMPLFLN (Long working distance Plan SemiApochromat) family (Olympus, Tokyo, Japan), which are optimized for infra-red (IR) observations. The light source is a 12 V, 100 W halogen lamp. This system was previously employed in histological HS analysis of brain cancer samples [17], breast tumor cell detection [6] and disease biomarker identification in plasma [44]. Furthermore, to create our DB, 125 plant and animal histology samples (rat histology, stems, leaf structure and blood smears) from the company Brunel microscopes Ltd

Optics EXPRESS

(England, UK) were employed (Fig. 3). These microscope slides are characterized by being theoretically flat, so no roughness is expected in the specimens.

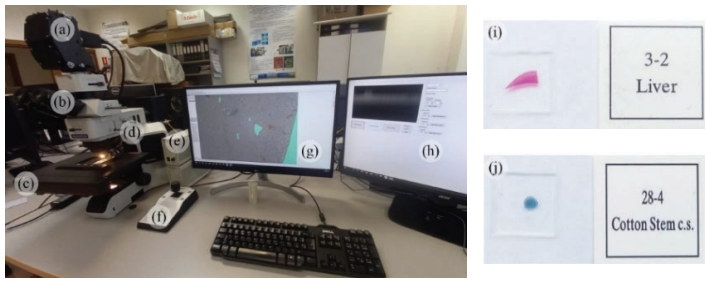


Fig. 3. Laboratory HS microscope employed in this work. (a) HS camera. (b) Eyepieces. (c) Motorized stage. (d) Objectives. (e) Light sources. (f) Joystick to move the stage. (g) RGB imaging software. (h) HS imaging software where a single HS frame can be visualized for focusing purpose. (i) Example of a histology rat liver tissue sample. (j) Example of histology cotton stem sample.

3.2. Database description

A dataset for focus analysis was obtained by capturing several images at different *WDs* to produce different levels of blurriness. Samples to be captured included rat intestine (Fig. 4(a)), bird blood (Fig. 4(b)), cotton leaf (Fig. 4(c)) and cladophores (Fig. 4(d)). However, the last ones were the least suitable for this experiment since they have a homogeneous spatial frequency and differences in blurriness are difficult to differentiate. Moreover, algae samples were also tested for this experiment, but it was discovered that they were non-flat and so, they cannot be included in these experiments (Fig. 4(d)).

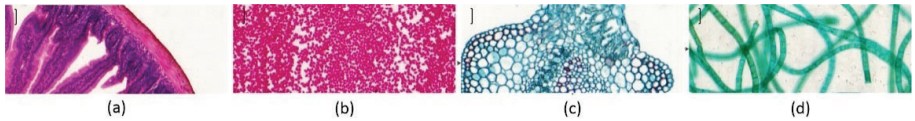


Fig. 4. Focused ROI images captured at 10× of different samples: (a) Rat intestine tissue; (b) Bird blood smear; (c) Cotton Leaf; and (d) Cladophores Algae. Scale bar: 50 µm.

Concerning the objectives lenses, in our HS acquisition system light attenuation is inversely proportional to the magnification. Thus, the higher the magnifying power, the higher the power of light it is needed to capture the sample. Using the default light system, it is not able to capture clear images using 50× objective lens. So, knowing that tissue structures captured at different magnifications can resolve different classification problems, three objective lenses (creating different pixel sizes) were selected (5×, 10×, 20×).

To select the number of images in the database, we had to consider that they would have to be reviewed by different users. Thus, following the ITU recommendations cited in Table 3, the total time of the test should not exceed 30 minutes. A pair of images is evaluated in around two seconds [45] so, it is possible to perform 900 rounds in 30 minutes. However, to be able to include some more images, we opted to design the subjective test in two sections of 20 minutes each, being able to perform 1200 pair comparisons in total. Following the swizz tournament, 24 pair comparisons (11 images, 2 images per round, 4 round per sample) are needed for each

Optics EXPRESS

sample. In conclusion, no more than 50 samples could be tested per magnification (considering that each magnification is performed by different subjects).

A summary of the number of reference images captured at each magnification and using each sample is shown in Table 5. The total number of images refers to the total number of reference images multiplied by the 11 levels of distortions selected in this work (i.e., number of images captured at different WDs for each reference image).

Table 5. Summary of the number of images and samples included in the HIDFA database.

	No. of reference images				Total No. of images
	5×	10×	20×	Total	
Rat histology	8	16	16	40	440
Leaf structure	5	10	6	21	231
Stems	16	10	19	45	495
Blood smears	0[a]	7	5	12	132
Freshwater Algae	0[a]	3	3	6	66
Total	29	46	49	**124**	**1364**

[a]No images were captured of blood smears and freshwater algae at 5× because they are mostly homogeneous and blur quantification algorithms could not perform properly.

The focusing process on HS data capturing using push-broom cameras is based on the subjective approach of the user. The HS image is focused by looking into one frame given by the push-broom camera ($Y\lambda$ frame shown in the grey image on the right at Fig. 5) and finding the sharpest spatial frequency along the different WD from the sensor to the sample. For each specimen, the focus point, WD_0, was firstly found (Fig. 5). The procedure continued by moving upwards a distance of $\Delta z\frac{(N-1)}{2}$. From this point, N HS images were captured distancing Δz each time from the previous image. N was always odd, producing the same number of slices upwards and downwards of WD_0. For each set of N HS images captured, a white and a dark reference image were also obtained at WD_0.

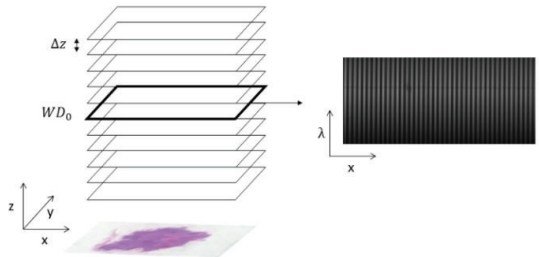


Fig. 5. Proposed HS DB capture procedure. Firstly, WD_0 was found and then N HS images were captured distancing Δz from each other.

Conventional HS calibration was performed over the HS cubes [8]. This is a standard procedure for flat field correction to transform captures to normalized transmittance. A push-broom frame is a capture of one spatial line containing the whole spectra of each pixel (1004 pixels × 826 spectral bands using the system of this work). During the capture process, a white reference image is recorded by capturing one push-broom frame of the light incident on the HS sensor (See Fig. S3 on Supplement 1). Similarly, the dark reference is obtained by turning off the light from

Optics EXPRESS

the system. Thus, an upper and lower light range are set. Then, Eq. (2) is performed over each push-broom frame (R) separately, to normalize the light.

$$I = \frac{R - DR}{WR - DR} \tag{2}$$

Then, preprocessing of the images were performed to reduce noise by removing the extreme bands as studied in [46], resulting HS cubes of 645 bands. After capturing and preprocessing the HS cubes, monochromatic images and synthetic RGB images were extracted. Monochromatic images were obtained by averaging the 645 bands of the preprocessed HS cubes. Synthetic RGB images were obtained by applying the normal probability density function for each RGB channel (R = 470 ± 0.04, G = 560 ± 0.06 and, B = 590 ± 0.08 nm), trying to imitate the HVS (see section S4.2. of Supplement 1).

Summarizing, 125 samples were selected for this specific application, captured at different *WD* from the focused plane and then, HS images were calibrated and preprocessed. Finally, a total of 1,375 HS images conformed the HIDFA and were used for testing different OSs.

3.3. Metrics for blur assessment

3.3.1. Subjective WD-based score (SS_{WD})

This score was given to each image according to their different *WDs* between the specimen and the objective lens. Thus, *WD* of each image can be calculated following Eq. (3), knowing that WD_0 is the point where the HS image is focused by visual inspection and, Δz was the increment distance between captures (0.2, 0.1 and 0.05 mm for 5×, 10× and 20×, respectively).

$$SS_{WD} = \frac{WD_0 \pm WD}{\Delta z} \tag{3}$$

3.3.2. Subjective MOS score (SS_{MOS})

This score was obtained in the form of MOS. All-play-all blurriness comparison between these images is practically non-viable, thus, the Swiss system tournament was employed. The principle of a Swiss competition is that each player will be pitted against another player who has performed as well (or as poorly) as themself. The first round is either drawn at random or seeded according to rating. The advantages of this method are twofold: it can find a clear winner between many competitors (the one that reach to the end without a single bad result), and it has a relatively small number of total rounds [47]. In our case, tournaments were organized between 11 different blur versions of a reference image, where users had to select the image closest to the reference one (Fig. 6). Ties were allowed. The first round was set randomly. Tournaments were developed in the scope of Microsoft Access in Windows 10. Moreover, subjective tests took 45 minutes, making a 5-minute pause in the middle of the experiment.

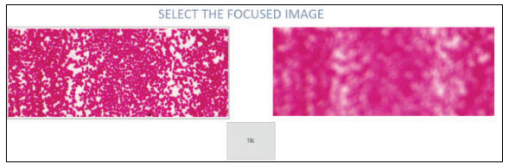


Fig. 6. Example of a round following the Swiss tournament system.

After all tests were completed MOS values for each single image were calculated following Eq. (4), where R_n is the rating of a subject over such image and N the total number of subjects

Optics EXPRESS

that went through the subjective test.

$$SS_{MOS} = \frac{\sum_{n=1}^{N} R_n}{N} \tag{4}$$

3.3.3. Objective scores

The main OSs employed for blur assessment were explained in section 2.5. In our experiments, we are concentrated on studying learning free methods, which can be further divided as shown in Fig. 7. From each of the four learning-free NR-IQA algorithm categories, the best available algorithm for monochromatic images was chosen. First, these algorithms were evaluated in a set of monochromatic images (computed from the mean of all wavelengths of the preprocessed HS image) at different *WDs*, since there is dependency between *WD* and focus. Then, they were applied to each band of each image, independently, to show the dependency between the wavelength and the focus (to prove the existence of any chromatic aberration). Experiments were carried out in MATLAB R2020a in a Windows environment (Microsoft Windows 10 Pro) with an Intel i7-10700 K 3.80 GHz and 64 GB RAM.

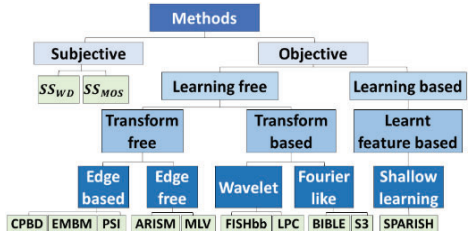


Fig. 7. Summary of all scores employed in this work.

3.3.4. Results evaluation

Once all the scores employed in this project have been explained in detail, next step was to correlate their results to test the quality of the different objective methods. To evaluate the obtained results, the Video Quality Experts Group (VQEG) [48] proposes mapping the *SS* to the *OS*. Since two different *SSs* (SS_{WD} and SS_{MOS}) were computed for each capture, three different correlations were performed: 1) SS_{WD} to SS_{MOS}; 2) SS_{WD} to *OS*; and, 3) SS_{MOS} to *OS*.

VQEG performs the mapping following Eq. (5), where $\tau_1 = \max \mathrm{F(OS)}$, $\tau_2 = \min \mathrm{F(SS)}$, $\tau_3 = mean(SS)$, *and* $\tau_4 = \frac{std(SS)}{4}$.

$$(SS) = \frac{\tau_1 - \tau_2}{1 + e^{\frac{SS - \tau_3}{\tau_4}}} + \tau_2 \tag{5}$$

Afterwards, three evaluation metrics were calculated to evaluate the method's performance: 1) Spearman's Rank-Order Correlation Coefficient (SRCC), which employs a monotonic function; 2) Pearson's Linear Correlation Coefficient (PLCC), which is a measure of the linear correlation after nonlinear mapping; 3) and Root-Mean-Square Error (RMSE), which is used to measure the differences after the nonlinear mapping. In competitive NR-IQA algorithms, the values of SRCC and PLCC are close to one, while the value of RMSE is close to zero.

4. Experimental results

4.1. HIDFA creation

Following the common procedures to create a database for IQA, the HIDFA database was generated (section S3, Fig. S2, Fig. S3 and Fig. S4 from Supplement 1) and it is now publicly

Optics EXPRESS

available (more information in data availability). HIDFA is characterized by increasing the number of reference images to 125 and the number of bands to 826 (Table 6). At the end, a total of 1,375 HS images were captured. ITU recommendations, explained in Table 2, indicated that at least 15 subjects are needed but no more than 40. Thus, a number of 24 participants was selected. Since each magnification would need a different reviewer, only images at 10× were employed for SS_{MOS} calculation. Concluding, twenty-four subjects evaluated a total of 46 reference images at 11 levels of distortion (506 images), resulting in 12,144 ratings.

Table 6. Comparison between HIDFA and other NR-IQA DBs

	No. reference images	Spectral range (nm)	Spatial range	Types of distortion	Levels of distortion	Total No. of images
TID2013 [21]	25	RGB	Macro	24	5	3000
LIVE [22]	29	RGB	Macro	5	5	779
CSIQ [23]	30	RGB	Macro	6	4 o 5	~810
SIDQ [24]	9	410–1000 (160 bands)	Macro	5	1	45
HIDFA [14]	29 at 5× 46 at 10× 49 at 20×	400–1000 (826 bands)	Micro	1	11	1364

Sociodemographic data of the subjects were studied to determine if SS_{MOS} values were independent of the subject characteristics (Table 1 of Supplement 1). Subjects aged between 20 and 76 years (Fig. S1 of Supplement 1), being 15 men and 9 women. Two thirds of them were qualified workers (only two did not go to university) while the rest were students. Thus, the average intellectual activity presented medium-high values. Regarding to the clinical variables, except for two subjects, they did not have a psychiatric history. None of the subjects presented hypertension or type II mellitus diabetes. In terms of visual acuity, we found some variability: two thirds had normal vision, while the rest had some vision problem, and so, wore glasses (except for one person with astigmatism who did not wear glasses). One of the subjects with normal vision had an eye surgery (more information on section S1. of Supplement 1).

Moreover, to determine whether sociodemographic data could be related to SS_{MOS} results, a statistical analysis was performed. To this aim, SS_{MOS} scores were calculated for each categorical value (e.g., SS_{MOS} for men and women). Standard deviation was computed over the 11 levels of distortions for a single reference image. The result were vectors of 46 values one per reference image. Finally, t-test and ANOVA [49] methods were performed between the standard deviation vectors of a feature with two or more categories, respectively. Statistical tests were employed to evaluate, at the 5% of significance level, if the null hypothesis is rejected (each variable is independent to the SS_{MOS} results). None of the features presents a $p<0.05$, so, not statistically significance was found. According to this statistical analysis, we can argue that SS_{MOS} results are independent to the sociodemographic data of the subjects.

4.2. SS analysis

SS_{MOS} values of the images taken at 10× (46 set of images) were calculated. To this aim, synthetic RGB images were generated by imitating the human eye spectral response [50]. Figure 8 shows the distribution of SS_{MOS} (from 1 to 6) along the SS_{WD} for four sample z-stack groups. Same results were obtained for all set of images of HIDFA. Mean correlation between SS_{MOS} and SS_{WD} was measured, achieving $SRCC = 0.89$, $PLCC = 0.8$ and $RMSE = 0.77$.

Optics EXPRESS

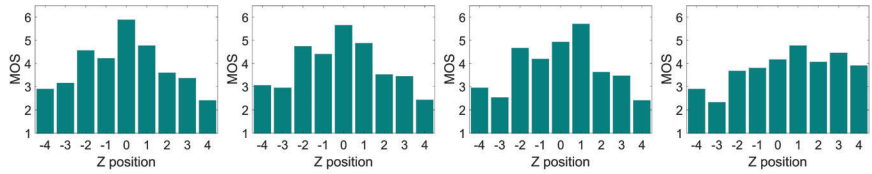


Fig. 8. Results of SS_{MOS} values for sample z-stack groups: (a) rat intestine tissue; (b) bird blood smear; (c) cotton leaf; and (d) cladophores.

4.3. *OS* score analysis

OSs were calculated over the synthetic RGB images, and the results were correlated to the SS_{WD} and the SS_{MOS}, as displayed in Table 7. Overall results are similar using both SSs. The best performance was achieved employing MLV, FISHbb, LPC and S3 algorithms. Same procedure was repeated for the monochromatic HS images obtained averaging all the bands of the HS cube (Table 8). In these results, ARISM, BIBLE and S3 algorithms were not computed since they only work for RGB images. The best performance was achieved employing MLV, FISHbb and LPC algorithms. Since there is a high correlation between the results obtained for SS_{WD} and SS_{MOS}, the latter was used in the next experiments to simplify the presentation of the results.

Table 7. Correlation Results for RGB Images

	Method		CPBDM [35]	EMBM [36]	PSI [20]	ARISM [37]	MLV [38]	FISHbb [39]	LPC [40]	BIBLE [41]	S3 [42]	SPARISH [43]
SS_{WD}	SRCC↑	Mean	0.62	0.69	0.72	0.50	0.84	**0.90**	0.83	0.72	**0.87**	0.7
		Std	0.26	0.24	0.25	0.26	0.15	**0.06**	0.18	0.23	**0.08**	0.22
	PLCC↑	Mean	0.63	0.64	0.70	0.51	0.81	**0.84**	**0.82**	0.76	0.81	0.76
		Std	0.24	0.22	0.22	0.28	0.12	**0.05**	**0.13**	0.13	0.06	0.13
	RMSE↓	Mean	2.32	2.33	2.04	2.50	**1.80**	1.98	**1.72**	2.12	2.05	2.14
		Std	0.41	0.34	0.18	0.52	**0.18**	0.09	**0.21**	0.18	0.10	0.18
SS_{MOS}	SRCC↑	Mean	0.61	0.67	0.69	0.5	0.80	**0.85**	0.82	0.66	**0.85**	0.65
		Std	0.27	0.27	0.28	0.26	0.25	**0.21**	0.23	0.28	**0.19**	0.28
	PLCC↑	Mean	0.62	0.59	0.67	0.5	0.77	**0.80**	**0.80**	0.70	0.77	0.7
		Std	0.24	0.22	0.27	0.27	0.24	**0.17**	**0.21**	0.23	0.17	0.22
	RMSE↓	Mean	2.33	2.33	1.94	2.57	**1.63**	1.80	**1.50**	1.98	1.90	2
		Std	0.63	0.47	0.44	0.71	**0.40**	0.31	**0.45**	0.30	0.22	0.3

4.4. *OS* analysis over single bands

After the evaluation of the focus *OS* on synthetic RGB and monochromatic images, we tested the focus measurement on the different bands of the HS cubes independently. The OS were computed over single band images (except for ARISM, BIBLE and S3 algorithms). Figure 9 shows the comparison of the results obtained for the monochromatic (mean over bands before metric computation) versus single band images (mean over bands after metric computation). Results show that methods can be divided in two: the ones with not similar means (CPBD, EMBM, PSI and, SPARISH) and the ones with similar means (MLV, FISHbb and, LPC).

For clarification, metric values were plotted against wavelength for focus visualization (Fig. 10). Some metrics seem to find differences between bands (CPBD, EMBM, PSI and, SPARISH),

Optics EXPRESS

Table 8. Correlation Results for Monochromatic Images

	Method		CPBDM [35]	EMBM [36]	PSI [20]	MLV [38]	FISHbb [39]	LPC [40]	SPARISH [43]
SS_{WD}	SRCC↑	Mean	0.6	0.73	0.47	**0.89**	**0.9**	0.83	0.61
		Std	0.31	0.13	0.27	**0.09**	**0.07**	0.14	0.25
	PLCC↑	Mean	0.61	0.74	0.49	**0.83**	**0.84**	0.81	0.69
		Std	0.29	0.06	0.27	**0.07**	**0.06**	0.12	0.22
	RMSE↓	Mean	2.44	2.18	2.34	1.99	1.96	**1.78**	2.17
		Std	0.54	0.18	0.47	0.09	0.09	**0.18**	0.2
SS_{MOS}	SRCC↑	Mean	0.6	0.72	0.45	**0.85**	**0.85**	0.8	0.56
		Std	0.31	0.18	0.29	**0.24**	**0.23**	0.24	0.29
	PLCC↑	Mean	0.62	0.69	0.45	0.8	**0.81**	0.8	0.64
		Std	0.28	0.16	0.3	0.17	**0.18**	0.19	0.26
	RMSE↓	Mean	2.46	2.11	2.45	1.8	1.76	**1.57**	2.06
		Std	0.78	0.33	0.66	0.33	0.33	**0.46**	0.36

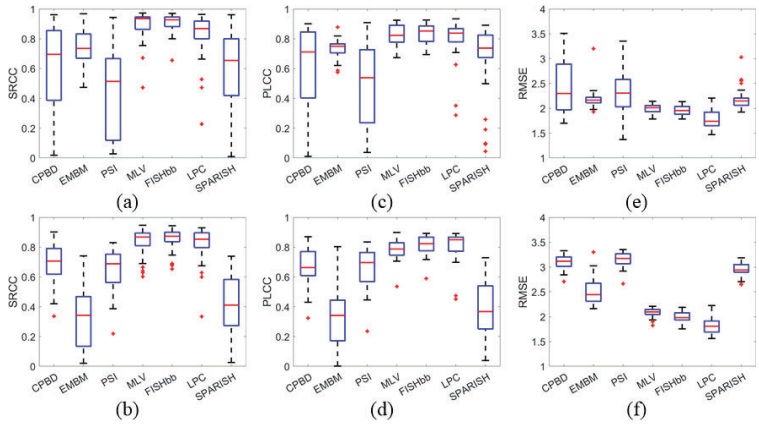


Fig. 9. Results of SS_{MOS} values versus *OS* for monochromatic images (a) SRCC; (c) PLCC; and the RMSE. Results of SS_{MOS} values vs *OS* for single band images (b) SRCC; (d) PLCC; and (f) RMSE.

Optics EXPRESS

while others keep a constant focus metric over the different wavelengths (MLV, FISHbb and, LPC). These results agree with the ones previously obtained.

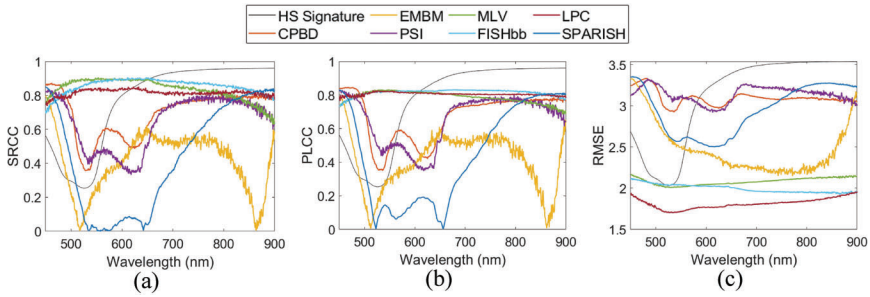


Fig. 10. Results of SS_{MOS} values versus OS for each single band image using the selected algorithms: (a) SRCC; (b) PLCC; and (c) RMSE.

4.5. *OS* analysis over single sample groups

To make a reliable wavelength comparison, samples must be divided into samples groups (blood smears, leaf structures, rat histology and stems). Figure 11 presents the results of the OS for each sample group and their mean HS signature. Overall, it can be deduced that worst metric values are obtained at lower information bands (HS signature lower peaks). The low blur value may be due to the few edges found in images with almost no color in such wavelength.

4.6. *OS* analysis over single frames

As it is not clear which algorithms are behaving accordingly to reality, further analysis was performed. Specific cases were studied by displaying a single band image from HS cubes for visual examination. Figure 12 show the monochromatic picture and several single band images of a rat histology capture at their most focused point ($SS_{WD} = 6$). PSI was the method selected to represent the ones that seem to find differences between bands, while FISHbb was representing the ones that keep a constant focus metric over the different wavelengths. PSI values decrease along with the loss of information in the image, while FISHbb values keep constant values. Thus, for an image with the same SS_{WD} for all wavelengths, FISHbb better recognizes the focus level along all bands.

4.7. Summary of the experimental results

In this work, several results have been obtained. Firstly, in order to test different blur quantification algorithms, the HIDFA was generated. Then a subjective test was carried out over 24 participants in order to recollect the MOS values of each synthetic RGB image of the dataset. Using these data, several state-of-the-art algorithms have been employed to select the most appropriate one for HS microscopy samples. MLV, FISHbb, LPC and S3 were the algorithms that provided the best results on synthetic RGB images. Afterwards, since HSI is intended for data collection of data across the light spectrum, each wavelength image was analyzed independently (ARISM, BIBLE and S3 methods could not be used because they only work over RGB images). From the results obtained, algorithms can be divided in two groups: methods that obtain similar values for the entire spectrum (G1: MLV, FISHbb and, LPC) and the ones that obtain different values (G2: CPBD, EMBM, PSI and, SPARISH). Afterwards, the images were divided into their original sample groups and the main HS signature was calculated for each one of them. In Fig. 11 it can be appreciated that G2 algorithms behave according to the information found in each band.

Optics EXPRESS

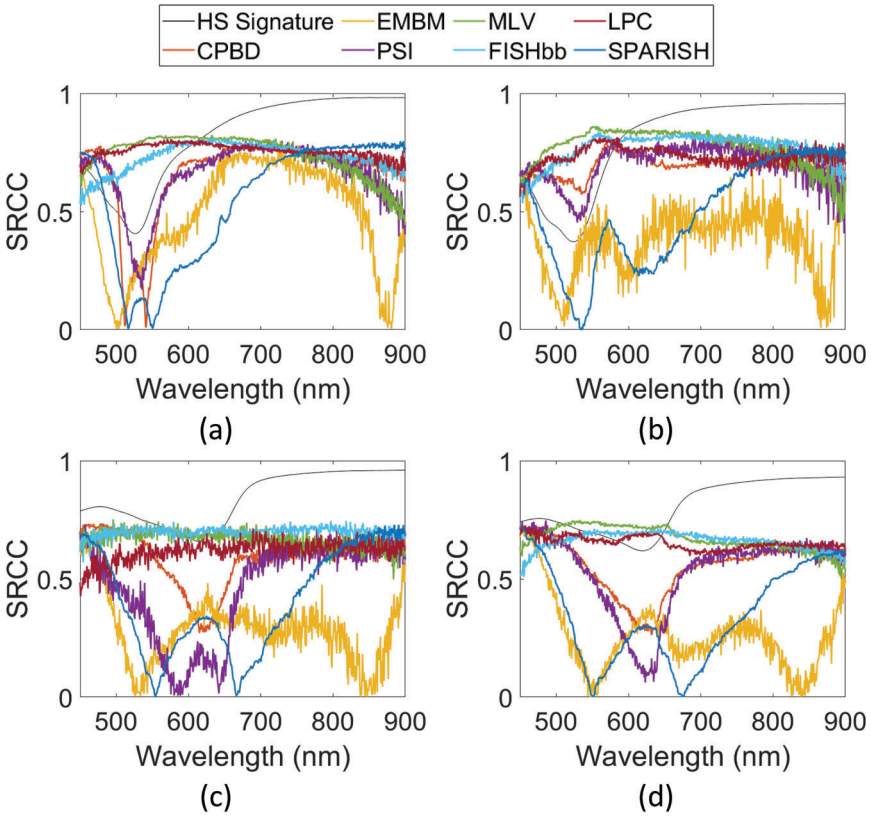


Fig. 11. SRCC correlation to SS_{MOS} values for: (a) rat histology; (b) blood smears; and (c) leaf structure; and (d) stems.

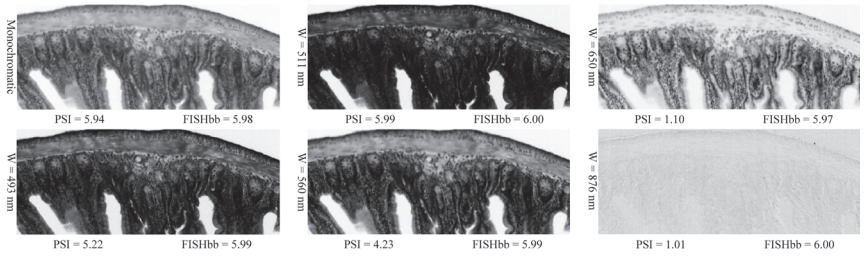


Fig. 12. Image of a rat intestine tissue at different wavelengths (W) and their corresponding *OS* values for the PSI and FISHbb algorithms, being $SS_{WD} = 6$.

Optics EXPRESS

Thus, we can conclude that G1 algorithms are sample independent while G2 methods are not. An example is found on Fig. 12, where different wavelength frames are shown next to their scores for one algorithm of each group. It can be appreciated that, although for the chromatic image PSI and FISHbb scores are similar, they vary along wavelengths. It is remarkable how FISHbb remains almost constant, although the spatial information is very different in each of the frames. Since all the spectrum of a HS cube is focused at the same WD, it would be coherent for all the bands to have the same score, making G1 algorithms more reliable.

5. Conclusion

Histopathology is the gold standard for cancer diagnosis. HSI is a novel technique that is showing high potential to improve the identification of tumor tissue and other diseases. High quality images are required to obtain an accurate automatic tissue classification. However, optimal focus is not always achieved due to the irregular surfaces of the samples. Thus, focus assessment becomes a relevant step before classifying histology images. In this paper, the state-of-the-art related to the existent *OSs* that quantify blurriness has been studied. Existing NR-IQA databases have also been reviewed.

One relevant contribution of this work is the creation of a public HS database for focus assessment based on 125 reference frames taken at 11 different WDs (a total of 1,375 HS cubes). SS_{MOS} scores were obtained by asking 24 subjects to complete a subjective test, where images were rated following the Swiss tournament system. Afterwards, such *SSs* were correlated to the previously studied *OS*s. The evaluation of these techniques, applied to synthetic RGB and monochromatic images generated from the HS cubes, agreed that the algorithms with better performance are MLV, FISHbb and LPC ($SRCC > 0.85$, $PLCC > 0.8$ *and* $RMSE < 1.8$). Then, OSs were also calculated over single band images, creating two groups of algorithms: the ones that obtain similar values for all the spectrum (G1: MLV, FISHbb and, LPC) and the ones that obtain different values (G2: CPBD, EMBM, PSI and, SPARISH). To select one group, the database was divided in the four origin groups (blood smears, leaf structures, rat histology and stems) and the HS signatures were compared to the values given by the algorithms. Since the entire spectrum of a HS cube is focused to the same WD, it would be coherent for all bands to have the same score. Thus, G1 algorithms should be selected since their score remain constant to information change along bands. Concluding, the most important contribution of this work is the finding of G1 algorithms to be more reliable for this application. Future work includes the use of these algorithms in z-stacking applications. The idea is to correct the blurriness, caused by different factors (Fig. 1) over the HS cubes. To this aim, the selected algorithms will be further studied to select the most focused areas and spectral bands over a range of HS cubes captured within the system's DOF.

Funding. Ministerio de Ciencia e Innovación (PID2020-116417RB-C42, TALENT-HExPERIA); Agencia Canaria de Investigación, Innovación y Sociedad de la Información (Conocimiento y Empleo, Consejería de Economía, Eje 3 Tema Prioritario 74 (85%), POC 2014-2020); Agencia Estatal de Investigación (European Union "NextGenerationEU/PRTR", FJC2020-043474-I, MCIN/AEI/10.13039/501100011033).

Acknowledgments. The authors would like to thank all twenty-four volunteers who participated in filling the MOS questionnaires for the subjective evaluation of the blurriness of the images of the dataset. This work was supported by the Spanish Government and European Union as part of the TALENT-HExPERIA (HypErsPEctRal Imaging for Artificial intelligence applications) project (PID2020-116417RB-C42). Moreover, this work was completed while Laura Quintana was beneficiary of the pre-doctoral grant given by the "Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI)" of the "Consejería de Economía, Conocimiento y Empleo", which is part-financed by the European Social Fund (FSE) (POC 2014-2020, Eje 3 Tema Prioritario 74 (85%)) and, Himar Fabelo was beneficiary of the FJC2020-043474-I funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR".

Disclosures. The authors declare no conflicts of interest.

Data availability. The datasets generated during the current study are available under reasonable request [14].

Optics EXPRESS

Supplemental document. See Supplement 1 for supporting content.

References

1. L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading," Comput. Methods Programs Biomed. **107**(3), 538–556 (2012).
2. G. Moallem, A. A. Pore, A. Gangadhar, H. Sari-Sarraf, and S. A. Vanapalli, "Detection of live breast cancer cells in bright-field microscopy images containing white blood cells by image analysis and deep learning," J. Biomed. Opt. **27**(07), 1 (2022).
3. A. Mora-Zuniga, S. Quiros-Barrantes, and F. Siles, "M-Phase Feature Extraction Algorithm for Phenotype Classification from Cancer Brightfield Microscopy," in *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)* (IEEE, 2018), pp. 1–8.
4. M. D. Zarella, D. Bowman, F. Aeffner, N. Farahani, A. Xthona, S. F. Absar, A. Parwani, M. Bui, and D. J. Hartman, "A Practical Guide to Whole Slide Imaging: A White Paper From the Digital Pathology Association," Arch Pathol Lab Med **143**(2), 222–234 (2019).
5. S. Ortega, M. Halicek, H. Fabelo, G. M. Callico, and B. Fei, "Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review," Biomed. Opt. Express **11**(6), 3195 (2020).
6. S. Ortega, M. Halicek, H. Fabelo, R. Guerra, C. Lopez, M. Lejeune, F. Godtliebsen, G. M. Callico, and B. Fei, "Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images," in *Medical Imaging 2020: Digital Pathology*, J. E. Tomaszewski and A. D. Ward, eds., (SPIE, 2020), p. 30.
7. M. Halicek, S. Ortega, H. Fabelo, C. Lopez, M. Lejaune, G. M. Callico, and B. Fei, "Conditional generative adversarial network for synthesizing hyperspectral images of breast cancer cells from digitized histology," in *Medical Imaging 2020: Digital Pathology*, J. E. Tomaszewski and A. D. Ward, eds., (SPIE, 2020), p. 29.
8. S. Ortega, R. Guerra, M. Diaz, H. Fabelo, S. Lopez, G. M. Callico, and R. Sarmiento, "Hyperspectral Push-Broom Microscope Development and Characterization," IEEE Access **7**, 122473–122491 (2019).
9. G. Wang and N. Fang, "Detecting and Tracking Nonfluorescent Nanoparticle Probes in Live Cells," in (2012), pp. 83–108.
10. L. Palmieri, G. Scrofani, N. Incardona, G. Saavedra, M. Martínez-Corral, and R. Koch, "Robust Depth Estimation for Light Field Microscopy," Sensors **19**(3), 500 (2019).
11. E. A. El-Gabry, A. V. Parwani, and L. Pantanowitz, "Whole-slide imaging: widening the scope of cytopathology," Diagn Histopathol **20**(12), 456–461 (2014).
12. C. Meng, P. An, X. Huang, C. Yang, L. Shen, and B. Wang, "Objective Quality Assessment of Lenslet Light Field Image Based on Focus Stack," IEEE Trans. Multimedia **24**, 3193–3207 (2021).
13. D. Li and T. Jiang, "Blur-Specific No-Reference Image Quality Assessment: A Classification and Review of Representative Methods," in (2019), pp. 45–68.
14. L. Quintana-Quintana, S. Ortega, H. Fabelo, F. J. Balea-Fernández, and G. M. Callico, "Hyperspectral Image Dataset for Focus Analysis (HIDFA)," https://hsidatabase.iuma.ulpgc.es/.
15. Elizabeth Allen and Sophie Triantaphillidou, *The Manual of Photography* (2011).
16. N. Burningham, Z. Pizlo, and J. P. Allebach, "Image Quality Metrics," in *Encyclopedia of Imaging Science and Technology* (John Wiley & Sons, Inc., 2002).
17. S. Ortega, M. Halicek, H. Fabelo, R. Camacho, M. de la, L. Plaza, F. Godtliebsen, G. M. Callicó, and B. Fei, "Hyperspectral Imaging for the Detection of Glioblastoma Tumor Cells in H&E Slides Using Convolutional Neural Networks," Sensors **20**(7), 1911 (2020).
18. A. Hore and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," in *2010 20th International Conference on Pattern Recognition* (IEEE, 2010), pp. 2366–2369.
19. Guangtao Zhai, Xiaolin Wu, Xiaokang Yang, Weisi Lin, and Wenjun Zhang, "A Psychovisual Quality Metric in Free-Energy Principle," IEEE Trans. on Image Process. **21**(1), 41–52 (2012).
20. C. Feichtenhofer, H. Fassold, and P. Schallauer, "A Perceptual Image Sharpness Metric Based on Local Edge Gradient Analysis," IEEE Signal Process. Lett. **20**(4), 379–382 (2013).
21. N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," Signal Process Image Commun **30**, 57–77 (2015).
22. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. on Image Process. **13**(4), 600–612 (2004).
23. D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," J. Electron. Imaging **19**(1), 011006 (2010).
24. S. le Moan, S. George, M. Pedersen, J. Blahová, and J. Y. Hardeberg, "A database for spectral image quality," in M.-C. Larabi and S. Triantaphillidou, eds. (2015), p. 93960P.
25. R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," Multimed Syst **22**(2), 213–227 (2016).
26. J. R. Lewis, "Psychometric Properties of the Mean Opinion Scale," Proceedings of HCI International **1**, 149–153 (2001).
27. ITU-T Recommendation P.911, *Subjective Audiovisual Quality Assessment Methods for Multimedia Applications* (1998).

Optics EXPRESS

28. Rec. ITU-R BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," (2002).
29. S. Jumisko-Pyykkö and J. Häkkinen, "Profiles of the Evaluators – Impact of Psychographic Variables on the Consumer-oriented Quality Assessment of Mobile Television," Multimedia on Mobile Devices **6821**, 68210L (2008).
30. F. Speranza, F. Poulin, R. Renaud, M. Caron, and J. Dupras, "Objective and subjective quality assessment with expert and non-expert viewers," in *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)* (IEEE, 2010), pp. 46–51.
31. O. Köster, M. Jessen, F. Khairi, and H. Eckert, "Auditory-perceptual identification of voice quality by expert and non-expert listeners," Proceedings of the XVI International Congress of the Phonetic Sciences (ICPhS) **1**, 1845–1848 (2007).
32. H. Choi, "Subjective video quality comparison using various displays," Opt. Eng. **48**(3), 037002 (2009).
33. "Methodologies for the subjective assessment of the quality of television images," 103 (2019).
34. G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psychol Rev **63**(2), 81–97 (1956).
35. N. D. Narvekar and L. J. Karam, "A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)," IEEE Transactions on Image Processing **20**(9), 2678–2683 (2011).
36. J. Guan, W. Zhang, J. Gu, and H. Ren, "No-reference blur assessment based on edge modeling," J Vis Commun Image Represent **29**, 1–7 (2015).
37. Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang, "No-Reference Image Sharpness Assessment in Autoregressive Parameter Space," IEEE Trans. on Image Process. **24**(10), 3218–3231 (2015).
38. K. Bahrami and A. C. Kot, "A Fast Approach for No-Reference Image Sharpness Assessment Based on Maximum Local Variation," IEEE Signal Process. Lett. **21**(6), 751–755 (2014).
39. P. V. Vu and D. M. Chandler, "A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation," IEEE Signal Process. Lett. **19**(7), 423–426 (2012).
40. R. Hassen, Zhou Wang, and M. M. A. Salama, "Image Sharpness Assessment Based on Local Phase Coherence," IEEE Trans. on Image Process. **22**(7), 2798–2810 (2013).
41. L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, "No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments," IEEE Trans. Cybern. **46**(1), 39–50 (2016).
42. C. T. Vu, T. D. Phan, and D. M. Chandler, "S3: A Spectral and Spatial Measure of Local Perceived Sharpness in Natural Images," IEEE Trans. on Image Process. **21**(3), 934–945 (2012).
43. L. Li, D. Wu, J. Wu, H. Li, W. Lin, and A. C. Kot, "Image Sharpness Assessment by Sparse Representation," IEEE Trans. Multimedia **18**(6), 1085–1097 (2016).
44. H. Fabelo, R. Leon, S. Ortega, F. Balea-Fernandez, C. Bilbao, G. M. Callico, and A. Wagner, "Novel Methodology for Alzheimer's Disease Biomarker Identification in Plasma using Hyperspectral Microscopy," in *2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS)* (2020).
45. M. Perez-Ortiz, A. Mikhailiuk, E. Zerman, V. Hulusic, G. Valenzise, and R. K. Mantiuk, "From Pairwise Comparisons and Rating to a Unified Quality Scale," IEEE Trans. on Image Process. **29**, 1139–1151 (2020).
46. H. Fabelo, S. Ortega, and A. Szolna, *et al.*, "In-Vivo Hyperspectral Human Brain Image Database for Brain Cancer Detection," IEEE Access **7**, 39098–39116 (2019).
47. L. Csató, "On the ranking of a Swiss system chess team tournament," Ann Oper Res **254**(1-2), 17–36 (2017).
48. VQEG, Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment (2000).
49. P. Mishra, U. Singh, C. Pandey, P. Mishra, and G. Pandey, "Application of student's t-test, analysis of variance, and covariance," Ann Card Anaesth **22**(4), 407 (2019).
50. L. Giannoni, F. Lange, and I. Tachtsidis, "Hyperspectral imaging solutions for brain tissue metabolic and hemodynamic monitoring: past, current and future developments," J. Opt. **20**(4), 044009 (2018).

C1. Assessing Processing Strategies on Data from Medical Hyperspectral Acquisition Systems

<table>
<tr><td>Title:</td><td colspan="2">Assessing Processing Strategies on Data from Medical Hyperspectral Acquisition Systems</td></tr>
<tr><td>Authors:</td><td colspan="2">Laura Quintana-Quintana, Carlos Vega, Raquel Leon, Guillermo V. Socorro-Marrero, Samuel Ortega, and Gustavo M. Callico</td></tr>
<tr><td>Conference:</td><td colspan="2">27th Euromicro Conference on Digital System Design (DSD)</td></tr>
<tr><td>Organizer:</td><td colspan="2">Euromicro</td></tr>
<tr><td>Date: 2024</td><td>Vol.: 1</td><td>Pages: 464-471</td></tr>
<tr><td colspan="2">Place: Paris, France</td><td>International</td></tr>
<tr><td>doi:</td><td colspan="2">https://doi.org/10.1109/DSD64264.2024.00068</td></tr>
<tr><td colspan="3">Open Access, Corresponding Author</td></tr>
</table>

Assessing Processing Strategies on Data from Medical Hyperspectral Acquisition Systems

Laura Quintana-Quintana[1*], Carlos Vega[1], Raquel Leon[1], Guillermo V. Socorro-Marrero[1], Samuel Ortega[2,1,3*], Gustavo M. Callico[1]
[1]Research Institute for Applied Microelectronics (IUMA), University of Las Palmas de Gran Canaria (ULPGC), Spain
[2]Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima), Tromsø, Norway
[3]Department of Mathematics and Statistics, UiT The Arctic University of Norway, Tromsø, Norway
*{lquintana, sortega}@iuma.ulpgc.es

***Abstract*— Hyperspectral imaging (HSI) has gained prominence in medical diagnostics due to its ability to capture and analyse detailed spectral information beyond human visual capabilities. Processing of HSI data is essential to enhance subsequent analysis and ensure the accuracy of results by reducing noise and unwanted artifacts. This paper provides an overview of state-of-the-art processing methods for HSI data, focusing on smoothing, normalization, and spectral derivatives. The efficacy of these methods is evaluated using root mean square error (RMSE) to compare pre-processed data with wavelength reference standard, alongside execution time considerations. Results indicate that certain algorithms, such as smoothing based on moving average, standard normal variate, and first spectral derivatives, yield superior performance across different medical HSI systems. Additionally, combining these processing techniques further improves data fidelity to the wavelength reference standard. Overall, this study offers insights into optimal processing strategies for enhancing the accuracy and reliability of HSI data.**

Keywords*— *hyperspectral imaging, processing methods, smoothing, normalization, spectral derivatives.

I. INTRODUCTION

Hyperspectral (HS) imaging (HSI) is a technique that captures and analyses a portion of the light spectrum in fine detail, often beyond human visual capabilities. This allows for the detection of substances with different spectral signatures. While initially used in remote sensing [1], HSI is increasingly applied in medical contexts for pathology identification, such as breast tumour [2], brain histological samples [3], colon and esophagogastric cancer [4], or skin and lung cancers [5]. After HSI capturing and prior to the development of tissue classification algorithms, HS data is pre-processed to remove non-tissue-related variability (e.g. inhomogeneities in illumination and noise in acquisition systems). The pre-processing algorithms usually aim to enhance subsequent information extraction from the data, such as compensate data according to Lambert-Beer's law or improving classification and regression models.

Traditional works were centred on spectral correction of HSI data. Numerous studies have investigated near-infrared (NIR) spectroscopy pre-processing, highlighting its critical role for subsequent analysis. It has been widely used in chemometrics, a field dedicated to correlating chemical measurements with properties of interest, such as the concentration of specific molecules. In 2009, Rinnan *et al.* published a review of the most common processing techniques for near-infrared spectra [6]. The authors classified pre-processing algorithms into two categories: scatter-correction and spectral derivatives. The theoretical foundations of the algorithms were explained, and a quantitative experiment was performed over near-infrared data of marzipan samples. Root Mean Square Error (RMSE) results indicated that the maximum improvement of any pre-processed model compared to the unprocessed one was approximately 25 %. These findings underline the challenge of identifying the best pre-processing method, mainly due to the impact of incorrect parameter settings, such as window size for smoothing algorithms. Additionally, it is noted that combining processing techniques is common practice, with guidelines typically recommending smoothing before normalization.

Likewise, in 2013, Engel *et al.* reviewed several pre-processing algorithms for chemometric data analysis (e.g. noise removal, base line offset and slope, light scatter, temporal and spectral misalignment, normalization, etc.) [7]. The authors focused on selecting the optimal data-analysis strategy to choose the combination of pre-processing methods that effectively remove artifacts and enhance the necessary information for various data-analysis goals. They explored three main approaches: trial and error based on subsequent analysis results, visual examination for artifacts, and quantifying data quality parameters to assess artifacts. These strategies were tested on a labelled dataset of beers with different alcohol content. Results indicated that choosing an effective pre-processing method, or combination of methods, significantly impacts analysis outcomes, with a variance of over 20 % in model accuracy between the best and worst pre-processing strategies. The study concluded that suboptimal pre-processing can greatly hinder data analysis goals, highlighting the need for a robust, quantitative approach to determine the optimal pre-processing strategy.

Following these lines, in 2004, Ezenarro *et al.* developed a MATLAB® toolbox named ProSpecTool to aid in selecting optimal pre-processing techniques for regression models in infrared (IR) spectroscopy [8]. ProSpecTool addressed the challenge by using objective criteria to filter and iterate pre-processing methods, including the classic set of pre-processing techniques (smoothing, normalization and derivatives) over raw data. The Forages dataset [9] was used as an example and performance evaluation metrics were calculated using pre-processed data. Results indicated that ProSpecTool can produce robust models comparable to those from trial-and-error approaches, but in significantly less time. Authors suggested that ProSpecTool could be an invaluable exploratory tool for IR

2024 27th Euromicro Conference on Digital System Design (DSD) | 979-8-3503-8038-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/DSD64264.2024.00068

2771-2508/24/$31.00 ©2024 IEEE
DOI 10.1109/DSD64264.2024.00068

spectroscopy practitioners, although final model validation and interpretation remain being the responsibility of the analyst. Recently, in 2023, Cozzolino *et al.* provided a broad overview of pre-processing methods for HSI data [10]. The article discussed several pre-processing methods beyond the former explained scatter correction and spectral derivatives algorithms (e.g., dimension reduction, resolution enhancement, geometric correction). As previous authors have concluded, they stated that pre-processing techniques are crucial to ensure the accuracy and reliability of the results by reducing interferences such as noise and unwanted artifacts. They also highlighted the need for iterative evaluation and simplicity in pre-processing techniques.

Other authors have centred their efforts just on removing noise from HS images, since it can disrupt accuracy of HS spectral-spatial feature classification. Singh *et al.* work proposed a new method involving interpolation to recover lost band information [11]. It employs Principal Components Analysis (PCA) [12] and Locality Preserving Projection (LPP) [13] to extract hybrid features, blending local and global spatial data. They tested it on standard datasets (Indian Pines, Salinas, Pavia University, and Kennedy Space Centre) and results showed that the accuracy of the classification increased significantly (313.81 %, 1448.33 % 153.21 % and 139.21 %, respectively) when the proposed framework is blended with state-of-the-art classifiers. Additionally, Bharath Bhushan *et al.* introduced a processing algorithm for band selection in HSI analysis, focused on removing noisy bands [13]. This algorithm calculates average inter-band block-wise correlation coefficients and applies a simple thresholding strategy based on their standard deviation. Validation using AVIRIS data [14] demonstrates its ability to identify noisy and water absorption bands accurately. The algorithm achieves a 94.73 % probability of correctly detecting noisy bands and a 3.18 % probability of false detection. The authors concluded that this processing algorithm successfully eliminates noise, improves band quality, and reduces computation complexity in band selection.

Out of all the spectral data processing documents found in the state-of-the-art, just two of them specifically addressed processing techniques for medical HS data. In 2015, Koprowski introduced a source code that can be freely used for preliminary HS data analysis on MATLAB® and discusses problems encountered when analysing medical HS images [15]. The application facilitates preliminary analysis of HS images, including reading data in various formats, calibration, filtering, and visualization. The graphical user interface and source code are available for download without restrictions, enabling users to modify and utilize them as needed. However, the authors did not investigate the impact of various standard processing methods on medical HS data. Thus, in 2022, Witteveen *et al.* conducted a comparison of different processing algorithms to reduce variations in spectra, caused by external factors, while preserving contrast between tissues with different optical properties [16]. They created a synthetic database by simulating tissue with differences in blood volume fraction, presence of different absorbers, scatter amplitude, and scatter slope. Furthermore, in order to test the methods for clinical validation, two clinical databases were collected (breast [17] and colorectal [18] tumour) in the Visible and Near Infrared (VNIR) and NIR ranges. Then, they analyse several processing algorithms which can be divided into normalization and spectral derivatives. The evaluation metrics included the overlap coefficient, which was calculated over pre-processed data belonging to different classes (e.g. 0.5 % or 2 % of blood volume fraction). They conclude that, overall, normal standardization (SNV), min–max, area under the curve, and single wavelength normalization are the most appropriate methods for processing data intended for the development of a tissue classification algorithm.

After this review of the state-of-the-art, it can be extracted that the current research faces limitations due to the absence of a clear standard for processing medical HS data. This lack of standard hinders the development of precise HSI intraoperative margin assessment techniques needed in medical context (e.g., complete tumour removal while preserving healthy tissue in cancer surgery) [19]. However, these requirements for precision are not as stringent in fields such as agriculture, food science, or chemometrics. Witteveen *et al.* did a review of eight processing algorithms, but only normalization and spectral derivative methods were studied. The aim of this work is to preliminary design a standard for preprocessing HS data from medical HS acquisition systems. State-of-the-art processing methods are going to be assessed, including noise removal techniques (smoothing), intensity misaligns of signals (normalization) and spectral derivatives. Each algorithm will be applied to HS images of a wavelength reference polymer captured using several medical HS systems. Finally, the data resulting from each processing method, or combination of them, will be evaluated against the wavelength reference standard of the material using RMSE, while also considering execution time.

II. Material and Methods

The whole set of medical HS systems included in this work will be described, the spectral characterisation target will be shown and finally, the different spectral processing methods and evaluation metrics will be presented.

A. Hyperspectral Medical Systems

Several medical HS systems, customized for various applications (gynaecology (G_VNIR) [20], [21], histology (H_VNIR) [22], [23], [24], and neurosurgery (N_VNIR and N_NIR) [25], [26]) are going to be used in this study. Table I provides a summary of the medical HS systems, their integrated HS camera, and their key optical features.

Table I Summary of Medical HS Systems

	G_VNIR	H_VNIR	N_VNIR	N_NIR
Application	Gynaecology	Histology	Neurosurgery	Neurosurgery
Optical System	Colposcope OP-C5 (OPTOMIC ESPAÑA, S.A., Spain)	Microscope Olympus BX53 (OLYMPUS Corp., Tokyo, Japan)	Custom	Custom
HS Camera	SnapScan VNIR	Pushbroom VNIR	Pushbroom VNIR	Pushbroom NIR
Magnification	0.4×, 0.6×, 1×, 1.6×, 2.5×	5×, 10×, 20× and 50×	1×	1×
Halogen Lamp Power (W)	50	100	150	150

More information about the HS sensor employed by each medical HS system is available in Table II. These setups are shown in Fig. 1. In addition, it must be mentioned that, for each medical HS system, custom software was developed to set image acquisition parameters.

TABLE II SUMMARY OF HS CAMERAS

	SnapScan VNIR	Pushbroom VNIR	Pushbroom NIR
HS Camera	Imec SNAPSCAN VNIR (Imec, Leuven, Belgium)	Hyperspec® VNIR A-Series, (Headwall Photonics, Inc., MA, USA)	Hyperspec® NIR 100/U, (Headwall Photonics, Inc., MA, USA)
HS Technology	SnapScan	Pushbroom	Pushbroom
Sensor	CMOS sensor, ams CMV2000 (ams OSRAM AG, Munich, Germany)	CCD sensor, Adimec-1000m (Adimec Advanced Image Systems B.V., Eindhoven, Netherlands)	InGaAs sensor, Xeneth XEVA 5052 (Xenics nv, Leuven, Belgium)
Spectral range (nm)	470 - 900	400 - 1000	900–1700
Spectral Sampling (nm)	2.86	0.72	4.65
Spectral Bandwidth / FWHM (nm)	10-15	2.5	4
# Spectral Channels	158	826	172
FOV (pixels)	1000 x 900	1×1004	1×320
Pixel Size (µm)	5.5	7.4	4.8

*CMOS: complementary metal-oxide semiconductor. CCD: charge-coupled device. InGaAs: Indium gallium arsenide. FOV: Field of View.

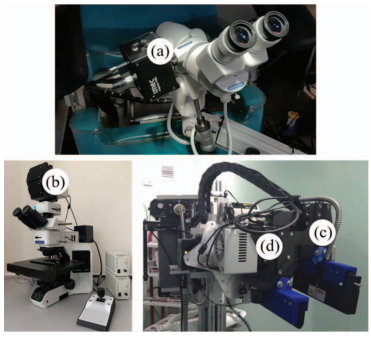


Fig. 1. Set-ups: (a) G_VNIR, (b) H_VNIR, (c) N_VNIR, and (d) N_NIR.

B. *Spectral Characterization Target*

The spectral characterization target which is going to be employed in this study is the NIST (National Institute of Standards and Technology) traceable wavelength calibration standard, which is composed by a mixture of three pure, rare-earth oxides (holmium, erbium, and dysprosium), mixed into Zenith Polymer® (LabSphere, Inc., NH, USA). The wavelength reference polymer exhibits distinct absorption peaks which provides a stable wavelength reference standard for validating spectrophotometers data. In Fig. 2 (a) it can be seen an image of the polymer itself which was captured once by each of the medical HS systems generating HS cubes of different sizes (G_VNIR (captured at 0.6×): 900×1000×158, H_VNIR (captured at 10×): 100×1004×826, N_VNIR: 741×1004×826, and N_NIR: 253×320×172). Fig. 2 (b) displays the spectral wavelength provided by the manufacturer (250 to 2450 nm, 2200 bands at 0.1 nm spectral resolution).

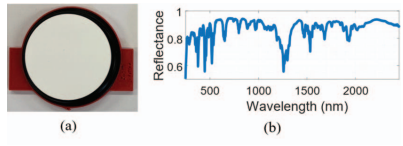


Fig. 2. Wavelength reference standard (a) picture and (b) spectral signature.

C. *Processing Methods*

During processing of HSI data, first flat-field correction is performed over the data to address variations caused by sensor and environmental factors [27]. This involves transforming digital numbers into radiation intensity [28] or true reflectance values [29], [30], [31]. The processes to convert raw sensor data (R) into calibrated images (CI), indicated in equation (1), requires a dark current (DC) capture, acquired with the shutter closed, and a white reference (WR), obtained using a diffuse reflectance standard from SphereOptics, which is produced using a Zenith Polymer® (LabSphere Inc., NH, USA) reflectance material [32]. This material is based on polytetrafluoroethylene and ensures a reflectivity above 99 %, making it an excellent choice for reflectance standards as it can withstand high levels of heat, humidity, and radiation exposure. These standards have a matte Lambertian reflecting surface, ensuring nearly equal intensity in all directions, with total reflection close to 100 % [32].

$$CI = \frac{R - DC}{WR - DC} \tag{1}$$

Flat-field correction is a well-established initial step before any further HSI processing [29], [30], [31], [33], [34]. However, further processing is subjected to the HSI instrumentation and data being studied. State-of-the-art HSI processing algorithms can be divided into smoothing, normalization, and spectral derivatives. Additionally, as emphasized by Rinnan *et al.* [6], the common practice involves combining processing techniques, with recommendations suggesting smoothing preceding normalization.

a. Smoothing Algorithms

Smoothing algorithms, aiming at reducing noise and irregularities, are conducted in the spectral domain of the data. These algorithms consider a symmetric window comprising k neighbours to the right- and left-hand side of a certain sample in the data sequence. All these values together comprise a smoothing window of odd value W, defined by (2). Since the signal at each point is affected by its neighbours' values, smoothing algorithms help to enhance signal and mitigate fluctuations, making patterns or trends more discernible but

attenuating high-frequency characteristics. The most common smoothing algorithms were selected for this study (Table III), where each algorithm is applied over a single spatial pixel R, producing a smoothed spectrum R_S . $\boldsymbol{\lambda}$ symbolizes the wavelength parameter.

$$W = 2k + 1; k \in \mathbb{N} \quad (2)$$

TABLE III SMOOTHING ALGORITHMS SUMMARY

Name	Acronym	Equation ($R_S(\lambda)$)	Description
Average	MOVMEAN	$\frac{1}{W}\sum_{i=-k}^{k} R(\lambda + i\,\Delta\lambda)$	Mean value over a window to preserve the central tendency of the data.
Median	MOVMEDIAN	$Median$	Median value over a window to preserve the central tendency of the data when outliers are present.
Gaussian-weighted average	GAUSSIAN	$\frac{\sum_{i=-k}^{k} w_i\, R(\lambda + i\,\Delta\lambda)}{\sum_{i=-k}^{k} w_i}$	Gaussian-shaped weights (w_i), with standard deviation set to $W/5$, emphasizing points closer to the centre of the window.
Linear regression [35]	LOWESS	$\sum_{n=1}^{N} \beta_n R(\lambda)^n + \epsilon$ $\beta_n \equiv regression\ coefficients$ $\epsilon \equiv residuals$	First degree polynomial fitting ($N = 1$).
Quadratic regression [36]	LOESS		Second degree polynomial fitting ($N = 2$).
Robust linear regression [37]	RLOWESS		Modifies LOWESS by iteratively adjusting the weights to give less influence on outliers in the regression ($N = 1$).
Robust Quadratic regression [37]	RLOESS		Modifies LOESS by iteratively adjusting the weights to give less influence on outliers in the regression ($N = 2$).
Savitzky-Golay filter [38]	SGOLAY	$\frac{1}{F}\sum_{i=-k}^{k} C_i R(\lambda + i\,\Delta\lambda)$	Least squares fit via quadratic convoluting integers (C_i) and normalizing factor (F).

b. Normalization Algorithms

Normalization methods are employed in HSI to standardize data across its spectral component, ensuring comparability and removing biases caused by scale differences. The six normalization methods chosen for this work were the ones provided by Witteveen *et al* [16] in their review. They are further explained in (Table IV). Each algorithm is applied over a single spatial pixel R , producing a normalized spectrum R_N . $\boldsymbol{\lambda}$ symbolizes the wavelength parameter.

TABLE IV. NORMALIZATION ALGORITHMS SUMMARY

Name	Acronym	Equation ($R_N(\lambda)$)	Description
Standard normal variate[39]	SNV	$\frac{R(\lambda) - \bar{R}}{\sigma}$	Each spectrum is subtracted its own arithmetic mean ($\bar{R}$) and divided by the unbiased estimation of its standard deviation (σ).
Multiplicative scatter correction [40], [41]	MSC	$\frac{R_i(\lambda) - c}{d}$ $R_i(\lambda) = c + d \cdot \bar{R};$	Each spectrum is scaled and offset to fit a reference (mean spectrum).
Min-max Normalization	MINMAX	$\frac{R(\lambda) - \min(R(\lambda))}{\max(R(\lambda)) - \min(R(\lambda))}$	Each spectrum is scaled and offset using its min and max values.
Mean centering	MC	$R(\lambda) - \bar{R}$	Each spectrum is subtracted its own mean.
Single wavelength	SW	$\frac{R(\lambda)}{R(\lambda_0)}$ $\lambda_0 = \arg\max(R(\lambda))$	Each spectrum is divided by its value at the reference wavelength with most reflectance (λ_0).
Area Under the Curve	AUC	$\frac{R(\lambda)}{\sum_{\lambda=1}^{\lambda_{end}} R(\lambda)}$	Each spectrum is divided by its AUC.

c. Spectral Derivatives Algorithms

Finally, spectral derivative algorithms analyse changes in spectral data, revealing variations in absorption, reflectance, or emission properties across wavelengths. The specificity of absorption ranges of human molecules, as noted by Steven Jaques [REF], typically contains crucial information for distinguishing tissues. However, detecting small peaks can be challenging. In this sense, derivative methods, are crucial for identifying subtle spectral features and patterns not readily apparent in calibrated data. For this preliminary study, just the first and second derivatives were assessed for simplicity. They are detailed in Table V, where $R_D(\lambda)$ is the value of the derived spectrum at that λ.

TABLE V DERIVATIVES ALGORITHS SUMMARY

Name	Acronym	Equation ($R_D(\lambda)$)	Description
First derivative	FD	$\frac{R(\lambda + \Delta\lambda) - R(\lambda)}{\Delta\lambda}$	Approximation of the first derivative of each spectrum from first differences.
Second derivative	SD	$\frac{R(\lambda + 2\,\Delta\lambda) - R(\lambda)}{(\Delta\lambda)^2}$	Approximation of the second derivative of each spectrum from second differences.

D. Evaluation Metrics

After calculating the different processing over the calibrated data, evaluation metrics are needed to assess the similarity between the pre-processed data and the wavelength reference standard. Additionally, execution time is measured to compare their overall performance.

a. RMSE

RMSE is a statistical measure that assesses the accuracy of a model by quantifying the differences between observed (O_i) and actual (A_i) values, as described in (3), where N is the common size of both, the observed and the actual sequences. A lower RMSE value, close to zero, indicates a stronger similarity between the spectral signatures.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - A_i)^2} \quad (3)$$

The methodology followed in this work includes the evaluation of pre-processed HS data against a wavelength reflectance standard. In terms of smoothing algorithms, it must

be considered that its aim is to reduce local noise while keeping main features of the signal intact. Thus, when evaluating spectral signatures pre-processed using a smoothing algorithm, the data is compared to the wavelength reflectance standard. Since in normalization and spectral derivative methods the HS data undergoes a transformation that alters its offset and scale, their inverse operation is applied over the pre-processed data relative to the wavelength reflectance standard. For instance, in the case of the MC normalization method, following the subtraction of each spectrum's individual mean, the average value of the wavelength reflectance standard is added back to all spectra.

b. Execution Time

To determine the execution time of each algorithm in MATLAB® 2023a, the processing algorithms will be timed 10,000 times across a dataset composed of 10 pixel-wise spectra of 3095 bands each. The processing methods was executed on an Intel(R) Core (TM) i7-10700K processor with 64 GB of RAM.

III. Results and Discussion

In this section, the reflectance data will be shown. Several processing methods are applied over the HS data, which are then compared to the wavelength reflectance standard.

A. Flat-field Correction

Flat-field correction was carried out over the captured HS data to compensate for variations due to sensor and environmental factors. This process involved converting digital numbers into radiation intensity or true reflectance values. Then, extreme bands were removed for each sensor, since they exhibit a low signal to noise ratio. In the H_VNIR and N_VNIR systems, since they use the same HS camera, the same spectral range was removed (400- 440 and 909-1000 nm), resulting two HS cubes composed by 645 spectral bands each. The HS cubes of N_NIR system were composed of 144 spectral bands, removing the 900-956 and 1638-1700 nm spectral range. Finally, a registration of the wavelength reflectance standard with the HS data of each setup was performed by selecting the closets values of the standard wavelength vector with respect to the one of each setup (mean produced error was 0.025 nm).

Fig. 3 illustrates, for each setup, one calibrated spectral signature and its registered wavelength reference standard. It is important to note that the wavelength calibration standard is made of is a mixture of three pure, rare-earth oxides (Holmium, Erbium and Dysprosium) mixed into Zenith Polymer®. Whereas the SphereOptics white reference exhibits diffused reflection, light can be specular reflected in the small crystals of the rare-earth polymer, resulting in higher values on the raw HS cube compared to those in the white reference. This effect can cause the calibrated cube to have values above one, as seen in Fig. 3 (a), (b) and (c). Other wavelengths calibration standards will be explored in future experiments.

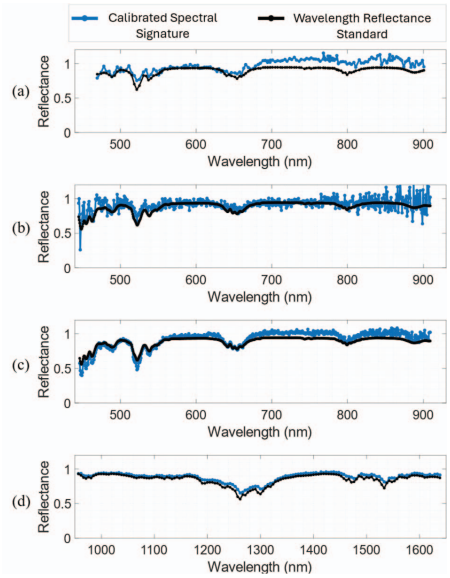


Fig. 3. Calibrated spectral signature of the different medical HS systems (blue), and the registered wavelength reference standard (black): (a) G_VNIR, (b) H_VNIR, (c) N_VNIR, and (d) N_NIR.

B. Processing Methods

a. Smoothing Algorithms

Smoothed pre-processed data from different setups were compared to the wavelength reflectance standard to find a smoothing method that balances the reduction of high-frequency noise and the retention of significant data. Smoothing was performed using a prescribed number of neighbours, 7, which resulted in window sizes of 20.02 nm for G_VNIR, 5.04 nm in H_VNIR and N_VNIR, and 32.55 nm in N_NIR (see Table II). These differences explain the dependency of results on the particular HS sensor used, as shown in Fig. 4 (a). H_VNIR and N_VNIR systems use the same pushbroom VNIR camera, and so their most effective algorithm correlates (MOVEMEAN). For the other two medical HS systems, GAUSSIAN algorithm works best for G_VNIR and LOESS for N_NIR. This last medical HS system consistently yields the best results, possibly due to its larger spectral sampling, since a wider spectral window is being considered when smoothing. Conversely, the H_VNIR system exhibits the opposite behaviour. This discrepancy suggests that the optimal smoothing algorithms and window sizes may need to be customized for different medical HSI systems. Furthermore, RMSE values of the unprocessed data with respect to the wavelength reflectance standard are also given for comparison (REF in Fig. 4 (a)). In terms of execution times (Fig. 4 (b)), it must be considered that RLOWESS and RLOESS have significantly longer execution times (~ 40.000 s) compared to others (~ 0.050 s), making the former ones less desirable. Also, it should be noted that MOVEMEAN is the fastest (0.035 s) since it is the simplest one.

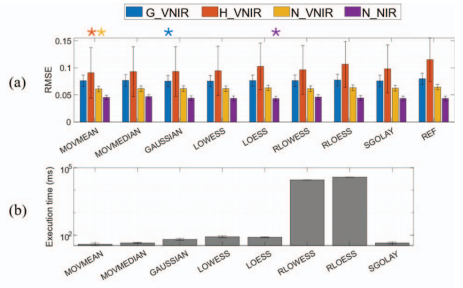


Fig. 4. Smoothing processing methods evaluation: (a) RMSE and (b) execution times. * represents the best values for each HS system.

b. Normalization Algorithms

Several normalization algorithms have been also examined, ranging from simple subtraction (MC) or scaling (AUC, SW) to combinations of these methods (SNV, MSC, MM). Results in Fig. 5 (a), indicate that SNV algorithm outperforms other normalization methods across all setups. The SNV algorithm is one of the combined methods that brings the pre-processed HS data closer to the wavelength reflectance standard. This algorithm becomes useful for uniform samples, but it should be further studied in real-world scenarios. Moreover, it can be observed that reference RMSE values of the unprocessed data with respect to the wavelength reflectance standard (REF in Fig. 5 (a)) are clearly improved using all methods but SW. Regarding execution times (Fig. 5 (b)), all algorithms take approximately 0.002 to 0.054 s. However, MSC is in the upper time bound, as it requires the fitting of a polynomial for each data point. The rest of the methodologies, based on basic operations, show no significant differences in their execution times.

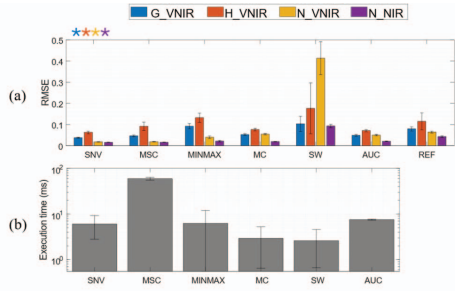


Fig. 5. Normalization processing methods evaluation: (a) RMSE and (b) execution times. * represents the best values for each HS system.

c. Spectral Derivatives Algorithms

Spectral derivative algorithms were also examined as an extra step of the processing chain. It is crucial to note that these methods are highly sensitive to noise, thus they must be applied to data that have already undergone noise reduction. Since MOVMEAN and SNV were the best performing algorithms, they are selected as the initial step before applying spectral derivatives. Results of comparing spectral derivate HS data with respect to the wavelength reflectance standard are shown in Fig. 6 (a). RMSE values show that FD performed better than the SD, although, not better than evaluating the unprocessed data with respect to the wavelength reflectance standard (REF in Fig. 6 (a)). This could be attributed to noise in the data. While derivatives are essential for identifying subtle spectral features and patterns that are not easily visible in calibrated data, they also amplify noise peaks. In terms of execution times (Fig. 6 (b)), SD algorithm is slower since it is composed of two FD in series.

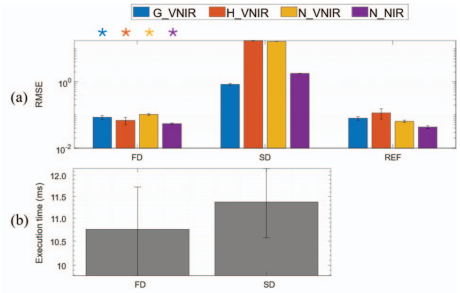


Fig. 6. Derivatives spectral processing methods evaluation: (a) RMSE and (b) execution times.* represents the best values for each HS system.

d. Combination of Algorithms

Furthermore, as recommended by Rinnan *et al.* [6], combining various processing algorithms is advisable, specifically suggesting smoothing prior to normalization. Following this guidance, all normalization algorithms were tested after applying each smoothing algorithm to the HS data, resulting in a total of 48 combinations. Table VI presents the RMSE values for each medical HS system, detailing the results for optimal smoothing, normalization, their combination, and the additional step of applying spectral derivative methods.

TABLE VI. RMSE OF BEST PERFORMING PROCESSING ALGORITHMS

	G_VNIR	**H_VNIR**	**N_VNIR**	**N_NIR**
Smoothing	GAUSSIAN	MOVMEAN	MOVMEAN	LOESS
	0.075	0.091	0.061	0.043
Normalization	SNV			
	0.037	0.063	0.018	0.016
Smoothing + Normalization	**SGOLAY + SNV**	**MOVMEAN + SNV**	**GAUSSIAN + SNV**	**LOESS + SNV**
	0.035	**0.036**	**0.015**	**0.015**
Smoothing + Normalization + Spectral Derivative	MOVMEAN + SNV + FD			
	0.045	0.077	0.0423	0. 0432

Although there is a consensus on the best normalization method (SNV), results show discrepancies in the best smoothing algorithms when applied alone versus in combination with normalization. This is due to the small differences between methods (~0.001). Additionally, applying derivatives to the data does not improve the evaluation results. Fig. 7 illustrates, for

each medical HS system, the processed data providing the best results (their best smoothing following by SNV normalization without the further step of derivatives). The calibrated and reference data are also provided for qualitatively evaluation. It can be observed how after the processing step, the H_VNIR data still shows high levels of noise, contrasting with the notably improved performance of G_VNIR and N_NIR data. This might be due to the lower spectral sampling of the pushbroom VNIR sensor, as the window size is not optimized to effectively reduce its high-frequency noise peaks. Further analysis is required to optimize the window size, adapting it to different spectral sampling sensors.

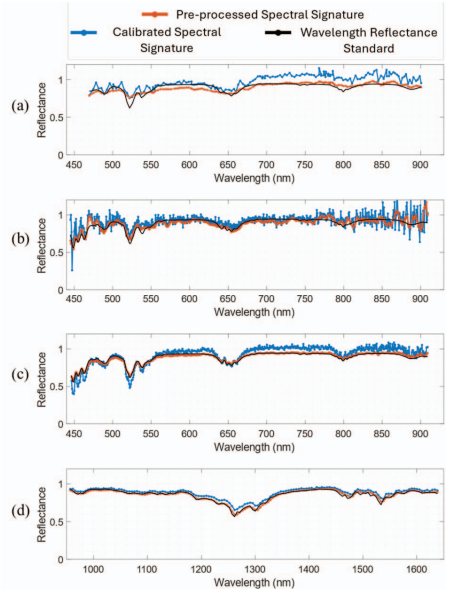


Fig. 7. Best pre-processed, smooth and normalized, spectral signature (orange) for each medical HS system, calibrated spectral signatures (blue) and wavelength reflectance standard (black): (a) G_VNIR, (b) H_VNIR, (c) N_VNIR, and (d) N_NIR.

IV. Conclusion

In this study, the need for effective processing methods in medical HSI data was addressed. The innovation lies in the generic optimization of spectral processing methods using a wavelength reflectance standard, instead of analysing a dataset for a specific application. The objective was to evaluate processing techniques to enhance spectral data quality and reliability by reducing noise and unwanted artifacts. First, a review of the state-of-the-art was performed, and various processing algorithms, focusing on smoothing, normalization, and spectral derivatives, were identified. Then, a wavelength reference standard was captured using four medical HS systems (G_VNIR, H_VNIR, N_VNIR and N_NIR). Each processing method was applied to the data recorded by the different medical HS systems. The performance of these processing methods was assessed using RMSE to compare processed data with the wavelength reference standard, and the execution time of each algorithm was considered to assess practical feasibility and efficiency.

Results highlight the effectiveness of certain processing techniques for each medical HS system. In terms of smoothing, the algorithms giving the best results were GAUSSIAN for G_VNIR, MOVMEDIAN for H_VNIR and N_VNIR and LOESS for N_NIR. The choice of smoothing algorithm varies across setups because of the different spectral sampling of HS camera employed. In terms of normalization, the SNV algorithm delivers optimal results across all medical HS systems, since it normalizes the scale and offset of the captured HS data which makes it more similar to the reference than the other algorithms. Furthermore, combining smoothing and normalization methods enhances performance, although different results are obtained for each medical HS system due to the non-optimized window size. Lastly, concerning spectral derivatives, FD performs better than SD, although none of them improves previous results since they enlarge the noise present on the data.

Regarding execution times, most algorithms are executed within the range of 10^{-2} to 10^{-3} s. However, RLOWESS and RLOESS, characterized as slow smoothing algorithms, exhibit longer processing times (~ 40 s). Nevertheless, future work will focus on analysing the performance of these algorithms on GPU or FPGA platforms in order to improve the algorithm performance, targeting real-time applications. Thus, the times obtained in different platforms allow to compare the proposed solutions against each other while providing a guideline for estimating the time required to process a given amount of data and for determining the appropriate algorithm for any particular application.

Despite the promising outcomes observed, our study has several limitations that require further attention. Concerning smoothing algorithms, optimizing the window size for noise reduction is necessary. Additionally, other processing algorithms should be considered to address spectral misalignment. Additionally, it is essential to acknowledge that the selection of evaluation metrics and benchmark datasets may impact result interpretation. Hence, further analysis using actual captured data is necessary to apply our findings to real-world HSI scenarios with increased complexity and variability.

Acknowledgment

This work was developed while Laura Quintana-Quintana, Carlos Vega, and Raquel Leon were beneficiaries of the pre-doctoral grant given by the “Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI)” of the “Consejería de Economía, Conocimiento y Empleo”, which is part-financed by the European Social Fund (FSE) (POC 2014-2020, Eje 3 Tema Prioritario 74 (85 %)). Moreover, this work is part of the TALENT (HypErsPEctRal Imaging for Artificial intelligence applications) project (PID2020-116417RB-C41), supported by the Spanish Government and European Union (FEDER Funds). This work has also been partially developed under the STRATUM project which received funding from the European Union’s Horizon Europe Programme HORIZON-IA action under grant agreement No. 101137416.

REFERENCES

[1] F. D. van der Meer *et al.*, "Multi- and hyperspectral geologic remote sensing: A review," *International Journal of Applied Earth Observation and Geoinformation*, vol. 14, no. 1, pp. 112–128, Feb. 2012, doi: 10.1016/j.jag.2011.08.002.

[2] E. Kho, B. Dashtbozorg, L. L. de Boer, K. K. Van de Vijver, H. J. C. M. Sterenborg, and T. J. M. Ruers, "Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information," *Biomed Opt Express*, vol. 10, no. 9, p. 4496, Sep. 2019.

[3] S. Ortega, H. Fabelo, R. Camacho, M. de la Luz Plaza, G. M. Callicó, and R. Sarmiento, "Detecting brain tumor in pathological slides using hyperspectral imaging," *Biomed Opt Express*, vol. 9, no. 2, p. 818, Feb. 2018, doi: 10.1364/BOE.9.000818.

[4] T. Collins *et al.*, "Automatic Recognition of Colon and Esophagogastric Cancer with Machine Learning and Hyperspectral Imaging," *Diagnostics*, vol. 11, no. 10, p. 1810, Sep. 2021, doi: 10.3390/diagnostics11101810.

[5] L. A. Zherdeva *et al.*, "Hyperspectral imaging of skin and lung cancers," J. Popp, V. V. Tuchin, D. L. Matthews, and F. S. Pavone, Eds., Apr. 2016, p. 98870S. doi: 10.1117/12.2227602.

[6] Åsmund Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, Nov. 2009, doi: 10.1016/j.trac.2009.07.007.

[7] J. Engel *et al.*, "Breaking with trends in pre-processing?," *TrAC Trends in Analytical Chemistry*, vol. 50, pp. 96–106, Oct. 2013, doi: 10.1016/j.trac.2013.04.015.

[8] J. Ezenarro, D. Schorn-García, O. Busto, and R. Boqué, "ProSpecTool: A MATLAB toolbox for spectral preprocessing selection," *Chemometrics and Intelligent Laboratory Systems*, vol. 247, p. 105096, Apr. 2024, doi: 10.1016/j.chemolab.2024.105096.

[9] I. Ruisánchez *et al.*, "Preliminary results of an interlaboratory study of chemometric software and methods on NIR data. Predicting the content of crude protein and water in forages," *Chemometrics and Intelligent Laboratory Systems*, vol. 63, no. 2, pp. 93–105, 2002, doi: https://doi.org/10.1016/S0169-7439(02)00039-4.

[10] D. Cozzolino, P. J. Williams, and L. C. Hoffman, "An overview of pre-processing methods available for hyperspectral imaging applications," *Microchemical Journal*, vol. 193, p. 109129, Oct. 2023, doi: 10.1016/j.microc.2023.109129.

[11] S. Singh and S. S. Kasana, "A Pre-processing framework for spectral classification of hyperspectral images," *Multimed Tools Appl*, vol. 80, no. 1, pp. 243–261, Jan. 2021, doi: 10.1007/s11042-020-09180-2.

[12] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.

[13] D. B. Bhushan, V. Sowmya, M. S. Manikandan, and K. P. Soman, "An effective pre-processing algorithm for detecting noisy spectral bands in hyperspectral imagery," in *2011 International Symposium on Ocean Electronics*, 2011, pp. 34–39. doi: 10.1109/SYMPOL.2011.6170495.

[14] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens Environ*, vol. 44, no. 2–3, pp. 127–143, May 1993, doi: 10.1016/0034-4257(93)90012-M.

[15] R. Koprowski, "Hyperspectral imaging in medicine: image pre-processing problems and solutions in Matlab," *J Biophotonics*, vol. 8, no. 11–12, pp. 935–943, Nov. 2015, doi: 10.1002/jbio.201400133.

[16] M. Witteveen, H. J. C. M. Sterenborg, T. G. van Leeuwen, M. C. G. Aalders, T. J. M. Ruers, and A. L. Post, "Comparison of preprocessing techniques to reduce nontissue-related variations in hyperspectral reflectance imaging," *J Biomed Opt*, vol. 27, no. 10, Oct. 2022, doi: 10.1117/1.JBO.27.10.106003.

[17] E. Kho *et al.*, "Feasibility of Ex Vivo Margin Assessment with Hyperspectral Imaging during Breast-Conserving Surgery: From Imaging Tissue Slices to Imaging Lumpectomy Specimen," *Applied Sciences*, vol. 11, no. 19, 2021, doi: 10.3390/app11198881.

[18] E. J. M. Baltussen *et al.*, "Hyperspectral imaging for tissue classification, a way toward smart laparoscopic colorectal surgery," *J Biomed Opt*, vol. 24, no. 01, p. 1, Jan. 2019, doi: 10.1117/1.JBO.24.1.016002.

[19] E. Kho *et al.*, "Hyperspectral Imaging for Resection Margin Assessment during Cancer Surgery," *Clinical Cancer Research*, vol. 25, no. 12, pp. 3572–3580, Jun. 2019, doi: 10.1158/1078-0432.CCR-18-2089.

[20] C. Vega *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)*, IEEE, Aug. 2022, pp. 863–870. doi: 10.1109/DSD57027.2022.00121.

[21] C. Vega García *et al.*, "Feasibility Study of Hyperspectral Colposcopy as a Novel Tool for Detecting Precancerous Cervical Lesions." May 2024. doi: 10.21203/rs.3.rs-4343232/v1.

[22] S. Ortega *et al.*, "Hyperspectral Superpixel-Wise Glioblastoma Tumor Detection in Histological Samples," *Applied Sciences*, vol. 10, no. 13, p. 4448, Jun. 2020, doi: 10.3390/app10134448.

[23] S. Ortega *et al.*, "Hyperspectral Imaging for the Detection of Glioblastoma Tumor Cells in H&E Slides Using Convolutional Neural Networks," *Sensors*, vol. 20, no. 7, p. 1911, Mar. 2020, doi: 10.3390/s20071911.

[24] L. Quintana-Quintana, S. Ortega, H. Fabelo, F. J. Balea-Fernández, and G. M. Callico, "Blur-specific image quality assessment of microscopic hyperspectral images," *Opt Express*, vol. 31, no. 8, p. 12261, Apr. 2023, doi: 10.1364/OE.476949.

[25] R. Leon *et al.*, "Evaluation of Hyperspectral Imaging Fusion for in-vivo Brain Tumor Identification and Delineation," in *2023 26th Euromicro Conference on Digital System Design (DSD)*, IEEE, Sep. 2023, pp. 493–499. doi: 10.1109/DSD60849.2023.00074.

[26] R. Leon *et al.*, "VNIR–NIR hyperspectral imaging fusion targeting intraoperative brain cancer detection," *Sci Rep*, vol. 11, no. 1, p. 19696, Oct. 2021, doi: 10.1038/s41598-021-99220-0.

[27] B. Boldrini, W. Kessler, K. Rebner, and R. W. Kessler, "Hyperspectral Imaging: A Review of Best Practice, Performance and Pitfalls for in-line and on-line Applications," *J Near Infrared Spectrosc*, vol. 20, no. 5, pp. 483–508, Oct. 2012, doi: 10.1255/jnirs.1003.

[28] G. Yang *et al.*, "The DOM Generation and Precise Radiometric Calibration of a UAV-Mounted Miniature Snapshot Hyperspectral Imager," *Remote Sens (Basel)*, vol. 9, no. 7, p. 642, Jun. 2017, doi: 10.3390/rs9070642.

[29] P. Geladi, J. Burger, and T. Lestander, "Hyperspectral imaging: calibration problems and solutions," *Chemometrics and Intelligent Laboratory Systems*, vol. 72, no. 2, pp. 209–217, Jul. 2004, doi: 10.1016/j.chemolab.2004.01.023.

[30] D. Nouri, Y. Lucas, and S. Treuillet, "Calibration and test of a hyperspectral imaging prototype for intra-operative surgical assistance," M. N. Gurcan and A. Madabhushi, Eds., Mar. 2013, p. 86760P. doi: 10.1117/12.2006620.

[31] A. Noviyanto and W. H. Abdulla, "Segmentation and calibration of hyperspectral imaging for honey analysis," *Comput Electron Agric*, vol. 159, pp. 129–139, Apr. 2019, doi: 10.1016/j.compag.2019.02.006.

[32] "Encapsulated Gray Scale Standards - Avian Technologies." Accessed: Jun. 03, 2021. [Online]. Available: https://aviantechnologies.com/product/encapsulated-gray-scale-standards/

[33] S. Ortega *et al.*, "Hyperspectral Push-Broom Microscope Development and Characterization," *IEEE Access*, vol. 7, pp. 122473–122491, 2019, doi: 10.1109/ACCESS.2019.2937729.

[34] L. Quintana *et al.*, "Instrumentation Evaluation for Hyperspectral Microscopy Targeting Enhanced Medical Histology," in *2021 XXXVI Conference on Design of Circuits and Integrated Systems (DCIS)*, IEEE, Nov. 2021, pp. 1–6. doi: 10.1109/DCIS53048.2021.9666188.

[35] W. S. Cleveland, "LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression," *Am Stat*, vol. 35, no. 1, p. 54, Feb. 1981, doi: 10.2307/2683591.

[36] W. S. Cleveland and S. J. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *J Am Stat Assoc*, vol. 83, no. 403, pp. 596–610, Sep. 1988, doi: 10.1080/01621459.1988.10478639.

[37] W. S. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots," *J Am Stat Assoc*, vol. 74, no. 368, pp. 829–836, Dec. 1979, doi: 10.1080/01621459.1979.10481038.

[38] Abraham. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures.," *Anal Chem*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964, doi: 10.1021/ac60214a047.

[39] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra," *Appl Spectrosc*, vol. 43, no. 5, pp. 772–777, 1989, doi: 10.1366/0003702894202201.

[40] P. Geladi, D. MacDougall, and H. Martens, "Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat," *Appl Spectrosc*, vol. 39, no. 3, pp. 491–500, May 1985, doi: 10.1366/0003702854248656.

[41] H. Martens, S. Jensen, and P. Geladi, "Multivariate Linearity Transformations for Near Infrared Reflectance Spectroscopy," *Applied Statistic*, pp. 205–234, 1983.

Chapter 4. Light - Histology Sample Interaction

While recent advancements in HS microscopy have focused primarily on sensor and system development, another critical factor influencing image quality is the preparation of the biological samples themselves. Histological analysis of biopsy samples is currently the gold standard for tumor detection and classification, relying on high-resolution imaging to distinguish cellular structures. To achieve optimal image sharpness in WSI microscopes, tissue sections are traditionally cut into thin slices, typically 3 to 5 μm thick, and stained with contrast-enhancing dyes such as H&E. However, in HS microscopy, obtaining diffuse transmittance from a sample, based on the principles of tissue-light interaction, may require different sample preparation techniques to acquire additional valuable spectral insights from biological tissues.

This chapter addresses objective **O4** of the Ph.D. thesis, which focuses on investigating how biological sample preparation, specifically tissue thickness, affects the spatial and spectral characteristics of HS microscopic images, and on proposing guidelines that improve image quality and diagnostic reliability. To fulfill these goals, a multi-phase approach was followed. First, experimental imaging was conducted on H&E-stained breast tissue sections prepared at 2 μm and 3 μm thicknesses. Based on H&E spectral transmittance peaks, HS images were segmented into stroma eosin-stained, nuclei hematoxylin-stained, and background unstained regions. Results showed that 3 μm samples not only provided a greater number of imaged cells but also exhibited improved spectral contrast, suggesting enhanced tissue-light interaction and better structure discrimination in HS data compared to thinner sections. On a second phase, a study was performed to explore sample thickness, beyond conventional histology protocols,

through computational analysis. A literature review revealed that approximately 40% of HS microscopy studies do not report tissue thickness, and among those that do, 98% use slices between 2 and 10 μm. To assess the influence of unstained tissue thickness on MS and HS image quality, a set of Monte Carlo light transport simulations was carried out across various tissue types (i.e., breast, liver, lung, and colorectal). The simulations demonstrated that thin samples yield poor spectral discrimination due to insufficient light-tissue interaction, while thicker samples approaching 500 μm enable improved tissue classification at the expense of light intensity.

The first approach was presented in the conference paper (**C2**) titled **"Analyzing Histology Hyperspectral Images, Does Tissue Thickness Matter"** presented in *SPIE Photonics Europe Conference* held in Strasbourg in 2024. It introduced the impact of sample thickness on spectral and spatial information quality. While the second approach was later developed during the Ph.D. Candidate's research stay at the Image-Guided Surgery Department (**M3**), The Netherlands Cancer Institute, in collaboration with experts in biomedical optics and cancer histology. The comprehensive results were published in the journal *Biomedical Optics Express* (**J3**) under the title **"Exploring the Role of Sample Thickness for Hyperspectral Microscopy Tissue Discrimination through Monte Carlo Simulations"**.

This chapter underlines the importance of interdisciplinary research in the development of reliable HS microscopic data, showing that technical advances in sensors and algorithms must be matched by optimized biological sample preparation. By addressing the often overlooked variable of tissue thickness, this work contributes to the establishment of improved protocols that promote accurate, reproducible, and clinically relevant HS imaging for medical diagnostics.

C2. Analyzing Histology Hyperspectral Images: Does Tissue Thickness Matter?

Title:	Analyzing Histology Hyperspectral Images: Does Tissue Thickness Matter?	
Authors:	Javier Santana-Nunez[¥], **Laura Quintana-Quintana**[¥], Himar Fabelo, Samuel Ortega, Esther Sauras-Colón, Noèlia Gallardo-Borràs, Daniel Mata-Cano, Carlos López-Pablo, and Gustavo M. Callico	
Conference:	SPIE Photonics Europe 2024	
Organizer:	SPIE	
Date: 07 - 11/04/2024	**Vol.**: 13006	**Pages**: 1-9
Place: Strasbourg, France		International
doi:	https://doi.org/10.1117/12.3017010	
Open Access, National Collaboration, Corresponding Author		

[¥]Equal Contribution

Analysing Histology Hyperspectral Images: Does Tissue Thickness Matter?

Javier Santana-Nunez[¥,*,1], Laura Quintana-Quintana[¥,*,1] , Himar Fabelo[2,1], Samuel Ortega[3,1,4], Esther Sauras-Colón[5,6], Noèlia Gallardo-Borràs[6], Daniel Mata-Cano[6], Carlos López-Pablo[5,6], Gustavo M. Callico[1].

[1]Research Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, Spain
[2]Fundación Canaria Instituto de Investigación Sanitaria de Canarias, Spain
[3]Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima), Tromsø, Norway
[4]University of Tromsø, Tromsø, Norway
[5]Dept. of Pathology, Hospital de Tortosa Verge de la Cinta, ICS, IISPV, Tortosa, Spain.
[6]Universitat Rovira i Virgili, Tortosa, Spain.
* jsnunez@iuma.ulpgc.es, lquintana@iuma.ulpgc.es
¥ These authors contributed equally to this work and are designated as co-first authors.

ABSTRACT

Cancer is one of the leading causes of death, thereby, contributing to their quick diagnosis or treatment is of greatest importance. Nowadays, tumours are mainly diagnosed and graded histologically using biopsies. Since the images need to be sharp to distinguish biological structures, samples are thinly sliced (3-5 µm) to avoid scattering and contrast is obtained using highly absorbance dyes (e.g., Haematoxylin and Eosin (H&E)). RGB (Red-Green-Blue) cameras have been widely employed to acquire those images, while new approaches, such as Hyperspectral (HS) Imaging (HSI), have been arising to obtain a greater amount of spectral information from the samples. However, in order to have diffuse light for the HS cameras to capture it, the thickness of the sample should be bigger than the ones employed in conventional microscopy. This work aims to characterize the influence of tissue thickness of histology breast samples sectioned at 2 and 3 µm on their spectral signatures. Based on the H&E transmittance spectra peaks, HS images were segmented into three structures: stroma (eosin-stained), nuclei (haematoxylin-stained), and background (non-stained). Results show that, spatially, in 3 µm samples there are more cells imaged than in 2 µm samples. Moreover, spectrally, 3 µm samples proportionate higher spectral contrast than 2 µm samples due the greater interaction of light with tissue, denoting them as more suitable for microscopic HSI.

Keywords- Hyperspectral Imaging, Breast Cancer, Histopathological Sample Thickness, Light Tissue Interactions.

1. INTRODUCTION

Breast cancer, a disease that annually afflicts over two million people, results in the death of more than half a million individuals each year [1]. This abnormal degree of mutated cells can be started by mammary cells (primary cancer) or by infiltrated cells into the mammary tissue (secondary cancer). Since the 1980s, the increase of early cancer detection programs in wealthy countries (e.g., United Kingdom) [1] shows a decrease in mortality, while in economically developing countries (e.g., Brazil) such tendency is not noticeable (Figure 1) [1]. These trends and bifurcations illustrate the necessity of more affordable early cancer detection programs.

Medical Hyperspectral Imaging (MHSI) is a field of Hyperspectral (HS) Imaging (HSI) that complements classic imaging techniques to obtain more data using the information given by interaction between light and biological samples [2], [3], [4]. The images obtained by HSI are called HS cubes. An HS cube is a tree-dimensional (two spatial and one spectral dimension) array of values where each one represents the irradiance of a pixel on a certain wavelength. The spectral signature is the vector of all the wavelengths of a single spatial pixel. A spectral signature is unique for each substance since light is absorbed or scattered differently by each kind of molecule depending on its chemical composition. This property allows the identification and classification of different materials [5]. The MHSI applied to cancer has shown promising results on the segmentation of normal and skin [6], breast [7], [8], and ovarian [9] cancerous tissue, among others.

Biomedical Spectroscopy, Microscopy, and Imaging III, edited by
Jürgen Popp, Csilla Gergely, Proc. of SPIE Vol. 13006, 1300611
 · 0277-786X · doi: 10.1117/12.3017010

Currently, the ground-truth for breast cancer diagnosis and grading is the histological study of biopsy samples extracted from the lumps detected in a patient's body. Histological tumour grade is based on the degree of differentiation of the morphological features of the tissue [10]. The breast biopsy samples are usually cut in thin slices (between 3 and 5 µm [11]) and dyed using Haematoxylin-Eosin (H&E) to highlight those differences and remark nuclei and stroma, which may point to abnormal cellular development. Afterwards, they may assess the identification of different biomarkers, (e.g., estrogenic receptors, progesterone receptors, HER2+, and Ki67) which may need the study of more samples stained with different immunochemistry [10]. In recent years, a new approach consisting in coupling HS cameras to microscopes have been researched [12], [13], [14]. This way we could extract specific information just from one slide from the patient biopsy.

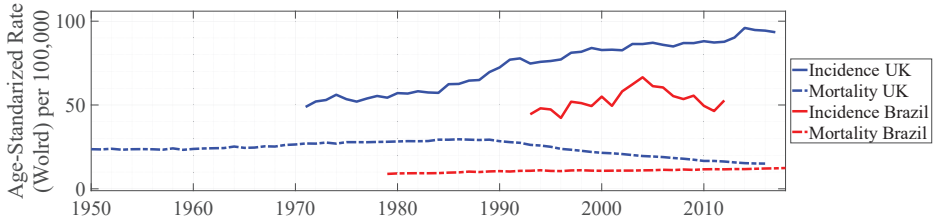


Figure 1. Incidence and Mortality evolution throughout the United Kingdom (UK) and Brazil [15].

When capturing breast biopsy samples under a RGB (Red-Green-Blue) microscope, the H&E dyes are the ones providing spatial contrast, since they absorb light in the Visible and Near Infrared (VNIR) range, from 450 to 700 nm (Figure 3c) [16]. Classic pathological approach is interested in obtaining sharp images; thus, it must reduce the scattering as much as possible. To do that the samples must be thinly sliced while maintaining the main morphological characteristics of the tissue studied [17], e.g., neurons are hundreds of thousands of times larger than red blood cells [18], so neurological biopsy samples should be sliced into thicker slides. In HS microscopy a different approach should be followed when slicing the samples since the main advantage of this technology is the capture of the diffused light obtained from the interaction of light with tissue. Following Beer-Lambert law, the thickness of the sample impacts its absorbance [19]. Since light in the VNIR range can penetrate the tissue up to 5 mm, samples should be thicker than the standard ones [20].

Previous studies have explored the influence of sample thickness on light penetration. However, these studies either used non-human samples [21], had thickness variation on the order of magnitude of the millimetres [22] (significantly higher than the traditional histology sample slide thickness and the intended thicknesses of this project), or their results were based on Monte Carlo simulations [20]. The main contribution of this work is the study of the influence of the conventional sample thicknesses over the captured diffuse light. To study it, breast samples are slices in two thicknesses (2 and 3 µm), captured with a microscopic HS system, and the spectral signatures of nuclei and stroma are analysed.

2. MATERIALS AND METHODS

2.1 Samples slides

A breast cancer biopsy from a patient was first fixed with formaldehyde and embedded into a paraffin block. Then two consecutively slides were extracted with two different thicknesses: 2 and 3 µm (Figure 2a and b). These slides were dyed using H&E and digitized using a Pannoramic 250 Flash III digital scanner (3D Histech Ltd., Budapest, Hungary). The digitized images allowed to annotate the areas where normal, Ductal Carcinoma in Situ (DCIS), or Infiltrating Ductal Carcinoma (IDC) tissue is present. The first slide has a thickness of 2 µm and contains 192 of those annotated areas, while the second slide has a thickness of 3 µm and has a larger number of annotated areas, 248 in total.

2.2 HS microscopic system and dataset

These biopsy sample slides were captured by a microscopic HS system (Figure 2c) [13], [14], [23]. This system is based on an OLYMPUS BX-53 microscope equipped with a 100 W T4Y4 Halogen Lamp from Olympus (Olympus, Tokyo, Japan) and a ×10 magnification lens (LMPLN-IR). In order to capture HS cubes, the pushbroom HS camera Hyperspec® VNIR A-Series (HeadWall Photonics, Fitchburg, MA, USA) is mounted. This camera is able to capture 826 bands within the spectral range of 400 to 1000 nm, allowing it to obtain a pixel size of 7.4 µm/pixels due to its 1004 spatial sensors. The dataset captured for this work is a set of four HS cubes of size 1004×1004×826. Due to the nature of the pushbroom camera, 1004 successive spatial lines were necessary to assemble each HS cube. The dataset is composed by two captures from the 2 µm slide and another two from the 3 µm slide, each with a field of view of 0.552 mm^2 of IDC tissue.

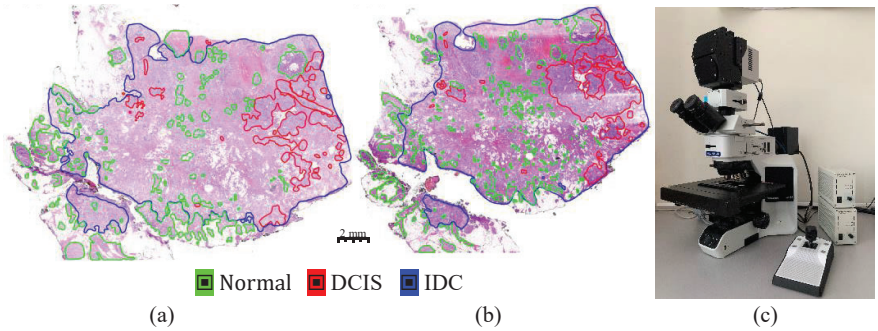


Figure 2. Materials used in this work: (a) 2 µm and (b) 3 µm biopsy slides, and (c) HS microscopic system.

2.3 Preprocessing chain

The HS cubes captured on transmittance are preprocessed in order to decrease the noise and are further manipulated to extract the absorbance and the absorption coefficient of the stroma and nuclei cellular structures. This preprocessing chain consists of: I) a flat-field correction and band reduction, II) segmentation of the cellular structures, III) filtering of outliers pixels, and IV) assembly of a synthetic RGB of each HS cube. In the following subsections, each preprocessing stage, developed using MATLAB R2022a, will be described.

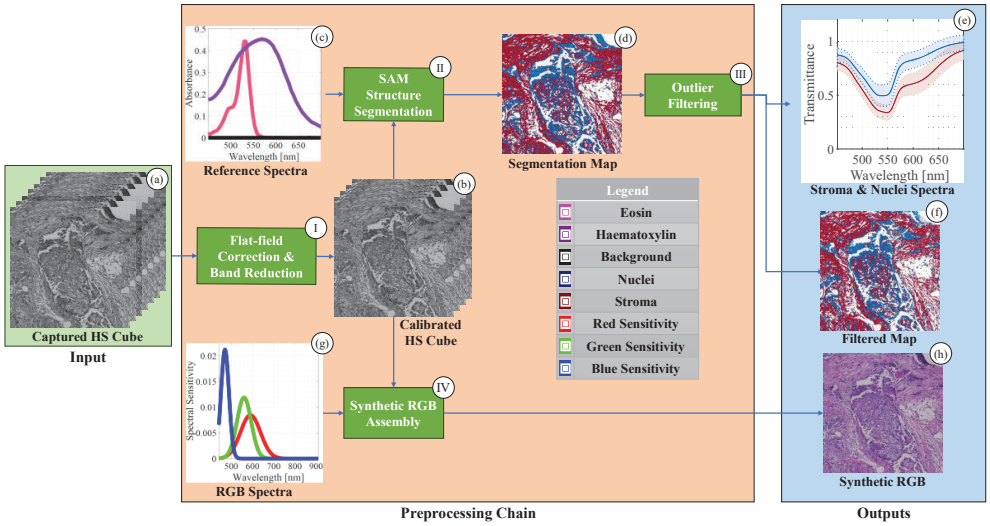


Figure 3. Preprocessing chain applied to each HS cube: (a) input HS cube, (b) calibrated HS cube, (c) absorbance reference signatures of haematoxylin, eosin, and background, (d) segmentation map, (e) stroma and nuclei filtered spectra, (f) filtered segmentation map, (g) red, green, and blue spectral sensitivity, and (h) synthetic RGB generated.

I. Flat-field correction and band reduction

The first step on the preprocessing chain is reducing the number of spectral bands of the captured HS cubes (Figure 3a). Previous works [24], [25] have shown that from the original 826 bands, only 645 bands are noise-free. From the selected bands, the H&E dyes have non-null absorbance values in the 450 to 700 nm range. So, to ensure that the signal use has the highest quality possible, the spectral range of this study has been narrowed down from 400 to 1000 nm to 450.685 to 699.516 nm (343 bands in total). Once the band reduction is performed, the flat-field correction of the remaining data is applied. In the microscopic framework used, the digital numbers reported by the camera need to be modify in order to correct the influence of the quantum efficiency and the background noise level of the sensors. The objective of the flat-field correction is to normalize the data into the 0 to 1 range, isolating the sample data from the influence of the camera

quantum efficiency, camera noise, and source light spectral non-uniformity [26]. This process consists of using a capture of the light used (*White*), a black level image (*Dark*) and a raw HS frame (*Raw*) in order to extract a calibrated HS cube (CC) where the transmittance values are stored.

$$CC = \frac{Raw - Dark}{White - Dark} \quad (1)$$

II. Spectral angle mapper structure segmentation

For the first step of the proposed methodology, a segmentation between haematoxylin-dyed (nuclei) and eosin-dyed (stroma and cytoplasm) tissue is needed [27]. Since cytoplasm represents a low percentage of each cell (10 - 40% in volume) [28] and for simplification of the argument, eosin-dyed pixels will be considered stroma. Moreover, the entire sample is not dyed, e.g., ducts where there is an absence of tissue or fat which is removed during the histological process, which is reflected as white areas on the image. From now on, these areas are considered as background. Thus, three different reference spectral signatures (RS_i) for each class (C_i) are considered: haematoxylin for nuclei, eosin for stroma, and pure light for the background (Figure 3c) [29].

The Spectral Angle Mapper (SAM) metric (2) allows to define a segmentation criterion. The SAM algorithm takes as inputs a reference spectral signature and a test spectral signature (TS) as q -dimensional vectors and outputs the angle between them [30]. A low SAM value indicates similarity between the reference and test spectral signatures. SAM values were obtained between the CC and the three references (*haematoxylin*, *eosin*, and *background*), each pixel was assigned to the class of the minimum achieved SAM (3). The result of this segmentation process is a nuclei and stroma spectra databases, that can be represented as segmentation maps where the colour of a pixel indicates its class (Figure 3d). Background pixels would not be included in later analysis since they are the absence of tissue.

$$SAM_{TS,RS_i} = cos^{-1}\left(\frac{\sum_k^q t_k r_k}{\sqrt{\sum_k^q t_k^2}\sqrt{\sum_k^q r_k^2}}\right); r_k \in RS_i; t_k \in TS \quad (2)$$

$$TS \in C_i : i = \underset{j}{\operatorname{argmin}}\, SAM_{TS,RS_j}, \forall\, j = \{\text{haematoxylin, eosin, background}\} \quad (3)$$

III. Outlier filtering

In order to reduce the variability of the nuclei and stroma spectra databases, the outlier pixels of each class are removed. A pixel (P) is considered an outlier if, at any band (b) from the studied spectrum (B), a value deviates more than three times the standard deviation (STD) above or below the mean value of the spectrum of each class at that band ($\overline{P_b}$) (4). The outliers pixels were removed from the later analysis. When the preprocessing chain is completed, the resulting spectral signatures (Figure 3e) can be used to extract the absorbance and absorption coefficients.

$$P \in Outlier \Leftrightarrow |P_b - \overline{P_b}| > 3 \cdot STD_b \forall\, b \in B \quad (4)$$

2.4 Results evaluation

I. Qualitative evaluation

A visual comparison between the HS cubes and the filtered segmentation maps (Figure 3f) can be done to evaluate the segmentation results qualitatively. Since not all frames from a HS cube can be displayed as an RGB image at the same time, a transformation is needed. A synthetic RGB image takes a spectral signature and simulates the sensitivities of the human eye cone cells (Figure 3g) to obtain the RGB values of each pixel [31]. With these synthetic RGB images (Figure 3h), the qualitative evaluation of the segmentation of the cellular classes can be performed.

II. Spatial difference

Once a successful segmentation has been confirmed, it is possible to evaluate the differences between the biopsy samples with thicknesses of 2 and 3 μm. As the sample thickness increases, the volume of tissue per sample is expected to increase too, and so, the number of cells per slide. This spatial difference between the several thicknesses can be obtained by calculating the nuclei to stroma pixel ratio (5) .

$$Nuclei/Stroma = \frac{\#Nuclei\ Pixels}{\#Stroma\ Pixels} \cdot 100\% \quad (5)$$

III. Spectral difference

Until now, the values captured, preprocessed, and stored on the *CC* were transmittance values. Transmittance (*T*) represents the ratio of light that has been able to get through a sample (*I*) in relation to the light emitted (I_0). However, its main issue is its non-linearity with the optical pathlength (*Z*) (6). Beer-Lambert law [19] states that the absorbance (*A*) is linearly directly related to the molar absorptivity (ϵ), the concentration of the sample (*c*) and the optical pathlength (7). Moreover, further analysis includes the comparison of the absorption coefficient (μ_a) of the different structures. The absorption coefficient is a measure of the capacity of a medium to limit the light that can pass through, per unit of optical pathlength [32], [33] (8). These resulting absorbance and absorption coefficient values allow to determine their contrast using equation (9) [34], where *X* is the mean absorbance or absorption coefficient values of a certain class and sample thickness.

$$T = \frac{I}{I_0} = e^{-\mu_a Z} \tag{6}$$

$$A = -\log_{10} T = \epsilon c Z \tag{7}$$

$$\mu_a = -\ln T / Z \tag{8}$$

$$Contrast = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 , \forall x_i \in X} \tag{9}$$

3. RESULTS

This section will present the results obtained following the previously exposed methodology. First, the results of the nuclei and stroma segmentation and its qualitative evaluation are shown. Then, the spatial and spectral differences between the captures of the 2 and 3 µm slides are presented.

3.1 Nuclei and stroma segmentation

First, using the dataset captured, the segmentation of nuclei, stroma, and background can be performed using the SAM metric (2) and the proposed segmentation criterion (3). In Figure 4, the synthetic RGBs, segmentation maps, and the mean and STD of spectral signatures of the nuclei and stroma pixels per capture are shown.

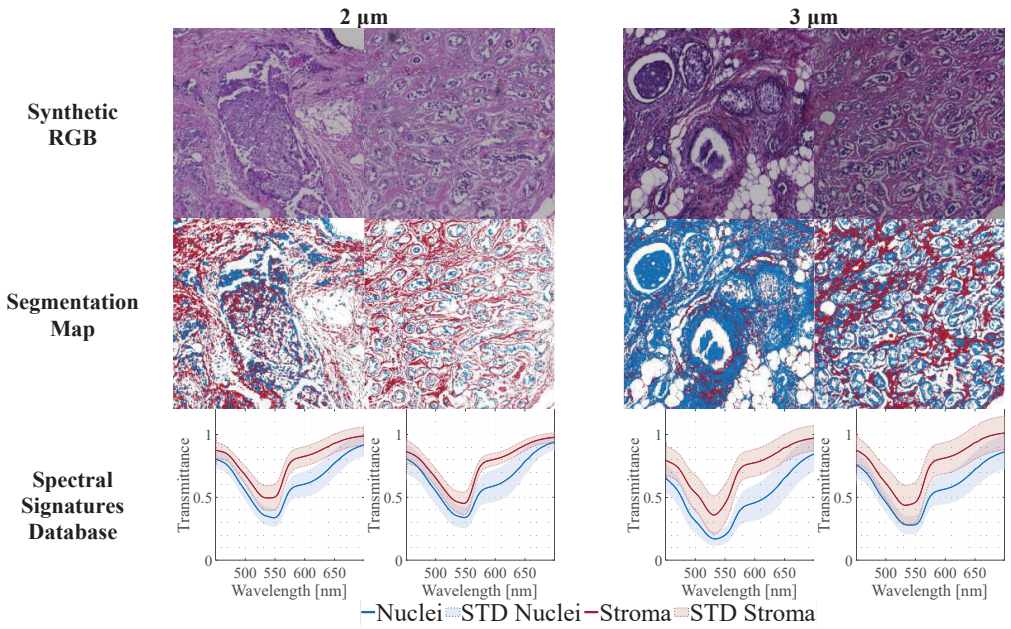


Figure 4. Synthetic RGB, Segmentation maps, and mean and STD of the spectral database of each HS cube.

3.2 Qualitative evaluation

Applying the previous segmentation maps to the synthetic RGB images, it is possible to generate a set of RGB masks where only pixels classified as nuclei or stroma are visible. Figure 5 shows that the stroma pixels are lighter than the nuclei ones, which correspond to the eosin and haematoxylin dyes, respectively (Figure 3d). This qualitative evaluation supports using H&E spectra as reference signatures to divide the HS cubes into their cellular structures.

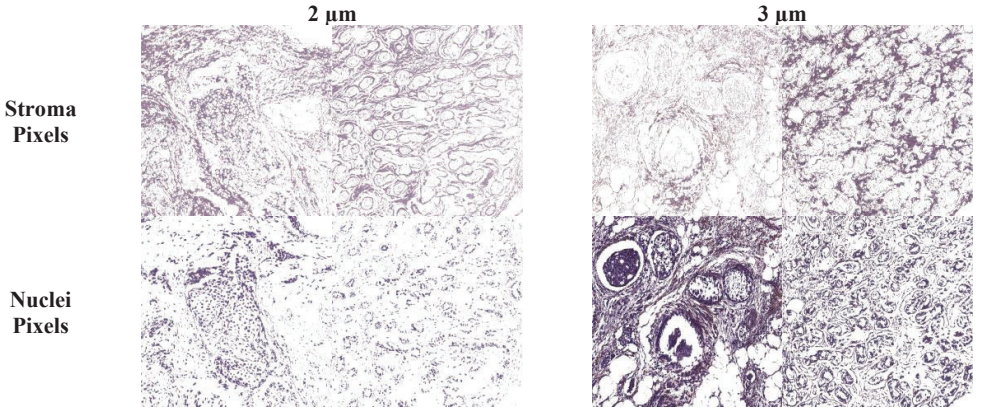


Figure 5. Stroma and Nuclei pixels on (a) 2 and (b) 3 µm HS captures.

3.3 Spatial difference

After successfully separating the stroma pixels from the nuclei pixels on each of the four images captured, we can compute the ratio of nuclei to stroma pixels. By applying equation (5) and adding up the pixel count for each class, the results shown in Figure 6 are obtained. The bar plots in Figure 6 reveal that the 2 µm captures contain fewer nuclei pixels compared to stroma pixels, with a ratio of 46.41 nuclei pixels for every 100 stroma pixels. However, the 3 µm captures exhibit a higher proportion of nuclei pixels relative to stroma pixels, with a ratio of 164.29 nuclei pixels for every 100 stroma pixels.

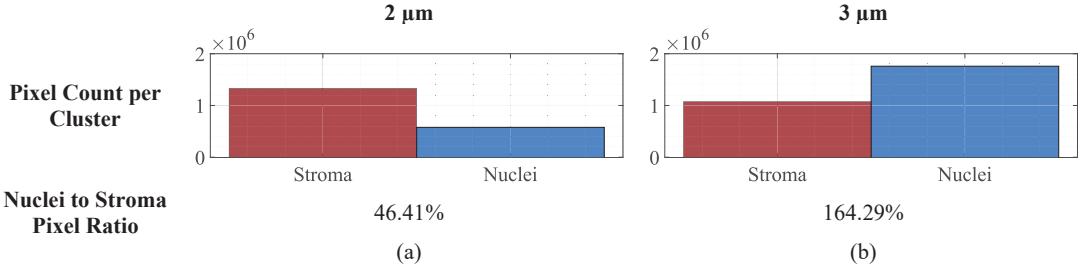


Figure 6 Count of Nuclei and Stroma pixels and ratio on (a) 2 and (b) 3 µm slides.

3.4 Spectral difference

Once the pixels corresponding to each cellular structure are identified, the absorbance and absorption coefficient of the different structures can be obtained using equation (7) and (8), respectively. A comparison between the absorbance and absorption coefficients of the different cellular structures and thicknesses is shown in Figure 7a and b, respectively. A quantitative evaluation of the contrast of each of these results is shown in Table 1 using equation (9). While the absorbance values of the 3 µm captures show a higher contrast, this effect is less appreciable on the absorption coefficient contrast values. The wider range of values of the 3 µm absorbance indicates its higher contrast between classes, making it more suitable for differentiating tissues in MHSI applications.

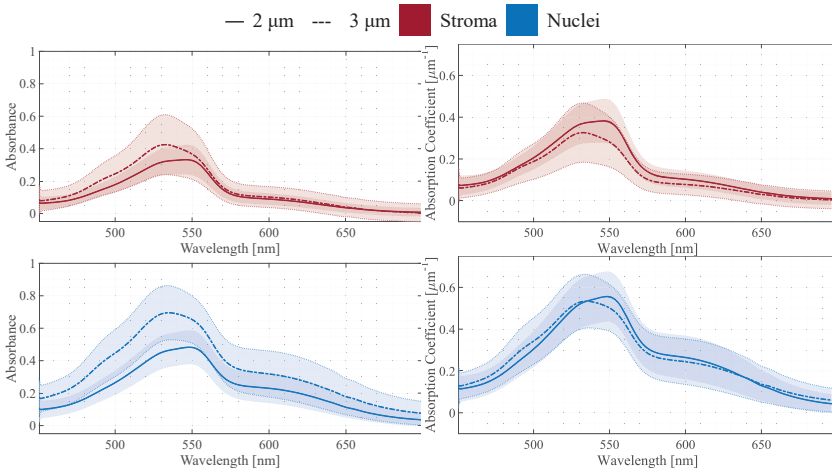


Figure 7. Mean and STD of (a) the absorbance and (b) the absorption coefficient of (top) stroma and (bottom) nuclei pixels in 2 and 3 µm slides.

Table 1 Contrast on the Absorbance and Absorption Coefficient mean values.

Cellular Structure	Thickness	Absorbance Contrast	Absorption Coefficient Contrast
Stroma	2 µm	0.1113	0.1281
	3 µm	**0.1411**	0.1083
Nuclei	2 µm	0.1523	0.1753
	3 µm	**0.2064**	0.1584

4. CONCLUSIONS

The recent decades have shown that early cancer detection programs are key to increase the survival rate. In this concern, the use of MHSI techniques in combination with microscopic systems are highly promising tools to approach histopathological diagnosis. In traditional histopathology, samples are prepared to optimize the RGB image quality based primarily on the morphological properties of the biological sample (samples are thin to enhance the clearness of the cellular structures in the visible spectrum). However, in MHSI these characteristics are key factors to obtain distinguished spectral signatures between pathological and normal tissue. As tissue thickness has been understudied for its effective selection in microscopic HS applications, this work aims to characterize its influence on histopathologic samples for microscopic HSI. A breast cancer biopsy sample was cut into 2 and 3 µm slides, dyed using H&E, and imaged to obtain four HS cubes. The nuclei and stroma pixels of each capture were segmented using the SAM metric and processed to obtain a transmittance spectra database of each cellular structure. These segmented pixels, in combination with the synthetic RGB image of each HS cube, allowed to qualitatively ensure that the pixels had been correctly categorized.

Spatially, the pixel count on each spectral database indicates how the stroma and nuclei are represented depending on the slide thickness. Results showed that the number of nuclei pixels in relation with the number of stroma pixels imaged in the 3 µm HS cubes is three times higher than in the 2 µm HS cubes, meaning that there is a higher number of cells being imaged. Spectrally, the segmented pixels allowed to obtain the absorbance values and absorption coefficients of the stroma and nuclei cellular structures. The higher contrast of the 3 µm absorbance, in comparison with the 2 µm ones, show that 3 µm samples absorb more light than 2 µm samples. This increase in the overall absorbed light with the tissue sample thickness aligns with the biphotonic literature that states that as the light pathlength increases, the probability of a photon to be absorbed is higher [33], [35]. Moreover, subtracting the influence of the tissue thickness the absorption coefficient shows that it is independent of the sample. In conclusion, the use of 3 µm biopsy sample slides is recommended over the 2 µm ones, as they show higher cell count and higher spectral contrast in the studied spectrum.

However, the main limitation of this work is the low variability on and the reduced number of the slides that have been studied, as they came from the same biopsy sample, the same type of tissue, and they were stained using the same dyes. In conclusion, these preliminary results open the discussion about the possibility of transforming the classic pathological workflow of fixing, cutting, and dyeing a biological sample, to be accommodated to use HS techniques in microscopy. Further research of other thickness is needed to stablish the optimal one for analysing histology HS images.

ACKNOWLEDGEMENT

This manuscript was developed within TALENT-HExPERIA (HypErsPEctRal Imaging for Artificial intelligence applications) project (PID2020-116417RB-C42) supported by the Spanish Government and European Union. It was completed while Javier Santana was beneficiary of the Investigo Program (Reference: 32/39/2022-0923131539) of the Canary Islands Employment Service ("Fondos del Plan de Recuperación, Transformación y Resiliencia" - Next Generation EU). Moreover, this work was developed thanks to the colaboration with the Netherlands Cancer Institute (NKI), while Laura Quintana was doing a research stay with the support of the "Consejería de Universidades, Ciencia e Innovación y Cultura" and the "Fondo Social Europeo plus". Laura Quintana was also beneficiary of the pre-doctoral grant given by the "Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI)" of the "Consejería de Economía, Conocimiento y Empleo", which is part-financed by the European Social Fund (FSE) (POC 2014-2020, Eje 3 Tema Prioritario 74 (85%)).

REFERENCES

[1] C. Coleman, 'Early detection and screening for breast cancer', *Semin Oncol Nurs*, vol. 33, no. 2, pp. 141–155, 2017, doi: 10.1016/j.soncn.2017.02.009.

[2] G. Lu and B. Fei, 'Medical hyperspectral imaging: a review', *J Biomed Opt*, vol. 19, no. 1, p. 010901, Jan. 2014, doi: 10.1117/1.JBO.19.1.010901.

[3] B. Fei, 'Hyperspectral imaging in medical applications', *Data Handling in Science and Technology*, vol. 32, pp. 523–565, Jan. 2019, doi: 10.1016/B978-0-444-63977-6.00021-3.

[4] R. Cui *et al.*, 'Deep Learning in Medical Hyperspectral Images: A Review', *Sensors 2022, Vol. 22, Page 9790*, vol. 22, no. 24, p. 9790, Dec. 2022, doi: 10.3390/S22249790.

[5] H. Fabelo *et al.*, 'Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations', 2018, doi: 10.1371/journal.pone.0193721.

[6] D. T. Dicker *et al.*, 'Differentiation of normal skin and melanoma using high resolution hyperspectral imaging', 2006, doi: 10.4161/cbt.5.8.3261.

[7] S. V. Panasyuk *et al.*, 'Medical hyperspectral imaging to facilitate residual tumor identification during surgery', *Cancer Biol Ther*, vol. 6, no. 3, pp. 439–446, 2007, doi: 10.4161/CBT.6.3.4018.

[8] S. Ortega *et al.*, 'Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images', *Proc SPIE Int Soc Opt Eng*, vol. 11320, p. 30, Mar. 2020, doi: 10.1117/12.2548609.

[9] T. E. Renkoski, K. D. Hatch, and U. Utzinger, 'Wide-field spectral imaging of human ovary autofluorescence and oncologic diagnosis via previously collected probe data', *J Biomed Opt*, vol. 17, no. 3, p. 036003, 2012, doi: 10.1117/1.JBO.17.3.036003.

[10] E. A. Rakha *et al.*, 'Breast cancer prognostic classification in the molecular era: the role of histological grade', *Breast Cancer Res*, vol. 12, no. 4, Aug. 2010, doi: 10.1186/BCR2607.

[11] M. Veta, J. P. W. Pluim, P. J. Van Diest, and M. A. Viergever, 'Breast cancer histopathology image analysis: A review', *IEEE Trans Biomed Eng*, vol. 61, no. 5, pp. 1400–1411, 2014, doi: 10.1109/TBME.2014.2303852.

[12] S. Ortega, M. Halicek, H. Fabelo, G. M. Callico, and B. Fei, 'Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review [Invited]', *Biomedical Optics Express, Vol. 11, Issue 6, pp. 3195-3233*, vol. 11, no. 6, pp. 3195–3233, Jun. 2020, doi: 10.1364/BOE.386338.

[13] S. Ortega *et al.*, 'Hyperspectral Imaging for the Detection of Glioblastoma Tumor Cells in H&E Slides Using Convolutional Neural Networks', *Sensors*, vol. 20, no. 7, p. 1911, Mar. 2020, doi: 10.3390/s20071911.

[14] S. Ortega *et al.*, 'Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images', *https://doi.org/10.1117/12.2548609*, vol. 11320, pp. 206–214, Mar. 2020, doi: 10.1117/12.2548609.

[15] 'Global Cancer Observatory'. Accessed: Feb. 16, 2024. [Online]. Available: https://gco.iarc.fr/en

[16] S. L. Gibbs *et al.*, 'Near-infrared fluorescent digital pathology for the automation of disease diagnosis and biomarker assessment', *Mol Imaging*, vol. 14, no. 4, Jun. 2015, doi: 10.2310/7290.2015.00005.
[17] C. Matenaers, B. Popper, A. Rieger, R. diger Wanke, and A. Blutke, 'Practicable methods for histological section thickness measurement in quantitative stereological analyses', 2018, doi: 10.1371/journal.pone.0192879.
[18] D. B. McMillan and R. J. Harris, 'An Atlas of Comparative Vertebrate Histology', *An Atlas of Comparative Vertebrate Histology*, pp. 1–604, Jan. 2018, doi: 10.1016/C2012-0-06909-8.
[19] I. Oshina and J. Spigulis, 'Beer–Lambert law for optical tissue diagnostics: current state of the art and the main limitations', *J Biomed Opt*, vol. 26, no. 10, Oct. 2021, doi: 10.1117/1.JBO.26.10.100901.
[20] C. Ash, M. Dubec, K. Donne, and T. Bashford, 'Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods', 2017, doi: 10.1007/s10103-017-2317-4.
[21] L. Kaub and C. Schmitz, 'Comparison of the Penetration Depth of 905 nm and 1064 nm Laser Light in Surface Layers of Biological Tissue Ex Vivo', *Biomedicines 2023, Vol. 11, Page 1355*, vol. 11, no. 5, p. 1355, May 2023, doi: 10.3390/BIOMEDICINES11051355.
[22] P. J. Muller and B. C. Wilson, 'An update on the penetration depth of 630 nm light in normal and malignant human brain tissue in vivo', *Phys Med Biol*, vol. 31, no. 11, p. 1295, 1986, doi: 10.1088/0031-9155/31/11/012.
[23] S. Ortega *et al.*, 'Hyperspectral Push-Broom Microscope Development and Characterization', *IEEE Access*, vol. 7, pp. 122473–122491, 2019, doi: 10.1109/ACCESS.2019.2937729.
[24] R. Leon *et al.*, 'VNIR–NIR hyperspectral imaging fusion targeting intraoperative brain cancer detection', *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–12, Oct. 2021, doi: 10.1038/s41598-021-99220-0.
[25] H. Fabelo *et al.*, 'In-Vivo Hyperspectral Human Brain Image Database for Brain Cancer Detection', *IEEE Access*, vol. 7, pp. 39098–39116, 2019, doi: 10.1109/ACCESS.2019.2904788.
[26] B. Boldrini, W. Kessler, K. Rebner, and R. W. Kessler, 'Hyperspectral Imaging: A Review of Best Practice, Performance and Pitfalls for in-line and on-line Applications', *J Near Infrared Spectrosc*, vol. 20, no. 5, pp. 483–508, Oct. 2012, doi: 10.1255/JNIRS.1003.
[27] J. K. C. Chan, 'The wonderful colors of the hematoxylin-eosin stain in diagnostic surgical pathology', *Int J Surg Pathol*, vol. 22, no. 1, pp. 12–32, Feb. 2014, doi: 10.1177/1066896913517939.
[28] A. T. Molines *et al.*, 'Physical properties of the cytoplasm modulate the rates of microtubule polymerization and depolymerization', *Dev Cell*, vol. 57, no. 4, pp. 466-479.e6, Feb. 2022, doi: 10.1016/J.DEVCEL.2022.02.001.
[29] T. Biswas, H. Suzuki, M. Ishikawa, N. Kobayashi, and T. Obi, 'Generative adversarial network based digital stain conversion for generating RGB EVG stained image from hyperspectral H&E stained image', *J Biomed Opt*, vol. 28, no. 05, May 2023, doi: 10.1117/1.JBO.28.5.056501.
[30] S. Chakravarty, B. K. Paikaray, R. Mishra, and S. Dash, 'Hyperspectral Image Classification using Spectral Angle Mapper', *Proceedings of 2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering, WIECON-ECE 2021*, pp. 87–90, 2021, doi: 10.1109/WIECON-ECE54711.2021.9829585.
[31] D. Cao and P. A. Barrionuevo, 'Chronobiology International The Journal of Biological and Medical Rhythm Research Estimating photoreceptor excitations from spectral outputs of a personal light exposure measurement device', 2014, doi: 10.3109/07420528.2014.966269.
[32] C. Hu, F. E. Muller-Karger, and R. G. Zepp, 'Absorbance, absorption coefficient, and apparent quantum yield: A comment on common ambiguity in the use of these optical concepts', *Limnol Oceanogr*, vol. 47, no. 4, pp. 1261–1267, 2002, doi: 10.4319/LO.2002.47.4.1261.
[33] M. H. Freeman, *Optics / M.H. Freeman, C. C. Hull.*, 11th ed. 2003.
[34] E. Peli, 'Contrast in complex images', *J. Opt. Soc. Am. A*, vol. 7, no. 10, 1990.
[35] E. Chlipala *et al.*, 'Optical density-based image analysis method for the evaluation of hematoxylin and eosin staining precision', *J Histotechnol*, vol. 43, no. 1, pp. 29–37, Jan. 2020, doi: 10.1080/01478885.2019.1708611.

J3. Exploring The Role of Sample Thickness for Hyperspectral Microscopy Tissue Discrimination Through Monte Carlo Simulations

Title:	Exploring The Role of Sample Thickness for Hyperspectral Microscopy Tissue Discrimination Through Monte Carlo Simulations		
Authors:	**Laura Quintana-Quintana**, Mark Witteveen, Behdad Dashtbozorg, Samuel Ortega, Theo J.M. Ruers, Henricus J.C.M. Sterenborg, and Gustavo M. Callico		
Journal:	Biomedical Optics Express		
Publisher:	Optica Publisher Group		
Date: 2025	**Vol.**: 16	**Pages**: 4644-4661	
doi:	https://doi.org/10.1364/BOE.563094		
JIF Quartile:	Q1 in Radiology, Nuclear Medicine & Medical Imaging	**IF**:	3.2
Open Access, International Collaboration, Corresponding Author			

Biomedical Optics EXPRESS

Exploring the role of sample thickness for hyperspectral microscopy tissue discrimination through Monte Carlo simulations

LAURA QUINTANA-QUINTANA,[1,*] MARK WITTEVEEN,[2] BEHDAD DASHTBOZORG,[2] SAMUEL ORTEGA,[1,3] THEO J. M. RUERS,[2] HENRICUS J. C. M. STERENBORG,[2] AND GUSTAVO M. CALLICO[1]

[1] *Institute for Applied Microelectronics (IUMA), University of Las Palmas de Gran Canaria (ULPGC), Spain*
[2] *Image-Guided Surgery, Department of Surgery, The Netherlands Cancer Institute, Amsterdam, Netherlands*
[3] *Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima), Tromsø, Norway*
[*] *lquintana@iuma.ulpgc.es*

Abstract: Recent advancements in multispectral (MS) and hyperspectral (HS) microscopy have focused on sensor and system improvements, yet sample processing remains overlooked. We conducted an analysis of the literature, revealing that 40% of studies do not report sample thickness. Among those that did report it, the vast majority, 98%, used 2–10 µm samples. This study investigates the impact of unstained sample thickness on MS/HS image quality through light transport simulations. Monte Carlo simulations were conducted on various tissue types (i.e., breast, colorectal, liver, and lung) using optical property parameters extracted from the literature. The simulations revealed that thin samples reduce tissue differentiation, while higher thicknesses (approximately 500 µm) improve discrimination, though at the cost of reduced light intensity. Although the results are based on idealized conditions and exclude certain real-world factors such as sample variability and instrument-specific effects, they highlight the need to study and optimize sample thickness for enhanced tissue characterization and diagnostic accuracy in MS/HS microscopy.

1. Introduction

Multispectral (MS) and hyperspectral (HS) imaging (MSI/HSI) are gaining increasing interest in medical applications due to the enhanced information they provide. These techniques capture diffuse spectra that arise from the interaction of light with tissues, which is unique since it depends on the inherent optical properties of the tissue itself. The difference in absorption (μ_a) and scattering (μ_s) coefficients of each tissue attenuates light differently across various wavelengths, providing distinct spectral signatures [1]. Specifically, MSI/HSI in microscopic histological analysis allows spatial and spectral examination of biological specimens in detail [2]. Some applications include the diagnosis of diabetic condition via retinal imaging [3], Alzheimer's disease biomarker identification in plasma samples [4], the detection of head and neck squamous cell carcinoma on histologic slides [5], or the classification of cholangiocarcinoma from pathology images [6].

MS/HS microscopy is typically performed by attaching a MS or HS camera to a bright-field microscope [7–9], enabling the capture of spectral features from tissues at a fine scale, such as mammalian cells (5 to 25 µm) or red blood cells (6 to 8 µm) [10]. While much of the development of these systems has focused on improving sensors and imaging technology [11–14], sample preparation has mostly followed the traditional methods used in histology analysis. Biopsies are either embedded in paraffin or frozen before being sliced into thin sections using a microtome or cryostat, respectively. These sections are typically cut to a thickness of 2 to 10 µm, matching

#563094 https://doi.org/10.1364/BOE.563094
Journal © 2025 Received 16 Apr 2025; revised 15 Sep 2025; accepted 18 Sep 2025; published 24 Oct 2025

Biomedical Optics EXPRESS

the scale of individual cells and allowing for the spatial differentiation of structures within the sample [10].

However, in MS/HS microscopy, special attention must be given to sample preparation, as MS/HS technology relies on the interaction of light with the sample, and light behaves differently at small tissue thicknesses compared to bulk tissue. According to diffusion theory, which approximates light propagation in highly scattering media such as tissue, the mean free path, defined as $l_s = 1/(\mu_a + \mu_s)$, represents the average distance a photon travels through tissue before its direction (scattering) or energy (absorption) is significantly altered [15]. In bulk tissue, the average distance travelled by photons before being absorbed or detected is higher or much higher than the mean free path (diffuse regime). However, in microscopy, if the thickness of a tissue slice is so short that does not exceed one mean free path, the detected photons will have nearly the same direction and energy as the emitted ones (sub-diffuse regime). In tissues, the mean free path typically ranges from 10 to 1000 µm, with around 100 µm being common in the visible spectrum [16]. As a result, when tissue is sliced to a thickness of 2–10 µm for conventional microscopy analysis, if captured without further processing (i.e., unstained), the detected light would be almost identical to the emitted light, resulting in low contrast and making the sample appear almost transparent to the human eye.

Thus, in bright-field microscopy, contrast enhancement of biological samples is performed using dyes (exogenous chromophores), such as hematoxylin and eosin (H&E) staining [17]. Staining techniques are designed to ensure that each structure of the sample absorbs different dyes, such as hematoxylin for the cell nuclei and eosin for the extracellular matrix and cytoplasm. Since each dye absorbs light at distinct wavelengths, contrast is created on the sample [18]. These samples are then examined either visually or by capturing RGB (Red, Green and Blue) images. Nonetheless, staining alters the intrinsic optical properties of tissues, confining spectral contrast to the specific dyes used, which are identical across all samples. As a result, MSI/HSI techniques cannot benefit from the spectral contrast provided by the endogenous chromophores inherent to each tissue type and human being.

Therefore, according to light propagation in tissue, to fully utilize MS/HS microscopy in the broadest context the focus should be on capturing these endogenous chromophores of tissues, which provide richer spectral data not only in the visible spectrum (VIS) but also in the near-infrared range (NIR). While a conventional tissue thickness of around 5 µm may serve this purpose, it may not always be optimal. Increasing thickness provides more detailed spectral signatures, until a certain point where light penetration is reduced, resulting in lower light intensity. To fully leverage the capabilities of MS/HS microscopy it is crucial to pay special attention to sample thickness, balancing the ability to differentiate cells on a sample (2-10 µm) with the challenge of capturing meaningful spectra from the sample itself (~100 µm). Selecting an optimal thickness ensures enough scattering and absorption events to enhance contrast while preserving the visualization of tissue structures.

This study aims to evaluate the impact of sample thickness in MS/HS microscopy spectral data and to establish a benchmark for optimizing tissue thickness in future studies. We begin with a comprehensive review and analysis of the literature to identify the sample thicknesses previously used in MS/HS microscopy studies. Next, a virtual specimen will be designed using a Monte Carlo (MC) Light Transport Simulator [19,20] to further investigate the effects of different sample thicknesses over the diffuse spectra obtained. Spectral signatures of normal and lesioned tissues (breast, colorectal, liver, and lung) at different sample thicknesses will be simulated using optical property parameters obtained from published studies. No new experimental hyperspectral data will be collected in this work. Finally, spectral contrast will be assessed by evaluating the accuracy of each thickness in distinguishing between normal and lesioned tissue. Thus, although the results represent an idealized theoretical investigation, since they do not account for sample heterogeneity, preparation artifacts, or instrument-specific effects beyond those simulated, this

Biomedical Optics EXPRESS

work still provides practical insights and guidelines for improving spectral imaging protocols in future studies.

2. State of the art

In the field of medical imaging, the integration of MS/HS microscopy has emerged as a powerful technique for real-time non-ionizing diagnostics. Ortega *et al.*, conducted a systematic review on MS/HS microscopy [2] using the following query: (Hyperspectral OR Multispectral) AND (histology OR pathology OR histopathology). From 2004-2019, 1776 documents were retrieved, and a rigorous selection process chose 192 documents that met the eligibility criteria established by the authors. Of those, 85 were based on the application of MS/HS microscopy technology for diagnosing various diseases. Building upon this foundation, we undertook an update to the review, extending the search period from August 13th, 2019, to December 12th, 2023. This yielded 759 additional documents, among which 100 met the eligibility criteria established by Ortega *et al.* (humans or mammals samples captured using MSI, HSI, or near-infrared (NIR) sensors coupled to an optical microscope) [2]. These works were meticulously scrutinized to identify and document the specific tissue thickness employed in the different MS/HS microscopy applications.

Tissue thickness was not reported as a parameter in 40% of the eligible studies, highlighting the limited attention given to this crucial aspect in existing literature. For the studies that did include this information (60 documents), the distribution of reported values is shown in Fig. 1. Notably, 5 µm is the most frequently documented tissue thickness, accounting for 28% of the studies specifying this parameter, since it is the thickness used for routine paraffin sections [21]. Similarly, another 28% reported a thickness of 10 µm, which corresponds to the maximum cutting capacity of most microtomes. Of these, 4 articles belong to the same research group, investigating normal and diabetic retina of rats [22–24]. They selected 10 µm-thick serial sections that passed through the optic nerve head and dyed them with hematoxylin and eosin (H&E) for HS microscopic examination. Chipala et al. [25] demonstrated that optical density (inversely proportional to transmittance) in H&E stained slides increases with tissue thickness (e.g., rising from 0.2 to 0.5 as thickness ranges from 2 to 10 µm), revealing the impact of thickness on optical properties. A notable exception is the study by Pertzborn *et al*. [26], which utilized 12 µm-thick frozen sections. These are slightly thicker than standard due to cryotome limitations. HSI data from these sections was used to train a 3D convolutional neural network for detecting oral squamous cell carcinoma. In summary, this analysis of the state of the art on MS/HS microscopy highlights the need for further research into how tissue thickness impacts light transmission and spectral contrast for MS/HS microscopy applications, emphasizing its importance for accurate imaging and analysis.

Biomedical Optics EXPRESS

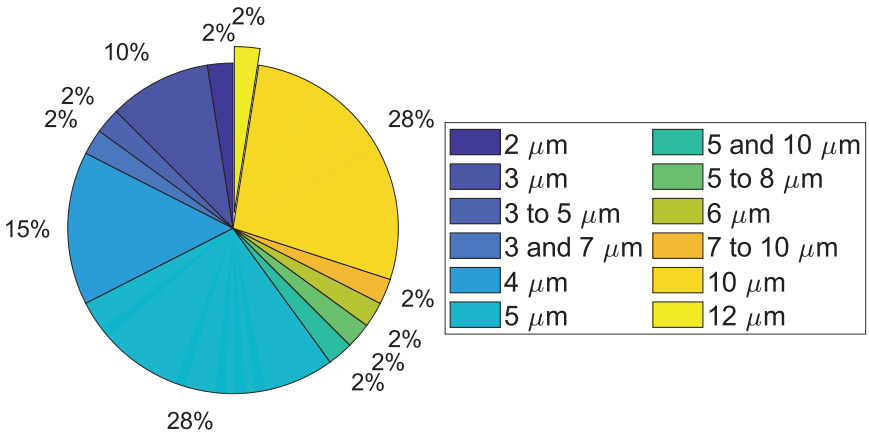


Fig. 1. Tissue thickness distribution in MS and HS microscopy studies from the state-of-the-art, with 5 µm and 10 µm being the most represented thicknesses at 28% each.

3. Materials and methods

In this section, the MC simulation framework is discussed to examine tissue-light interactions at different thicknesses in both normal and lesioned tissues (specifically breast, lung, colorectal, and liver). To achieve this, a comprehensive analysis of the instrumentation and optics of the HS microscopy system is conducted, along with an investigation of the optical properties of the tissues (absorption, scattering, scattering phase function, and refractive index). Additionally, spectral similarity metrics will be introduced to assess the similarity of the spectral signatures simulated for normal and lesioned tissues at a certain thickness. This approach will help to identify the thickness that most effectively distinguishes normal and lesioned tissue.

3.1. Microscopic HS system

The objective of this study is to investigate the influence of sample thickness on spectral signatures. Since there is no standard MS/HS microscope, we will focus on an HS microscope based on a widely used configuration [13,11]. A HS sensor (i.e., the Hyperspec VNIR -Visual and Near Infrared- A-Series HS camera -HeadWall Photonics, MA, USA-) integrated with a conventional bright-field microscope (i.e., the Olympus BX-53 -Olympus, Tokyo, Japan-) has been used. The system transmits light through a collimator to achieve uniform radiance of the sample, commonly referred to as Köhler illumination. Moreover, the light goes through one of the four objective lenses (10×, Numerical Aperture = 0.3 and Working Distance = 18 mm, HS sensor pixel pitch = 7.4 µm and thus spatial resolution = 0.74 µm) optimized for infrared (IR) observations. Figure 2 depicts a schematic of the tissue placement and the light propagation path through the tissue.

To eliminate the impact of the acquisition system on the HS images, a flat-field correction process is commonly applied to the raw data using both white and dark reference images [8]. In transmittance mode, a region of the transparent glass slide without specimen is used to capture the white reference spectrum (*White Ref*), while the dark reference spectrum (*Dark Ref*) is acquired by turning off the light source. The transmittance pixel (*Transmittance Pixel*) is then calculated using Eq. (1), with the raw pixel (*Raw Pixel*) as the initial data captured by the sensor. Through proper calibration, the transmission effects of the glass slide are compensated for. Consequently, in these simulations, the glass is not considered. Furthermore, the dark reference spectrum is not

Biomedical Optics EXPRESS

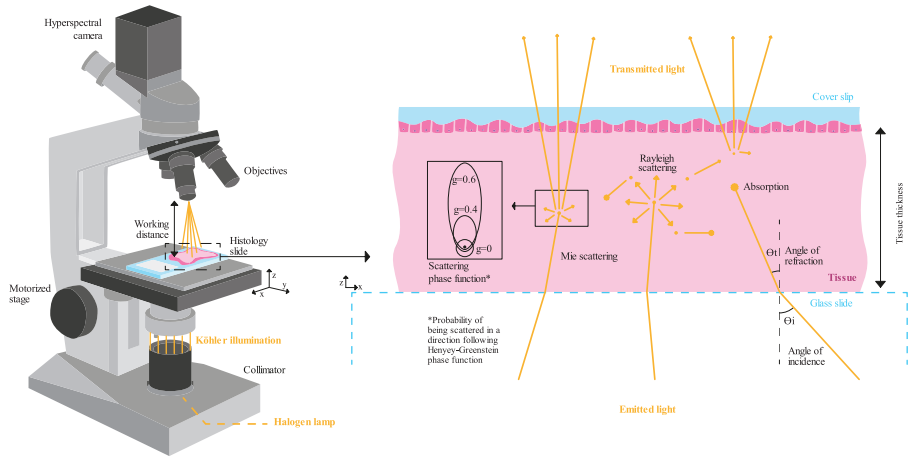


Fig. 2. Schematic of microscopic HS system capturing a histology slide in transmittance mode (left) and light-tissue interactions within the tissue being analyzed (right).

simulated, as sensor noise is not being simulated.

$$Transmittance\ Pixel = \frac{Raw\ Pixel - Dark\ Ref}{White\ Ref - Dark\ Ref} \tag{1}$$

3.2. Tissue optical properties

The next step in designing the MC framework is to define the optical properties of the tissues (more information in the Supplement 1):

- The absorption coefficient (μ_a) was computed, following Jaques [27], as a weighted sum of the main chromophores (oxygenated hemoglobin (μ_a.oxy), deoxygenated hemoglobin (μ_a.deoxy), water (μ_a.water), and fat (μ_a.fat)) according to Eq. (2):

$$\mu_a = BS\mu_{a.oxy} + B(1-S)\mu_{a.deoxy} + W\mu_{a.water} + F\mu_{a.fat} \tag{2}$$

  Here, B is the blood volume fraction, S is the hemoglobin oxygen saturation, W is the water content, and F is the fat content. Minor absorbers such as melanin, bilirubin and β-carotene were excluded due to their minimal concentration in the tissues under study [27].

- The reduced scattering coefficient (μ'_s) was modeled as a combination of Rayleigh and Mie scattering effects, with wavelength dependence relative to a reference wavelength $\lambda_{Ref} = 500\ nm$. The formulation is given in Eq. (3):

$$\mu'_s = a\left(f_{Ray}\left(\frac{\lambda}{\lambda_{Ref}}\right)^{-4} + (1-f_{Ray})\left(\frac{\lambda}{\lambda_{Ref}}\right)^{-b_{Mie}}\right) \tag{3}$$

  where a is a scaling factor, v is the Rayleigh contribution weight, and b_{Mie} is the Mie scattering exponent.

- The scattering phase function describes the angular distribution of scattered photons. While the standard Henyey-Greenstein (HG) function, defined in Eq. (4), is commonly

Biomedical Optics EXPRESS

used:

$$p_{HG}(g,\theta) = \frac{1}{4\pi}\frac{1-g^2}{(1+g^2+2g\cos\theta)^{3/2}} \tag{4}$$

we employed the more accurate Two-Term HG model, Eq. (5), to better capture forward and backward scattering in thin tissue samples:

$$p(\alpha, g_f, g_b, \theta) = \alpha \cdot p_{HG}(g_f,\theta) + (1-\alpha) \cdot p_{HG}(g_b,\theta) \tag{5}$$

Here, α is the weight of forward scattering, and $p_{HG}(g_f,\theta)$ and $p_{HG}(g_b,\theta)$ are the anisotropy parameters for forward and backward scattering, respectively.

- The refractive index (n) was fixed at 1.35, consistent with values reported for similar biological tissues, and defined in (6) and (7):

$$n = \frac{c}{v} \tag{6}$$

$$n = \frac{\sin\theta_i}{\sin\theta_r} \tag{7}$$

where c is the speed of light in vacuum, v in the medium, and $\sin\theta_i$ and $\sin\theta_r$ are the angles of incidence and refraction, respectively.

3.2.1. Optical properties of tissue under evaluation

A comprehensive review of the state-of-the-art literature pertaining to the eight tissues under investigation (breast, lung, colorectal, and liver, in both pathologies, lesioned and healthy) was conducted to extract their optical properties. No additional experimental hyperspectral data were acquired for this study. Absorption and scattering parameters for different tissues were extracted from several literature sources. According to Eq. (2), the key parameters for calculating absorption are the percentages of blood (B), fat (F), and water (W), as well as blood oxygen saturation (S). For scattering calculations, the required parameters include the reference wavelength (λRef), Rayleigh scattering fraction (f_Ray), and Mie scattering parameters (b Mie and α). For breast tissue, data from L. L. de Boer et al. [28] were used. Fat content was estimated as 90% of the tissue volume excluding blood, with the remaining 10% assigned to water. Lung tissue parameters were obtained from J. W. Spliethoff et al. [29], who directly provided blood and water percentages; fat content was then calculated as the residual fraction. All values for colorectal tissue were sourced from M. S. Nogueira et al. [30]. For liver tissue, parameters came from N. Reistad et al. [31], reporting fat as 10.1% of the total volume excluding blood, with water making up the remaining 89.9%. Scattering phase function parameters (forward scattering (gf), backward scattering (gb), and anisotropy (α)) were also taken from multiple studies: N. Ghosh et al. [32] for breast, R. Marchesini et al. [33] for lung, A. N. Bashkatov et al. [34] for colorectal, and both P. Saccomandi et al. and R. Marchesini et al. [35,33] for liver.

Mean values of absorption, scattering, and phase function optical properties are presented in Table 1, with their standard deviations available in Table S1 of the Supplement 1. Fig. 3 illustrates these mean values along with their corresponding standard deviations for the tissues under evaluation. It is important to note that, similar to regular tissue, simulated tissue on microscopy slides includes both fat and water, but these values might be different after histopathologic processing or frozen sectioning, where some water is lost. While these values are crucial in real-life scenarios for thin tissue slicing, they are included here to simulate pure, freshly tissue excised, and not processed tissue.

Biomedical Optics EXPRESS

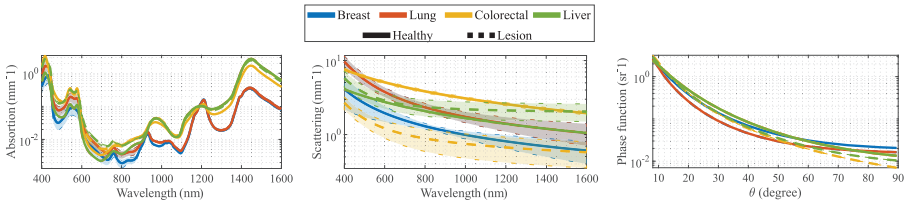


Fig. 3. Optical properties of tissues (breast, lung, colorectal, and liver) in both normal and lesioned states: (a) absorption coefficient, (b) scattering coefficient, and (c) scattering phase function.

Table 1. Optical Properties parameters for normal and lesioned samples of breast, lung colorectal, and liver tissues[a]

		Absorption					Scattering			R	Scattering Phase Function			R
		B (%)	S (%)	F (%)	W (%)	λRef (nm)	b Mie	f_Ray	a		gf	gb	α	
Breast	**N**	1,75	16*	88,43	9,83	800	0,69*	5	14	[28]	0,87	-0,09	0,82	[32]
	L	3,5	41*	86,85	9,65	800	0,69*	15	25		0,92	-0,06	0,84	
Lung	**N**	7,5	90	35,5	57	800	0,98*	0*	38	[29]	0,84	-0,5	0,92	[33]
	L	2,5	69	3,5	94	800	0,98*	0*	21		0,84	-0,5	0,92	
Colorectal	**N**	5	64	7,7	78,5	500	0,2	56	14,3	[30]	0,89	0	1	[34]
	L	6,6	66	1,4	84,9	500	0,5	40	15,9		0,89	0	1	
Liver	**N**	8,5	50	9,24	82,26	800	0,35	29	13,7	[31]	0,88	-0,46	0,93	[35,33]
	L	7,2	50	9,37	83,43	800	0,05	10,7	18,2		0,88	-0,46	0,93	

[a]N: Normal tissue, L: Lesioned tissue, B: Blood, S: Saturation, F: Fat, W: Water, R: Reference. *Extracted from [27].

3.3. MC simulation framework

To investigate light–tissue interactions under controlled and reproducible conditions, a custom MC simulation framework was developed. This framework was developed to simulate the HS imaging system described in Section 3.1, producing synthetic radiance data based on well-defined tissue conditions. This approach enables systematic analysis of how key parameters, such as tissue thickness, affect the resulting spectral signatures. An overview of the complete simulation pipeline is presented in Fig. 4.

3.3.1. Simulated volume

The MC simulation was set up to emulate a homogeneous single-layer sample with a spatial extent of 100×800 µm. To evaluate the influence of tissue thickness on the spectral response, he height of the simulated volume was varied across several depths: 5, 20, 50, 100, 200, 500, 1000, and 2000 µm. These values were selected to span from typical histological sections (5 µm) to depths where light transmittance becomes negligible due to increased tissue attenuation (2000 µm). However, simulating thick samples at high spatial resolution can be computationally intensive. To manage this, the voxel size and the number of voxels were adjusted according to the tested thickness:

- For 5 and 20 µm thicknesses, a voxel size of 1 µm was used, yielding volumes of $[5, 20] \times 100 \times 800$ voxels (depth, lines, pixels).

Biomedical Optics EXPRESS

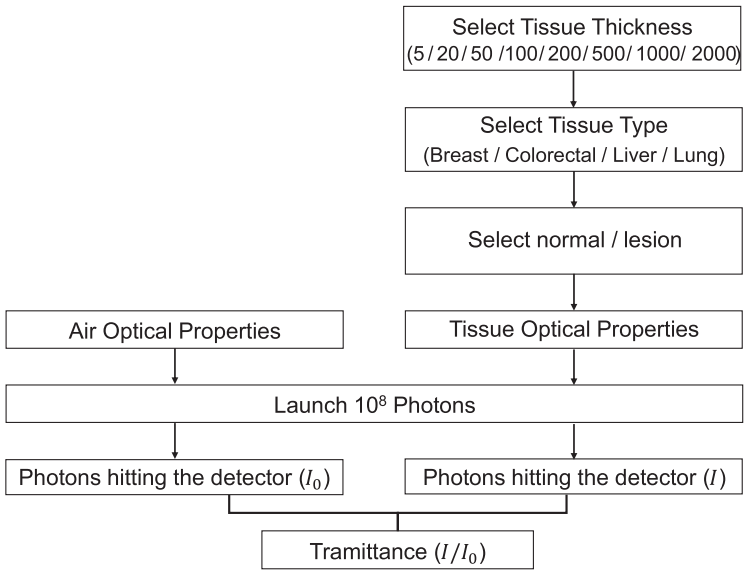


Fig. 4. Framework followed to perform the different MC simulations obtaining the spectral signatures of various organ tissues, at two tissue types at several tissue thicknesses.

- For 50, 100, and 200 µm thicknesses, a voxel size of 10 µm was used, resulting in volumes of $[5, 10, 20] \times 10 \times 80$ voxels.
- For thicker samples (500, 1000, and 2000 µm), a voxel size of 100 µm was used, leading to volumes of $[5, 10, 20] \times 1 \times 8$ voxels.

This adaptive resolution strategy ensured that the number of simulated voxels remained within a manageable range, while still allowing for meaningful comparisons across different thicknesses.

3.3.2. Simulated tissue

For each tissue type (breast, lung, colorectal, and liver) and its corresponding pathological states (normal, lesioned), the mean and standard deviation of the optical properties were collected from the literature (see Table 1 from Supplement 1). To introduce variation within each tissue type, and assuming a normal distribution, ten random values were selected for each tissue–pathology combination. In total, 80 biological tissues were simulated (10 per tissue–pathology combination).

3.3.3. Simulated spectra

Replicating the HS microscopy setup described in Section 3.1, each MC simulation covered the 400–1000 nm spectral range. A total of 105 spectral bands were simulated, corresponding to a distance between bands of 5.7 nm. While different number of bands with different spectral resolutions could have been selected to better represent a specific MS or HS application, 105 bands were chosen to provide a balanced and representative sampling of the spectral range defined.

3.3.4. Flat-field correction

Flat-field correction is a critical preprocessing step in HS imaging, used to compensate for non-uniformities in illumination, sensor response, and optical path variations [36,37]. To simulate

Biomedical Optics EXPRESS

the flat-field reference (I_0), the MC simulation was run without any tissue in the optical path, emulating the acquisition of an empty region on a microscope slide. Although photons were propagated through air (assumed to have negligible absorption and scattering) not all were expected to reach the detector. Instead, the simulation estimated the number of photons passing through the sensor slit at the detector's working distance (18 mm).

Next, the optical properties were replaced with those of biological tissue (e.g., breast tissue), and the corresponding radiance spectra (I) were generated across the wavelength range. Flat-field correction was then applied according to Eq. (8) by dividing, at each wavelength, the number of photons transmitted through the tissue by the ones detected in the absence of a sample. This process yielded normalized tissue transmittance values.

It is important to note that, since all simulations were conducted on spatially homogeneous volumes and no spatial variation was modeled, spatio-spectral calibration (typically required in real-world HS imaging systems) was not necessary in this study.

$$T = \frac{I}{I_0} \tag{8}$$

3.3.5. Software and hardware specifications

A total of 67,200 simulations were performed (8 tissue thickness × 80 biological tissues × 105 spectral bands), with each iteration simulating 10^8 photons. Due to the high computational demand, MC simulations were executed using Monte Carlo eXtreme (MCX) version 2023.11, developed by Qianqian Fang [38–40], which supports GPU acceleration. Simulations were run through MATLAB R2023a on a workstation equipped with an Intel Xeon Silver 4216 (16-core) CPU, 128 GB of RAM, and three NVIDIA Tesla T4 (TU104GL) GPUs.

3.4. Data augmentation

To establish a realistic database, in alignment with real-world samples (e.g. including the sensor signal-to-noise ratio, wavelength miscalibration, improper light source thermal management, or intensity fluctuation due to changes in the light source), data augmentation was performed over the simulated spectral signature database. The procedure outlined in [41] was followed, where each stage of the process builds upon the previous one.

First, two wavelength miscalibration processes were simulated. The spectra's wavelength range experienced reduction or extension by a random value ranging from -1 to 1 nm, with the spacing between bands adjusted in concordance. For instance, if the original vector spanned from 400 to 1000 nm with 100 bands and now is extended from 400 to 1001 nm, the spacing between bands is augmented from 6 to 6.01 nm. The second wavelength miscalibration technique was to shift the wavelength vector within the range ± 4.8 nm. This would represent measurements taken on different days, which assessed the shift in light spectra. Finally, the spectra were interpolated to the original wavelength vector to have all the data in the same wavelength vector.

Following these adjustments, two spectra intensity errors were introduced simulating fluctuations in light source intensity [41]. The first error involved wavelength-dependent intensity fluctuations, where the spectrum's intensity was randomly tilted following a linear shape, by up to 5% of the spectral signature mean. For instance, a spectral signature with a mean 0.8 would potentially be scaling the left tail up to 0.84 and the right tail down to 0.76. The second intensity error involves a constant intensity shift of up to 5% of the spectra mean.

In the final step, noise was introduced to the spectra to replicate the sensor's intrinsic noise floor. First, the dataset was expanded by interpolating five additional points between each simulated band pair, increasing the spectral resolution from 105 to 521 bands. This expansion provided a larger number of data points on which the noise could be applied. Additive white Gaussian noise was then applied to simulate a signal-to-noise ratio (SNR) of 35 dB, simulating the noise of a real HS microscopy system. Figure 5 visually represents the original spectral signature, which refers

Biomedical Optics EXPRESS

to the one obtained using the MC simulation framework, and each stage of the data augmentation process.

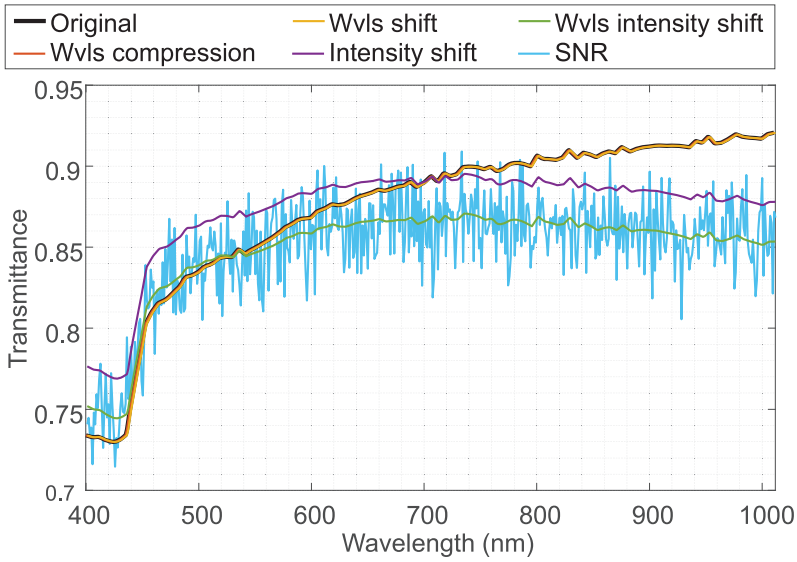


Fig. 5. Data augmentation applying different strategies to the original spectral signature database. The original transmittance spectrum is altered through wavelength (wvls) compression, wavelength shift, intensity shift, wavelength intensity shift, and finally white Gaussian noise is added.

3.5. Spectral evaluation metrics

Several spectral evaluation metrics were employed to assess the differences between the two spectral signatures. Let t be the test spectra, r the reference spectra and N be the number of bands, the following metrics are as follow:

3.5.1. Euclidean distance

The Euclidian distance (d) quantifies the distance between two vectors in an N-dimensional space. It is determined by Eq. (9), which computes the quadratic mean of variances between test and reference values. The Euclidean distance is bounded between 0 and 1 (for t and r in the [0, 1] range), with a value close to zero indicating a high similarity between the spectral signatures.

$$d = \sqrt{\sum_{i=1}^{N} (t_i - r_i)^2} \tag{9}$$

3.5.2. Spectral angle mapper

The Spectral Angle Mapper (SAM) assesses spectral similarity by computing the angle between spectra, treated as vectors in an N-dimensional space [42]. It is calculated using (10). This metric represents angles, so it ranges from 0 to $\pi/2$, when both, t and r are non-negative. A smaller

Biomedical Optics EXPRESS

SAM value indicates greater spectral similarity.

$$SAM = \arccos\left(\frac{\sum_{i=1}^{N} t_i r_i}{\sqrt{\sum_{i=1}^{N} t_i^2}\sqrt{\sum_{i=1}^{N} r_i^2}}\right) \tag{10}$$

3.5.3. Normalized spectral similarity score

The Normalized Spectral Similarity Score (NS^3) is a method for evaluating spectral similarity by considering both Euclidean and SAM distances between spectra. It offers a novel approach to spectral comparison by addressing potential misidentifications arising from ambiguous high-confidence scores in spectral amplitude and angle. Following (11), NS^3 normalizes spectral amplitudes and applies a custom function to match the spectral angle and amplitude difference scores, enhancing accuracy. Taking into account the ranges of d and SAM, this metric ranges from 0 to $\sqrt{2}$, where a low NS^3 score indicates a strong correspondence between the test and reference signatures.

$$NS^3 = \sqrt{d^2 + (1 - cos(SAM))^2} \tag{11}$$

3.5.4. SID

Spectral Information Divergence (SID) is an information-theoretic spectral measure introduced to assess the dissimilarity between the spectral signatures of two pixels in HS images [43]. It is developed based on the concept of divergence, specifically measured in the probabilistic behavior between the spectral characteristics of the compared pixels (12). SID values lie in the range $[0, \infty)$, where a smaller SID indicates low divergence between the spectral signatures of the compared pixels.

$$SID = \sum_{i=1}^{N} t_i log\left(\frac{t_i}{r_i}\right) + \sum_{i=1}^{N} r_i log\left(\frac{r_i}{t_i}\right) \tag{12}$$

3.5.5. SID-SAM

The SID-SAM mixed measure, calculated according to Eq. (13), leverages the strengths of both SID and SAM in spectral discriminability [44]. This implies that the spectral similarity and dissimilarity achieved through the mixed measure are significantly improved by multiplying the spectral capabilities of both measures. The choice of tangent or sine over cosine is made to compute the perpendicular distance between two vectors, rather than calculating the projection of one vector along the other. A SID-SAM score close to 0 signifies a robust match between the test signature and the reference signature and tends to infinity when the maximum dissimilarity is reached.

$$SID(TAN) = SID \times \tan(SAM) \tag{13}$$

3.6. Quantitative thickness evaluation

The previously presented evaluation metrics were calculated to measure the difference between inter-class (normal (N) vs lesion (L)) and intra-class (N vs N and L vs L) spectra. Figure 6(a) shows the Euclidean distance of inter-class (orange histograms) and intra-class (blue histograms) liver spectra. To obtain a metric that yields consistent results within the same class while emphasizing differences for spectra from different classes, the disparity between the inter-class and intra-class histograms for a given thickness serves as a qualitative indicator of the discrimination capability associated with that thickness.

However, a quantitative metric is needed to measure this discrimination. At each thickness, a threshold (th) can be defined to categorize the inter and intra-class distances as belonging to inter or intra-class evaluations. The accuracy metric can then be calculated over the true positive,

Biomedical Optics EXPRESS

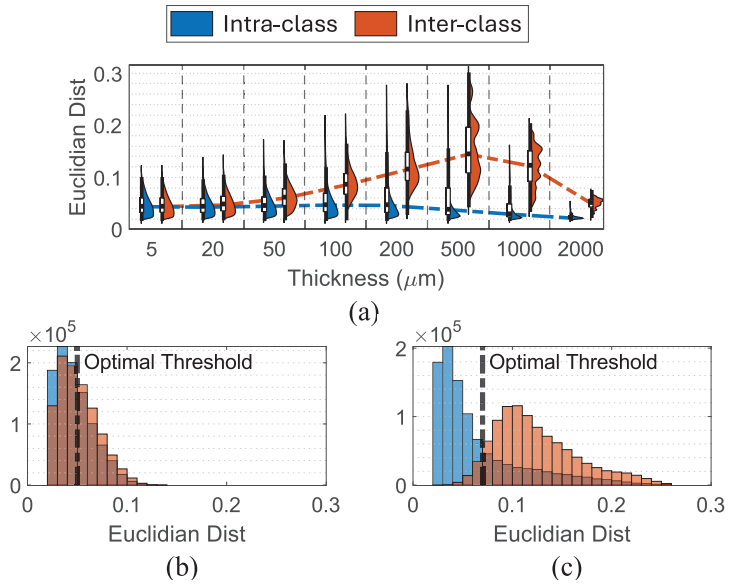


Fig. 6. (a) Euclidean distance scores between intra-class (blue) and inter-class (orange) spectra at different tissue thicknesses for liver tissue. Zoom over the histograms at (b) 20 µm and (c) 200 µm, red line showing threshold with maximum accuracy for each case.

true negative, false positive, and false negative values. Under a specified *th*, the accuracy of the classification is determined by (14), where each bin_i represents the individual values within a histogram. Since spectra within a class are more similar to each other (intra-class distances) than to the spectra of the other class (inter-class distances), intra-class distances should be lower than the inter-class ones. Therefore, the distributions are ordered such that values below the *th* are classified as intra-class, while values above it are classified as inter-class, providing most of the time accuracy values between 0.5 and 1. In scenarios where histograms overlap, the selection of any *th* would yield a substantial number of misclassifications, indicating a constrained discriminative capacity in these thicknesses (e.g., for Fig. 6(b) the threshold providing maximum accuracy (0.54) was found at an Euclidean distance of 0.05). Conversely, in thicker samples, where histograms manifest greater separation, establishing a *th* becomes more straightforward (e.g., for Fig. 6(c) the threshold providing maximum accuracy (0.83) was found at an Euclidean distance of 0.07).

The accuracy was calculated across all possible *th*s for a single set of histograms, covering the lowest to the highest obtained metric scores, with an increment of one bin at a time. The highest accuracy value obtained across all thresholds was chosen to be the discriminative power of that thickness. Subsequently, the methodology was repeated for all combinations of tissue type (breast, colorectal, liver, and lung), metric (Euclidean distance, NS3, SAM, SID, SIDSAM), and thickness (5, 20, 50, 100, 200, 500, 1000, and 2000 µm). This process allowed us to identify, at each tissue and metric, the thickness that yielded the best accuracy between lesion and normal

Biomedical Optics EXPRESS

tissue.

$$ACC = P(intra - class)\sum_{i=1}^{th} P(bin_i|intra - class) + P(inter - class)\sum_{i=th}^{end} P(bin_i|inter - class) \quad (14)$$

4. Results

In this section, the results obtained following the previous methodology are presented. The MC simulated dataset after data augmentation is shown in Fig. 7 for four of the studied thicknesses (all thicknesses shown in Fig. S1 in the Supplement 1), where green and red spectra correspond to normal and lesioned tissue, respectively.

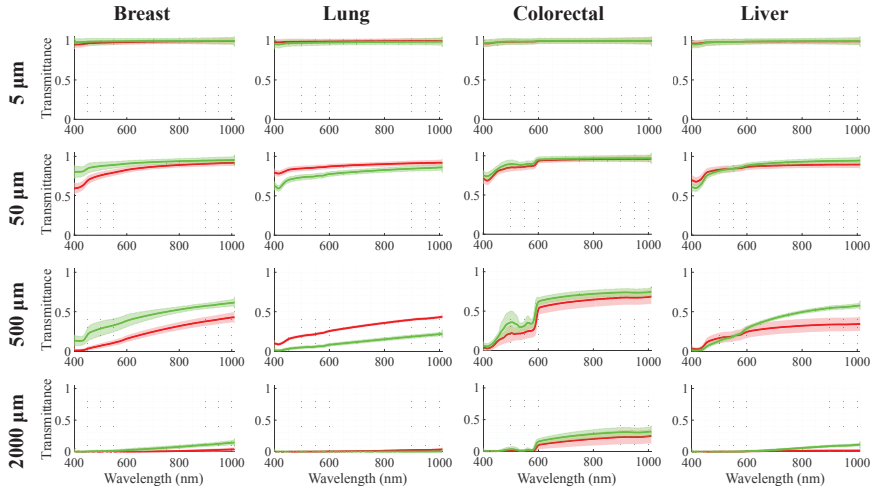


Fig. 7. Spectral signatures for normal (green) and lesion (red) tissues for each type of tissue and four slide thickness

As anticipated, tissue samples with a thickness of 5 µm exhibit flat transmittance spectra, with values approaching 1. This suggests that there are no distinct peaks associated with the absorption or scattering of endogenous chromophores within the tissue. These results confirm that samples thinner than the diffusion length (6 to 12 mm [45]) do not experience significant scattering or absorption. At this thickness, there is no notable difference in the spectral signatures of normal and lesioned tissues. However, as the sample thickness increases, the distinction between normal and lesioned tissue becomes more apparent. By 50 µm, differences in spectral signatures between tissue types start to emerge, indicating that conventional sample thicknesses of 2 to 10 µm may not provide sufficient contrast for pathology discrimination with HS microscopy. At 500 µm, the contrast between tissue types is most pronounced, although the mean transmittance drops below 0.5, a value that could be even lower in practical applications due to instrumental attenuation. Beyond this thickness, light penetration diminishes significantly, reducing the signal that reaches the sensor. The maximum simulated thickness of 2000 µm is well beyond the cutting capacity of standard microtomes, which are typically limited to around 60-70 µm. At these thick samples (>2000 µm), the spectral signatures appear nearly flat due to minimal light transmission through the sample, resulting in transmittance values approaching zero across all tissues and pathologies. This results in a lack of spectral resolution at this thickness.

Biomedical Optics EXPRESS

After simulating the entire dataset, the evaluation of the data was performed using the metrics described in the 3.5 Spectral Evaluation Metrics section. Inter- and intra-class distances were calculated using Euclidean distance, SAM, NS3, SID, and SID-SAM (see Fig. S2 from Supplement 1). Subsequently, in accordance with the methodology described in the 3.6 Quantitative Thickness Evaluation, the aim was to assess the discrimination power of each tissue thickness. This evaluation focused on identifying which tissue can be accurately distinguished between normal and lesioned tissues (inter-class spectra) while also accurately classifying them as equivalent when they belong to the same class (intra-class spectra). Figure 8 illustrates the accuracy results for each tissue type and metric.

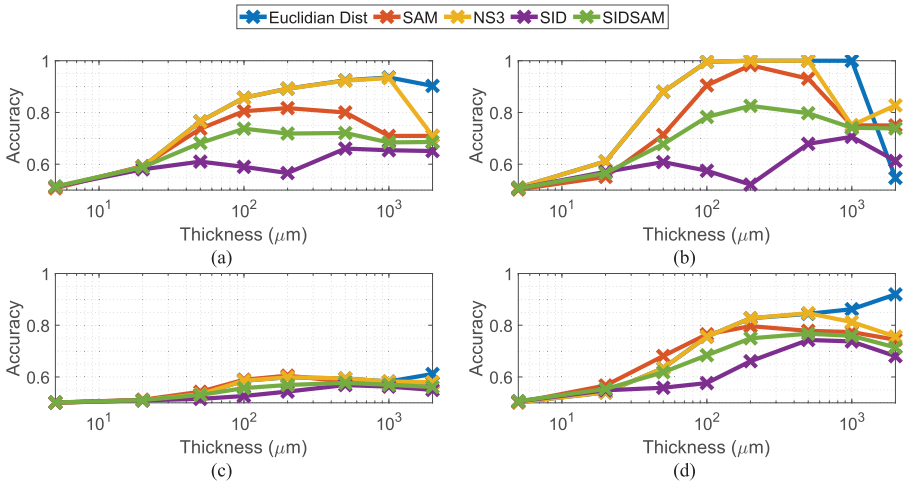


Fig. 8. Spectral difference between tumor and normal tissues fusing (a) breast, (b) lung, (c) colorectal and (d) liver. Thickness is represented on the x-axis on a logarithmic scale.

Different metrics exhibit varying levels of discrimination due to their inherent characteristics (e.g., SAM measures the relative difference between spectral signatures, while Euclidean distance is an absolute metric). However, all metrics consistently show that a thickness of 5 µm offers no discrimination between pathologies on unstained samples (accuracy around 0.5). The optimal discrimination thickness was determined by first identifying the thickness that yielded the highest accuracy for each evaluation metric individually. A majority voting approach was then applied to select the thickness at which the greatest number of metrics reached their maximum performance. This approach is metric-agnostic, as not all metrics may be equally relevant for a given application, allowing flexibility in selecting the most appropriate metrics based on the specific requirements of the intended use case. Based on this method, the optimal discrimination thicknesses were found to be 1000 µm for breast tissue, 200 and 500 µm for lung and colorectal tissues, and 500 µm for liver tissue.

5. Discussion

In recent years, there has been a growing adoption of MS/HS microscopic systems. However, it is essential to exercise caution in their application. In traditional histology, samples typically range from 2 to 10 µm in thickness, which is thinner than the diffusion length in tissue (6 to 12 mm [45]). Given the absence of absorption and scattering in these thin samples, dyeing becomes necessary to produce image contrast. However, employing sample dyeing in MS/HS microscopic imaging is suboptimal as the final goal is to capture information regarding the interaction of light

Biomedical Optics EXPRESS

with intrinsic biomarkers present in tissue (endogenous chromophores). An in-depth analysis of the current state-of-the-art was conducted to find out the tissue thickness utilized in previous MS/HS microscopic studies. From this review, it was found that 60% of the documents did not report the tissue thickness employed in the experiments, and from the other 40%, all works but one analyzed samples of 2 to 10 μm thicknesses, following traditional histology procedures. The objective of this work was to assess how the choice of sample thickness affects spectral contrast on MS/HS microscopy data and to suggest an approach for understanding the impact of thickness on the resulting spectral signatures.

A MC Light Transport Simulator framework was developed to investigate tissue-light interactions at various thicknesses (5, 20, 50, 100, 200, 500, 1000, and 2000μm) in both normal and lesioned tissues, including breast, lung, colorectal, and liver. This process involved a thorough analysis of a microscopic HS system, with particular attention to its instrumentation and optical components. Furthermore, the optical properties of the tissues, such as absorption, scattering, scattering phase function, and refractive index, were sourced from the literature to create a virtual model of the tissues for the simulations. After the development of the MC framework, ten values were selected for each tissue, pathology, and thickness, making a total of 640 simulated spectral signatures in the range of 400 to 1000 nm. Once the simulations were completed, data augmentation techniques were applied to the original dataset to better reflect real-world HS microscopic conditions. These techniques simulated practical factors such as sensor SNR, wavelength calibration errors, inadequate thermal management of the light source, and intensity fluctuations due to variations in the light source, thereby enhancing the simulation data with more realistic scenarios.

MC simulated spectra from the virtual specimen illustrated that conventional histology thicknesses (~5 μm) would produce uniform light responses, suggesting that contrast in standard histology images primarily originates from dye absorbance (exogenous chromophores). Conversely, excessively thick samples (1000 and 2000μm) would block light transmission, resulting in limited photon capture by the HS camera. Qualitatively, samples sliced at 500 μm demonstrated enhanced discrimination between normal and lesioned tissue, although they exhibit a relatively low maximum intensity (50% of the maximum achievable). To quantitatively evaluate the discrimination factor of each thickness, at which normal and lesioned tissue become distinguishable, several spectral evaluation metrics were employed (Euclidean Distance, SAM, NS^3, SID, and SID-SAM). For each thickness, these metrics evaluated all spectral signatures between them, categorizing the results into two groups: inter-class (N vs L) and intra-class (N vs N | L vs L) comparisons. The goal was to identify the thickness where a given metric best discerned between spectra from different classes, characterized by small intra-class distances and large inter-class distances. For each thickness, histograms of inter-class and intra-class distances were created, and a range of thresholds were tested to find the one providing the most accurate classification of classes. The accuracy of the classification at the optimal threshold for each thickness indicated the discrimination power of the spectral evaluation metric at that thickness. As expected from the qualitative analysis, results showed that a thickness of 5 μm did not provide adequate differentiation between spectra from the same or different classes. For other thicknesses, each tissue type exhibited a distinct maximum discrimination thickness: 1000 μm for breast tissue, 200 and 500 μm for lung and colorectal tissues, and 500 μm for liver tissue.

However, caution is required when interpreting the absolute values presented in this paper since several limitations must be considered in this study. The simulated tissue closely approximates in vivo conditions (based on the available values in the current literature), although histological analysis is performed on ex vivo samples. The biopsy procedure entails cutting and slicing, which results in blood loss from the sample, leaving an arbitrary residual volume. Furthermore, the remaining hemoglobin interacts with oxygen in the air, forming oxyhemoglobin. As a result, the observed saturation levels may approach 100%, making them irrelevant to the original tissue

Biomedical Optics EXPRESS

saturation. Additionally, the preparation of formalin-fixed stained slices may lead to the loss of water and fat. Other endogenous chromophores may also be missing, or the homogenized structure of the sample may not accurately resemble the original tissue. These factors should be considered in future simulations tailored to specific applications. Moreover, these simulations do not account for the spectral signatures associated with sample fixation methods involving chemicals such as formalin and paraffin. Variations in instrumentation (such as power loss in the optical system and the effects of the histology glass holding the samples) have also not been simulated. It is also important to note that the simulations were performed on a spatially homogeneous sample, without accounting for structural heterogeneity and its effects on the resulting MS/HS image (e.g., optical vignetting, image aberrations, focus variation, etc.). Thus, while the virtual specimen enables the simulation of light transmission through various tissues and thicknesses, the absolute spectral signature values obtained are not expected to perfectly reflect real-life scenarios. Nonetheless, the results of the simulation show that for unstained hyperspectral microscopy, the tissue thickness is a factor that should be considered during sample acquisitions.

6. Conclusions

In conclusion, this study underscores the critical importance of carefully considering sample preparation protocols in MS/HS microscopic applications, as conventional histological methods may not be able to provide the enhanced information derived from light-tissue interactions in thicker samples. The simulated spectral signatures presented here provide valuable insights into how light-tissue interactions vary with tissue thickness, serving as a useful reference point for future studies. Nevertheless, these results reflect a theoretical investigation, as they do not account for sample heterogeneity, preparation artifacts, or instrument-specific effects beyond those simulated. Consequently, further research is needed to determine the optimal sample thickness for each tissue type to ensure more accurate and reliable results in real-world applications. The methodology developed in this work can be adapted and extended to other systems and tissue types, allowing researchers in the field to identify the optimal thickness for their specific applications. Different MS/HS sensors with varying bandwidths and numbers of spectral bands may also be simulated. The discriminative power of each thickness depends on the tissue composition (which will be heterogeneous and more complex than the single-layer model proposed in this paper) and on the quality of the spectral signatures (which will depend on the instrumentation). Nonetheless, a correlation between the simulations and the experimental results is expected when the spectral resolution of the instrumentation used in the experiments aligns with the simulated one. However, this correlation must be validated empirically.

Thus, while simulated spectra offer insights into which tissue thicknesses provides enhanced spectral contrast, future studies should empirically validate these findings. The framework developed in this work can be used to perform a finer search for the optimal thickness of a specific tissue, which can then be validated using HS microscopy data acquired from real tissue sections. Validation would involve classification tasks, such as distinguishing between different tissue types (e.g., liver vs. lung) or pathological states (e.g., normal vs. tumor), to assess whether the selected thickness enhances classification performance. Afterwards, it would be crucial to explore how variations in sample thickness can be effectively achieved and incorporated into the clinical workflow. While the aim is not to drastically alter existing clinical processes, if unstained thicker samples lead to improved spectral data in MS/HS microscopy, their integration into clinical procedures could enhance diagnostic accuracy beyond the current state of the art.

Funding. Consejería de Educación, Universidades, Cultura y Deportes, Gobierno de Canarias (Research Stay, EST2023010017); Agencia Canaria de Investigación, Innovación y Sociedad de la Información (TESIS2021010084); European Health and Digital Executive Agency (STRATUM, 101137416); Ministerio de Ciencia, Innovación y Universidades (OASIS, PID2023-148285OB-C43).

Biomedical Optics EXPRESS

Acknowledgment. We gratefully acknowledge Óscar Quintana for his work in developing Fig. 2 presented in this study. His expertise in data visualization was instrumental in enhancing the clarity and impact of this work. The authors also thank Guillermo V. Socorro-Marrero for his support in the statistical analysis of this work.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," J. Biomed. Opt **19**(1), 010901 (2014).
2. S. Ortega, M. Halicek, H. Fabelo, *et al.*, "Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review," Biomed. Opt. Express **11**(6), 3195 (2020).
3. Q. Li, Y. Xue, J. Zhang, *et al.*, "Microscopic hyperspectral imaging studies of normal and diabetic retina of rats," Sci. China, Ser. C:Life Sci. **51**(9), 789–794 (2008).
4. H. Fabelo, R. Leon, S. Ortega, *et al.*, "Novel Methodology for Alzheimer's Disease Biomarker Identification in Plasma using Hyperspectral Microscopy," in *2020 XXXV Conference on Design of Circuits and Integrated Systems (DCIS)* (2020).
5. L. Ma, J. V. Little, A. Y. Chen, *et al.*, "Automatic detection of head and neck squamous cell carcinoma on histologic slides using hyperspectral microscopic imaging," J. Biomed. Opt **27**(04), 046501 (2022).
6. Y. Duan, J. Wang, M. Hu, *et al.*, "Leukocyte classification based on spatial and spectral features of microscopic hyperspectral images," Opt. Laser Technol. **112**, 530–538 (2019).
7. K. S. Banu, M. Lerma, S. U. Ahmed, *et al.*, "Hyperspectral microscopy- applications of hyperspectral imaging techniques in different fields of science: a review of recent advances," Appl Spectrosc Rev 1–24 (2023).
8. S. Ortega, R. Guerra, M. Diaz, *et al.*, "Hyperspectral Push-Broom Microscope Development and Characterization," IEEE Access **7**, 122473–122491 (2019).
9. J. Stergar, R. Hren, and M. Milanič, "Design and Validation of a Custom-Made Hyperspectral Microscope Imaging System for Biomedical Applications," Sensors **23**(5), 2374 (2023).
10. *An Atlas of Comparative Vertebrate Histology* (Elsevier, 2018).
11. L. Quintana, S. Ortega, R. Leon, *et al.*, "Instrumentation Evaluation for Hyperspectral Microscopy Targeting Enhanced Medical Histology," in *2021 XXXVI Conference on Design of Circuits and Integrated Systems (DCIS)* (IEEE, 2021), pp. 1–6.
12. Q. Zhang, Y. Wang, Q. Li, *et al.*, "An autofocus algorithm considering wavelength changes for large scale microscopic hyperspectral pathological imaging system," J Biophotonics **15**(5), e202100366 (2022).
13. L. Quintana, S. Ortega, H. Fabelo, *et al.*, "Blur-Specific No-Reference Image Quality Assesment for Microscopic Hyperspectral Image Focus Quantification," in *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (IEEE, 2021), pp. 1–5.
14. D. Cozzolino, P. J. Williams, and L. C. Hoffman, "An overview of pre-processing methods available for hyperspectral imaging applications," Microchem. J. **193**, 109129 (2023).
15. W. M. Star, "Diffusion Theory of Light Transport," in *Optical-Thermal Response of Laser-Irradiated Tissue* (Springer US, 1995), pp. 131–206.
16. C. R. Nave and J. Sheridan, "The microwave and infrared spectra and structure of hydrothiophosphoryl difluoride," J. Mol. Struct. **15**(3), 391–398 (1973).
17. A. T. Feldman and D. Wolfe, "Tissue Processing and Hematoxylin and Eosin Staining," in (2014), pp. 31–43.
18. H. A. Alturkistani, F. M. Tashkandi, and Z. M. Mohammedsaleh, "Histological Stains: A Literature Review and Case Study," Glob J Health Sci **8**(3), 72 (2015).
19. S. L. Jacques and L. Wang, "Monte Carlo Modeling of Light Transport in Tissues," in *Optical-Thermal Response of Laser-Irradiated Tissue* (Springer US, 1995), pp. 73–100.
20. Q. Fang and S. Yan, "MCX Cloud—a modern, scalable, high-performance and in-browser Monte Carlo simulation platform with cloud computing," J. Biomed. Opt **27**(08), 083008 (2022).
21. R. Lott, E. Janet Tunnicliffe, J. S. C. Sheppard, *et al.*, *Practical Guide to Specimen Handling in Surgical Pathology* (College of American Pathologists, 2015).
22. Q. Li, J. Zhang, Y. Wang, *et al.*, "Molecular Spectral Imaging System for Quantitative Immunohistochemical Analysis of Early Diabetic Retinopathy," Appl. Spectrosc. **63**(12), 1336–1342 (2009).
23. Q. Li, Y. Xue, G. Xiao, *et al.*, "New microscopic pushbroom hyperspectral imaging system for application in diabetic retinopathy research," J. Biomed. Opt **12**(6), 064011 (2007).
24. Qingli Li, Yiting Wang, Jingfa Zhang, *et al.*, "Quantitative Analysis of Protective Effect of Erythropoietin on Diabetic Retinal Cells Using Molecular Hyperspectral Imaging Technology," IEEE Trans. Biomed. Eng. **57**(7), 1699–1706 (2010).
25. E. Chlipala, C. M. Bendzinski, K. Chu, *et al.*, "Optical density-based image analysis method for the evaluation of hematoxylin and eosin staining precision," J. Histotechnol. **43**(1), 29–37 (2020).
26. D. Pertzborn, H.-N. Nguyen, K. Hüttmann, *et al.*, "Intraoperative Assessment of Tumor Margins in Tissue Sections with Hyperspectral Imaging and Machine Learning," Cancers **15**(1), 213 (2022).

Biomedical Optics EXPRESS

27. S. L. Jacques, "Optical properties of biological tissues: a review," Phys. Med. Biol. **58**(11), R37–R61 (2013).
28. L. L. de Boer, B. H. W. Hendriks, F. van Duijnhoven, *et al.*, "Using DRS during breast conserving surgery: identifying robust optical parameters and influence of inter-patient variation," Biomed. Opt. Express **7**(12), 5188 (2016).
29. J. W. Spliethoff, L. L. de Boer, M. A. J. Meier, *et al.*, "Spectral sensing for tissue diagnosis during lung biopsy procedures: The importance of an adequate internal reference and real-time feedback," Lung Cancer **98**, 62–68 (2016).
30. M. S. Nogueira, M. Raju, J. Gunther, *et al.*, "Tissue biomolecular and microstructure profiles in optical colorectal cancer delineation," J. Phys. D: Appl. Phys. **54**(45), 454002 (2021).
31. N. Reistad, J. H. Nilsson, M. Bergenfeldt, *et al.*, "Intraoperative liver steatosis characterization using diffuse reflectance spectroscopy," HPB **21**(2), 175–180 (2019).
32. N. Ghosh, S. K. Mohanty, S. K. Majumder, *et al.*, "Measurement of optical transport properties of normal and malignant human breast tissue," Appl. Opt. **40**(1), 176 (2001).
33. R. Marchesini, A. Bertoni, S. Andreola, *et al.*, "Extinction and absorption coefficients and scattering phase functions of human tissues in vitro," Appl. Opt. **28**(12), 2318 (1989).
34. A. N. Bashkatov, E. A. Genina, V. I. Kochubey, *et al.*, "Optical properties of human colon tissues in the 350–2500 nm spectral range," Quantum Electron. **44**(8), 779–784 (2014).
35. P. Saccomandi, V. Vogel, B. Bazrafshan, *et al.*, "Estimation of anisotropy coefficient of swine pancreas, liver and muscle at 1064 nm based on goniometric technique," J Biophotonics **8**(5), 422–428 (2015).
36. P. Geladi, J. Burger, and T. Lestander, "Hyperspectral imaging: calibration problems and solutions," Chemom. Intell. Lab. Syst. **72**(2), 209–217 (2004).
37. C.-I. Chang, *Hyperspectral Imaging* (Springer US, 2003).
38. S. Yan, S. L. Jacques, J. C. Ramella-Roman, *et al.*, "Graphics-processing-unit-accelerated Monte Carlo simulation of polarized light in complex three-dimensional media," J. Biomed. Opt **27**(08), 083015 (2022).
39. Q. Fang and D. A. Boas, "Monte Carlo Simulation of Photon Migration in 3D Turbid Media Accelerated by Graphics Processing Units," Opt. Express **17**(22), 20178 (2009).
40. S. Yan and Q. Fang, "Hybrid mesh and voxel based Monte Carlo algorithm for accurate and efficient photon transport modeling in complex bio-tissues," Biomed. Opt. Express **11**(11), 6262 (2020).
41. A. Scarbrough, K. Chen, and B. Yu, "Designing a use-error robust machine learning model for quantitative analysis of diffuse reflectance spectra," J. Biomed. Opt **29**(01), 015001 (2024).
42. F. A. Kruse, A. B. Lefkoff, J. W. Boardman, *et al.*, "The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data," Remote Sens Environ **44**(2-3), 145–163 (1993).
43. Chein-I Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," IEEE Trans. Inf. Theory **46**(5), 1927–1932 (2000).
44. C.-I. Chang, "New hyperspectral discrimination measure for spectral characterization," Opt. Eng. **43**(8), 1777 (2004).
45. A. E. Profio, "Light transport in tissue," Appl. Opt. **28**(12), 2216 (1989).

Chapter 5. HS microscopic Medical Databases

A critical step toward the clinical translation of HS microscopy lies in the establishment of benchmark datasets that accurately capture the variability inherent in biological sample preparation and image acquisition conditions. These datasets are essential for advancing HS imaging research because they provide standardized yet diverse examples of tissue morphology and spectral signatures, which are indispensable for developing, training, and validating robust image analysis algorithms. Despite the growing interest and rapid technological progress in HS microscopy, a major challenge remains: the scarcity of publicly available, well-annotated HS microscopic datasets. This gap significantly impedes the ability of researchers to objectively evaluate and compare the performance of advanced processing methods, including HS image classification, spectral unmixing, and data compression. Without public, standardized datasets, benchmarking becomes inconsistent, hindering reproducibility and complicating collective progress in the field.

This chapter addresses objective **O5** of this Ph.D. thesis, which is dedicated to the acquisition of HS images of histological samples across varying preparation and imaging conditions. The goal is to create representative datasets that reflect real-world laboratory practices and enable reproducible, comparable research. To this end, two comprehensive databases were developed, each tailored to support different clinical and research needs within the field of HS microscopy.

The first database (**J4**), **"Histological Hyperspectral Glioblastoma Dataset (HistologyHSI-GB)"**, was created in-house. The data have been published in *The Cancer Imaging*

Archive (TCIA) repository, and the data descriptor has been published in the journal *Scientific Data*. This dataset contains 469 hyperspectral images acquired from 13 patients diagnosed with glioblastoma, a highly aggressive brain tumor. Tissue sections stained with H&E were captured at 20× magnification using a custom HS microscope system with spectral sensitivity spanning 400 to 1000 nm. Expert histopathologists provided image-level annotations, ensuring clinical relevance and enabling robust algorithm training. By capturing both spectral and spatial tissue features, this dataset offers a rich resource for studying tumor heterogeneity and developing HS image analysis methods with application in neuro-oncology.

The second dataset was established through a collaborative effort with the Department of Pathology at the Hospital of Tortosa Verge de la Cinta in Tortosa, Spain. During her research stay (**M1**), the Ph.D. Candidate contributed to the acquisition and curation of this breast tissue dataset, focused on recurrence cases. This database (**J5**), **"Histological Hyperspectral Breast Cancer Recurrence Dataset (HistologyHSI-BC Recurrence)"**, also published the data in TCIA and the data descriptor in the journal *Scientific Data*. It offers valuable imaging data for studying diagnostic challenges associated with recurrent tumors. It was specifically designed to advance prediction of distant breast cancer recurrence. It integrates 677 HS images, 47 histopathological whole-slide images, and detailed clinical and demographic data from 47 breast cancer patients, nearly half of whom developed recurrence over a 12-year period. The histological slides were digitized via whole-slide scanning and expertly annotated, while HS images were acquired using a bright-field microscope coupled with an HS camera. This multimodal approach offers unprecedented opportunities to identify prognostic biomarkers by combining spectral imaging, histopathology, and clinical insights, thus supporting personalized treatment strategies.

The contents of this chapter offer significant contributions to the scientific community by releasing two publicly available HS microscopic datasets, which will lead to advancements in both research and clinical use of HS microscopic imaging. The *HistologyHSI-GB* and *HistologyHSI-BC Recurrence* databases

provide valuable, clinically relevant resources that support algorithm development and clinical translation. By addressing data variability and promoting standardization, this work helps accelerate reproducibility, collaboration, and the adoption of HS microscopy in routine medical practice.

J4. Histological Hyperspectral Glioblastoma Dataset (HistologyHSI-GB)

Title:	Histological Hyperspectral Glioblastoma Dataset (HistologyHSI-GB)		
Authors:	Samuel Ortega[¥], **Laura Quintana-Quintana**[¥], Raquel Leon[¥], Himar Fabelo, Maria de la Luz Plaza, Rafael Camacho and Gustavo M. Callico.		
Journal:	Scientific Data		
Publisher:	Nature Portfolio		
Date: 2024	**Vol.**: 11	**Pages**: 681	
doi:	https://doi.org/10.1038/s41597-024-03510-x		
JIF Quartile:	Q1 in Multidisciplinary Sciences	**IF**:	5.8
Open Access, Open Code, Open Data, National Collaboration			

[¥]Equal Contribution

scientific data

OPEN

DATA DESCRIPTOR

Histological Hyperspectral Glioblastoma Dataset (HistologyHSI-GB)

Samuel Ortega[1,2,3,7] ✉, Laura Quintana-Quintana[2,7], Raquel Leon[2,7], Himar Fabelo[2,4,5], María de la Luz Plaza[6], Rafael Camacho[6] & Gustavo M. Callico[2]

Hyperspectral (HS) imaging (HSI) technology combines the main features of two existing technologies: imaging and spectroscopy. This allows to analyse simultaneously the morphological and chemical attributes of the objects captured by a HS camera. In recent years, the use of HSI provides valuable insights into the interaction between light and biological tissues, and makes it possible to detect patterns, cells, or biomarkers, thus, being able to identify diseases. This work presents the HistologyHSI-GB dataset, which contains 469 HS images from 13 patients diagnosed with brain tumours, specifically glioblastoma. The slides were stained with haematoxylin and eosin (H&E) and captured using a microscope at 20× power magnification. Skilled histopathologists diagnosed the slides and provided image-level annotations. The dataset was acquired using custom HSI instrumentation, consisting of a microscope equipped with an HS camera covering the spectral range from 400 to 1000 nm.

Background & Summary

Hyperspectral (HS) imaging (HSI) is a technology able to measure both the spatial and spectral information of objects or substances, combining the features of spectroscopy and digital imaging in a single imaging modality. Because the absorption, reflection, transmission and scattering of light are unique to each material, this technology allows non-invasive identification of materials. The first use of HSI was for the remote sensing exploration of the Earth's surface in the 80s[1]. In recent years, this technology has been extended to a wide range of applications, such as precision agriculture[2,3], food quality inspection[4–6], industrial sorting of materials[7,8], art conservation[9,10], or forensic sciences[11,12]. In medicine, recent research has proven HSI technology to be useful for different clinical applications[13,14], for example, as a surgical guidance tool[15,16], as a tool for early diagnosis[17–19], or as a technology able to measure different biochemical parameters that can be useful for medical practitioners[20–23].

Digital and computational pathology techniques are intended to provide pathologists with a tool for the quantitative analysis of pathological specimens, reducing inter-observer variability among different pathologists and saving the time of manual examination of histological specimens[24,25]. Recently, some researchers have investigated HSI as a suitable technology for computational pathology in various fields, such as digital staining, colour enhancement, standardization of pathological slides or the exploitation of autofluorescence or immunohistochemistry of histological slides[26]. However, the primary use of HSI in computational pathology is currently in diagnostic research for routine clinical practice. In this context, recent applications have been focused on the diagnosis of cholangiocarcinoma[27,28], head and neck squamous cell carcinoma[29], membranous nephropathy[30], breast cancer[31], or the classification of leukocyes[32,33], among others.

The workflow in HS computational pathology research usually involves digitizing the histological slides using HSI instrumentation and extracting information from the HS images that could be useful for diagnostic

[1]Seafood Industry Department, Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima), Tromsø, Norway. [2]Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain. [3]Department of Mathematics and Statistics, UiT The Arctic University of Norway, Tromsø, Norway. [4]Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Las Palmas de Gran Canaria, Spain. [5]Research Unit, Hospital Universitario de Gran Canaria Doctor Negrín, Las Palmas de Gran Canaria, Spain. [6]Department of Pathological Anatomy, Hospital Universitario de Gran Canaria Doctor Negrín, Las Palmas de Gran Canaria, Spain. [7]These authors contributed equally: Samuel Ortega, Laura Quintana-Quintana, Raquel Leon. ✉e-mail: sortega@iuma.ulpgc.es

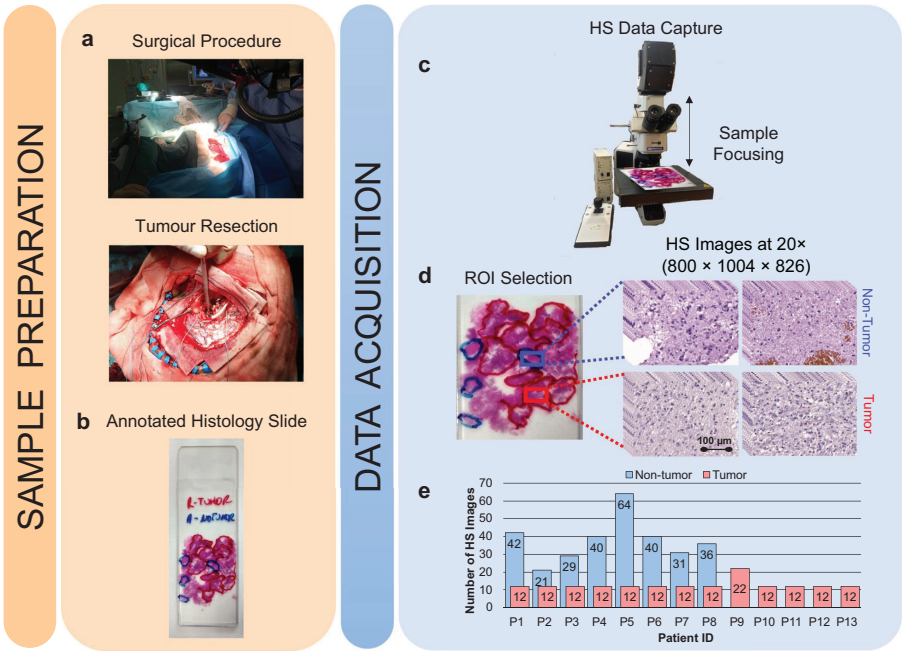


Fig. 1 Graphical abstract of the methodology followed. (**a**) Resection procedure. (**b**) Macroscopic annotations of the GB locations. (**c**) HS data capture using a microscopic HS system. (**d**) ROI selection. (**e**) Dataset summary.

purposes using various image processing methods. Although a wide variety of techniques are used in the literature to this end, it is difficult to compare the different approaches fairly, mainly due to the lack of publicly available datasets[34].

In this work, we provide a publicly available dataset of HS images of haematoxylin and eosin (H&E) stained histological slides corresponding to brain tumours, specifically Glioblastoma (GB)[35,36]. To the best of our knowledge there are other databases related to gastric cancer but, this is the first publicly available dataset of HS brain histological images[35]. This dataset is composed of 469 HS images from 13 different patients, with image-level annotations for two different classes (non-tumour or tumour) according to the manual examination of the histological samples. The HS images cover the spectral range from 400 to 1000 nm and were taken at 20× magnification. On the one hand, this dataset can be relevant for researchers interested in HS image classification and other HS image processing techniques, such as spectral unmixing or HS data compression. This dataset was acquired by our research group and all HS images were employed to train different classification algorithms for GB detection which were presented in previous research work[36–38]. In this manuscript, we exclusively present the curated version of the dataset, from which artifacts and labelling errors found in previous publications have been eliminated. Besides the proposed classification techniques, a broad range of potential methods could be explored to evaluate the effectiveness of HSI in enhancing the performance relative to the outcomes obtained with RGB (Red-Green-Blue) images. On the other hand, this dataset can be used by researchers in the field of computational pathology and pathology practitioners to envision the possibilities of this technology for routine clinical practice. In this work, we provide a repository with the HS data, its homologous RGB image and, a snapshot of the original slides showing the region of interest for each HS image. We also provide a comprehensive explanation of the microscopic HS system, its quality validation process, and how the dataset is organized.

Methods

This section provides a detailed explanation of the methodology employed in previous works[36–38], This includes a description of the methods used for collecting histological samples, an overview of the microscopic HS system, and the process of acquiring and processing the HS data.

Histological samples description. The research conducted in this study employs human biopsies extracted during brain tumour resection procedures (Fig. 1a). This research involved participants who were 18 years of age or older, all diagnosed with primary brain tumours and undergoing neurosurgical procedures at the University Hospital of Gran Canaria Doctor Negrín (Las Palmas de Gran Canaria, Spain). Prior to their involvement in the study, each participant provided written informed consent, which explicitly authorized the publication of any images or data obtained during the study. The Research Ethics Committee of the University Hospital of Gran Canaria Doctor Negrin (Comité Ético de Investigación Clínica-Comité de Ética en la Investigación,

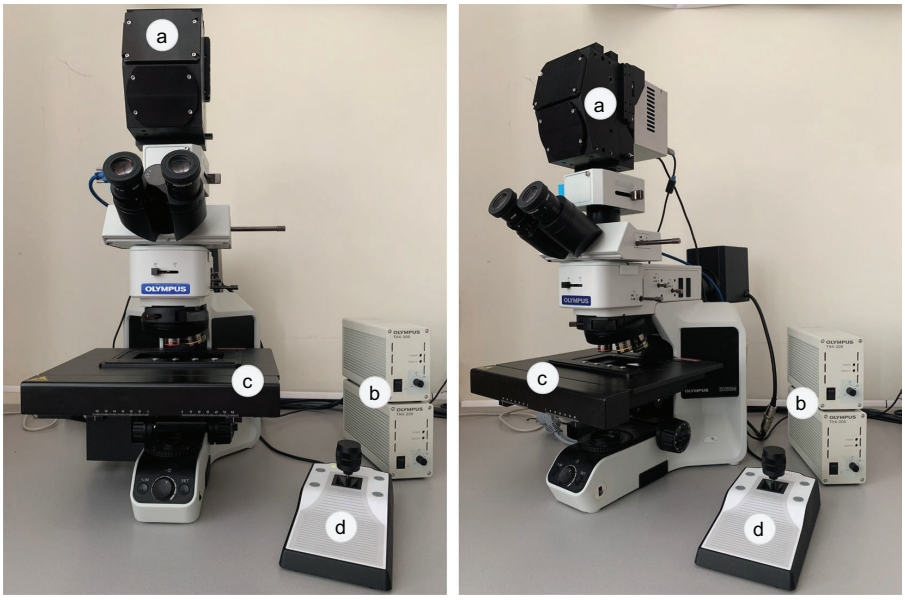


Fig. 2 Microscopic HS system. (**a**) HS camera. (**b**) Halogen light source. (**c**) Positioning joystick. (**d**) XY linear stage.

CEIC/CEI) approved the study protocol and consent procedures (reference 130069). All research procedures were conducted in strict compliance with applicable guidelines and regulations. The pathological slides used in this research were processed and analysed in the Pathological Anatomy Department of the same hospital. After the tumour tissue resection during neurosurgery, the biopsy samples underwent a series of standardized procedures. First, the samples were dehydrated to remove the excess of water, as it is immiscible with most embedding media. The samples were then embedded in paraffin blocks, mounted on microtomes and sliced into 4 μm thick slices. Finally, the slices were rehydrated and stained with H&E, a method commonly used in pathology.

The pathologists involved in the study analysed the stained sections using routine examination techniques. Each sample was evaluated and diagnosed as GB (a grade 4 primary brain tumour) according to the 2016 World Health Organization (WHO) classification of tumours of the central nervous system[39]. Macroscopic annotations of the GB locations on the physical pathological slides were made using a red marker pen (Fig. 1b). These annotations served as reference points for further analysis. In addition, non-tumour areas, where no discrete presence of tumour cells was observed, were annotated (blue marker pen) on the histology slides. The pen-marker annotations on the histological slide were deliberately outlined with wide borders to maintain a safety distance between tumour and non-tumour areas. Afterwards, regions of interest (ROIs) were selected from these pathologist-annotated areas for further study. These ROIs were subsequently digitized using the microscopic HS system (Fig. 1c), allowing for a detailed analysis of their spectral characteristics. Multiple HS images were acquired to cover the entire selected ROI. Figure 1d shows an example of the annotations within the pathological slide and the selection of different ROIs and the HS images (imaged at 20×). Finally, Fig. 1e summarizes the number of HS images acquired for each patient in the HistologyHSI-GB dataset.

Microscopic HS system. In this study, an HS camera coupled to a conventional brightfield microscope was employed to capture the HistologyHSI-GB dataset (Fig. 2). The HS camera (Fig. 2a) is a Hyperspec® VNIR A-Series from HeadWall Photonics (Fitchburg, MA, USA), which is based on an imaging spectrometer coupled to a CCD (charge-coupled device) sensor, the Adimec-1000 m (Adimec, Eindhoven, Netherlands). This HS camera works in the visual and near-infrared (VNIR) spectral range, from 400 to 1000 nm with a spectral resolution of 2.8 nm, sampling 1004 spatial pixels and, 826 spectral channels. The microscope is an Olympus BX-53 (Olympus, Tokyo, Japan), with four magnification lenses: 5×, 10×, 20× and 50×. The objective lenses are optimized for infrared (IR) observations and the light source is an halogen lamp (Fig. 2b). The HS camera is based on a push-broom technique, requiring a spatial scanning to acquire an HS cube. The system employs a mechanical stage (SCAN, Märzhäuser, Wetzlar, Germany) attached to the microscope for this purpose, which provides accurate movement in the 3 spatial axes directions (Fig. 2c-d). A more detailed description of the different parts of the acquisition system can be found in Table 1. A custom software was developed for synchronizing the scanning movement and the HS camera data acquisition. The optimal exposure time was configured to 40 *ms* (the maximum allowed by the HS camera). The scanning speed of the microscope platform was adjusted according to the ratio of pixel size to exposure time to obtain squared pixels in the resulting HS cubes. Since multiple images were captured from each ROI, the software was designed to enable the acquisition of consecutive HS cubes in a row.

	Component	Manufacturer	Model	Key Parameter	
Microscope	Microscope Model	Olympus, Tokyo, Japan	BX-53	Brightfield microscope	
	Trinocular Tube		U-TR30IR-1-2	30° inclination and FN22	
	Eyepiece		WHN10X-H-1-7	Wide field eyepiece FN22	
	Lenses*		LMPLN-IR	5× and 10×	
			LCPLN-IR	20× and 50×	
	Power supply unit		TH4 - 200	12 V 100 W	
	Lamp House		U-LH100IR-1-7	Transmittance and reflectance	
	IR Halogen Lamp	Philips, Amsterdam, Netherlands	7724 EVA	400–1800 nm	
	Stage	Märzhäuser, Wetzlar, Germany	SCAN 130 × 85	3D movement with ± 3 μm resolution	
	Joystick		M-HID-JS-3	Movement in the 3 axes	
	Camera Adapter	Olympus, Tokyo, Japan	U-CMAD3-1-7	C-mount	
HSI System	HS Camera	HeadWall Photonics, Fitchburg, MA, USA	Hyperspec® VNIR A-Series	Technology	Push-broom scanning
				Spectral range	400 to 1000 nm
				N° of bands	826 bands
				Spectral resolution	2.8 nm
				Spatial size	1 × 1004 pixels
	FPA Detector	Adimec, Eindhoven, Netherlands	RA1000m	CCD with 7.4 μm pixel pitch	
	Frame grabber	EPIX, Inc., Buffalo Grove, IL, USA	PIXCI® EL1	PCIe x1 Camera Link Frame Grabber	

Table 1. Description of the HS microscopic system components.

Whenever an HS cube was captured (composed by 800 lines), it was stored in memory while the camera and platform continuing to capture data until several cubes were captured. This approach helps save time during image acquisition and minimizes the need for human intervention. To prevent potential degradation of focus or errors caused by the platform while moving, the capture of consecutive HS cubes was limited to a maximum of ten.

Data acquisition methodology. As previously mentioned, relevant areas were identified on the slides and highlighted with a pen in blue (non-tumour) or red (tumour). The capture process of a sample starts by selecting a ROI, from different non-tumour and tumour highlighted areas, to be imaged. Since cells details are needed for further processing, a 20× magnification was chosen to capture the HS images. The coarse focus of the specimen (Fig. 3a) is performed using the microscope binoculars. The procedure relies on the user's subjective criteria. The final HS image is brought into focus by examining a specific frame captured by the push-broom camera, referred to as the $Y\lambda$ frame (e.g., $Y\lambda$ frame extracted from the yellow line in Fig. 3b). The λ axis of an $Y\lambda$ frame corresponds to the spectral information, while the Y axis represents the spatial information across the field of view (FOV) of the camera. The objective is to identify the sharpest spatial frequency along various working distances from the sensor to the sample (Fig. 3c shows a focused $Y\lambda$ frame while Fig. 3d shows an unfocused one). The working distance adjustment is performed precisely by using the Z movement with the joystick.

After achieving the optimal focus on the sample, the software was configured to capture several HS images consecutively, where the number of images is defined as an input parameter. The number of images should be kept relatively low to avoid the focus degradation throughout the specimen, due to the non-flat nature of microscopic samples and the platform error/vibration during movement. In this case, a maximum of 10 HS images were extracted consecutively from a selected ROI. The dataset was captured with the light power set to the maximum (100 W) and the exposure time to 40 *ms*. At 20× magnification, the pixel size is 0.373 μm, and the microscope platform was configured to scan the sample at a speed of 9.325 $\mu m/s$. Furthermore, to overcome the challenges posed by the high dimensionality of the HS images, the collected cubes were constrained to a spatial size of 800 lines, resulting in 1.23 GB data cubes. The HS cubes had a dimension of 800 × 1004 × 826 (number of lines × number of rows × number of bands), corresponding to a spatial size of 299 × 375 μm recorded over a span of 32 *s*.

After the HS images were captured, the reference images for calibration were acquired. In HS image processing, flat-field calibration is an essential step designed to correct the raw data recorded by an HS system. This method corrects the HS data for differences due to the environmental conditions and instrumentation. The flat-field calibration makes use of white (*WR*) and dark (*DR*) reference images. The WR recording is designed to capture data about the HS imaging system under the same conditions used for sample collection, without involving the sample itself. Therefore, the WR is obtained by scanning a section of the histological slide where no tissue is present. Since there is no sample material in such position of the slide, this HS frame contains the maximum values that the sensor is able to measure for each pixel and wavelength in the specified capturing conditions (exposure time, light intensity, the optical properties of the glass slide, etc.). Afterwards, the *DR* is captured by blocking the light transmission to the HS camera. This HS frame contains the minimum values that the system is able to provide for each pixel and band, and also information about the dark currents in the CCD. Ideally, the *DR* values should be very close to zero. However, higher values can be obtained, typically due to the intrinsic noise of the sensor. To ensure a robust measurement of the reference images for calibration, 100 $Y\lambda$ frames are captured for both *WR* and *DR*, allowing any potential errors to be averaged. These reference images were employed for the HS data calibration as detailed in next section. During the acquisition process of the HistologyHSI-GB dataset,

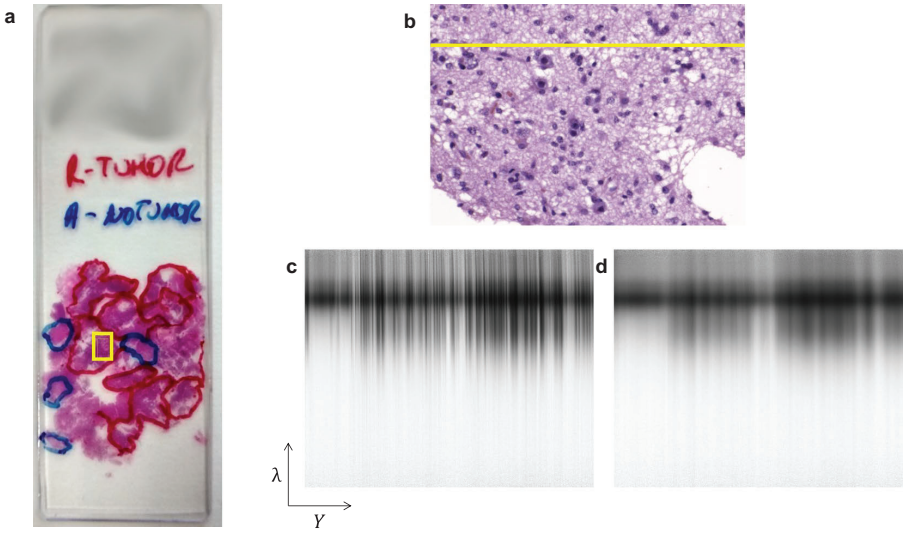


Fig. 3 Capture process to obtain focused HS cubes. (**a**) Example of a histology slide with tumour and non-tumour annotations. The yellow square identifies a ROI where the HS image was captured. (**b**) Synthetic RGB image where the $Y\lambda$ frame employed to focus the sample is marked in yellow. Examples of (**c**) focused and (**d**) unfocused $Y\lambda$ frames.

image-level annotations were applied. These annotations (tumour or non-tumour) remained consistent across the entire HS cube, indicating that all data within the cube shared the same annotation.

HS Data calibration. The goal of the HS microscopic system is to provide a spectral signature per spatial pixel of the captured scene. These spectral signatures indicate the percentage of incident radiation that the scanned object transmits or reflects at each captured wavelength. Various factors, including the inherent spectral response of the sensor, the transmission of light through lenses and optical components, and the spectral characteristics of the light source influence the spectral response of an HS acquisition system. To obtain spectral signatures that accurately indicate the percentage of transmitted or reflected radiation at each wavelength in the sample, the HS cubes need to be calibrated. This calibration consists of normalizing the captured HS pixels by linearly scaling their values considering the *WR* and *DR*. Equation (1) is employed to calibrate the HS data, where r_i and Raw_i refer to each $Y\lambda$ frame from the calibrated and the raw image, respectively. Figure 4 shows an example of how the spectral signatures of different pixels (Fig. 4a) are scaled to transmittance using the aforementioned calibration. The shape of the *WR* and *DR* is shown in Fig. 4b, and several pixels from a ROI before (Fig. 4c), and after calibration (Fig. 4d).

$$r_i = \frac{Raw_i - DR}{WR - DR} \tag{1}$$

Furthermore, the calibration process also helps to remove the stripping noise effect, which typically appears when acquiring HS images using push-broom scanners[40]. The stripping noise consists in spatially coherent lines that appear in the spatial scanning axis due to static artifacts produced in the sensors, which are repeated in each push-broom frame, as shown in Fig. 5a. In the calibrated images, the effect of the stripping noise disappears (Fig. 5b). The stripping noise is mainly produced due to the fact that different photo-receptors of the sensor have slightly different sensibilities, producing slightly different values when measuring exactly the same amount of incident radiation. The effect of stripping noise and light influence can also be observed in the synthetic RGB shown in Fig. 5c and how this effect disappears after performing the calibration (Fig. 5d).

Greyscale images (Fig. 5a,b) were generated by averaging all spectral bands of the HS image, while the synthetic RGB image (Fig. 5c) was obtained closely mimicking the spectral response of the human eye[41]. For modelling the human eye spectral response, the method employed the normal probability density function following Eq. (2) over the HS data, where μ is the mean ($\mu_R = 590$, $\mu_G = 560$, and $\mu_B = 470$) and σ is the standard deviation ($\sigma_R = 0.08$, $\sigma_G = 0.06$, and $\sigma_B = 0.04$). In Fig. 6, we can observe that, after the normal probability density function, RGB channels take the following central values and bandwidths: $R = 590 \pm 44\ nm$, $G = 560 \pm 79\ nm$ and, $B = 470 \pm 111\ nm$.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{2}$$

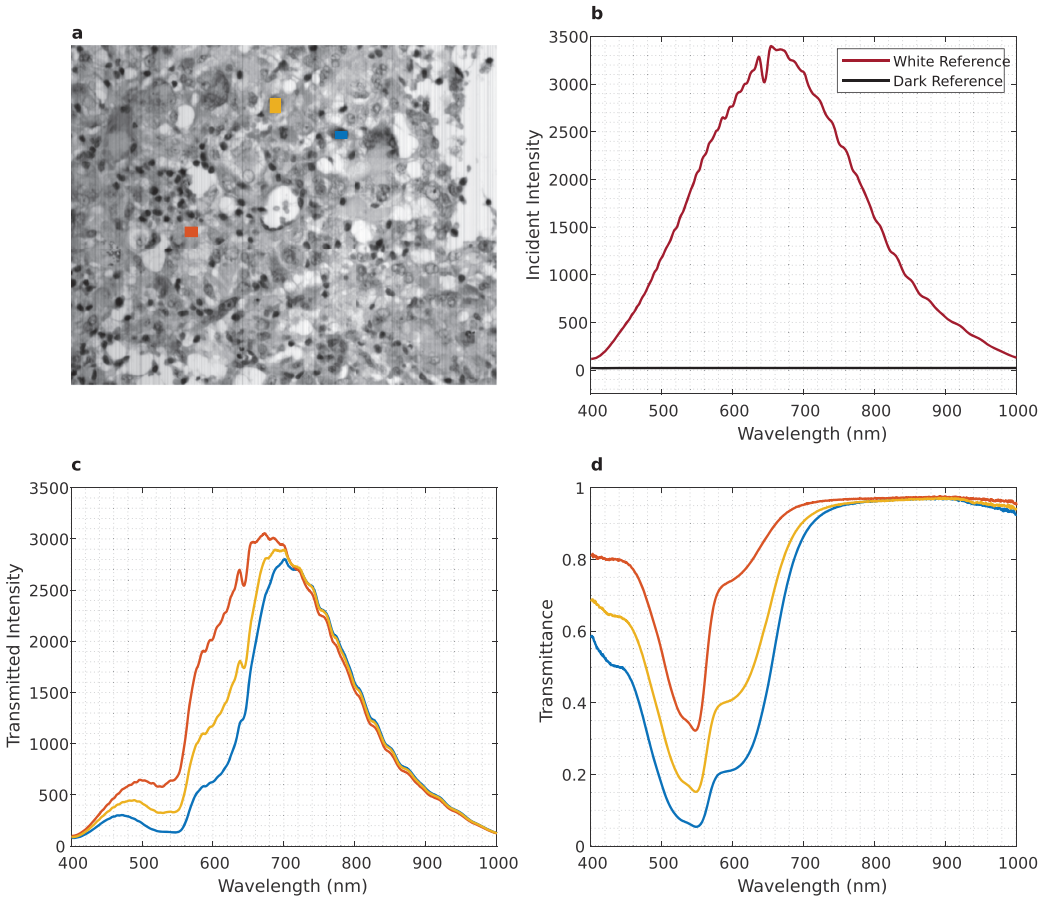


Fig. 4 Effect of calibration in the spectral signatures. (**a**) Grayscale image (generated by averaging each spectral band) and selecting pixels corresponding to different materials. (**b**) *WR* and *DR* spectral signatures. (**c**) Uncalibrated spectral signatures from the selected pixels. (**d**) Calibrated spectral signatures from the selected pixels. Colours in (**c,d**) correspond to selected pixels in (**a**).

Finally, in order to present some examples of the HistologyHSI-GB dataset, Fig. 7 shows the synthetic RGB images, as well as the different calibrated spectral bands found in several HS cubes. The contribution of the sensor noise can be observed in the extreme bands.

Data Records

The HistologyHSI-GB dataset[42] has been deposited in The Cancer Imaging Archive (TCIA) repository[43] for cancer imaging. The dataset is structured in a hierarchy of folders, as shown in Fig. 8. At the top level of the hierarchy there is a single folder associated with each one of the patients comprising the dataset. At the patient level, the folder names correspond to *Pi*, where $\{i \in \mathbb{N} | 1 \leq i \leq 13\}$. For each patient, we can find several folders containing the different HS images for that patient. There is a different number of folders per patient, and the name of each folder encodes the information about which ROI of the histological slide the data was acquired from (*ROI_j*) and another field indicating an image identifier within that ROI (*Ck*). The folders in the image level also contain information about the image-level annotations according to the diagnosis, which can be tumour (*T*) or non-tumour (*NT*). The number of ROIs and image identifiers varies depending on the patient, but the total number of images from each class can be found in Fig. 1e. A conventional image of the slide with the macroscopic annotations and the location of the different ROIs within the slide is available for each patient (*Pi.png*).

Finally, each folder within the image-folder level contains an HS image from the histological slide, the necessary files for the calibration (dark and white references), and a synthetic RGB image extracted from the HS cube. The HS cubes are stored in ENVI format[44] (the standard format for storing HS images). The ENVI format consists of a flat-binary raster file with an accompanying ASCII (American Standard Code for Information Interchange) header file. A more detailed description of the different files in each image folder can be found

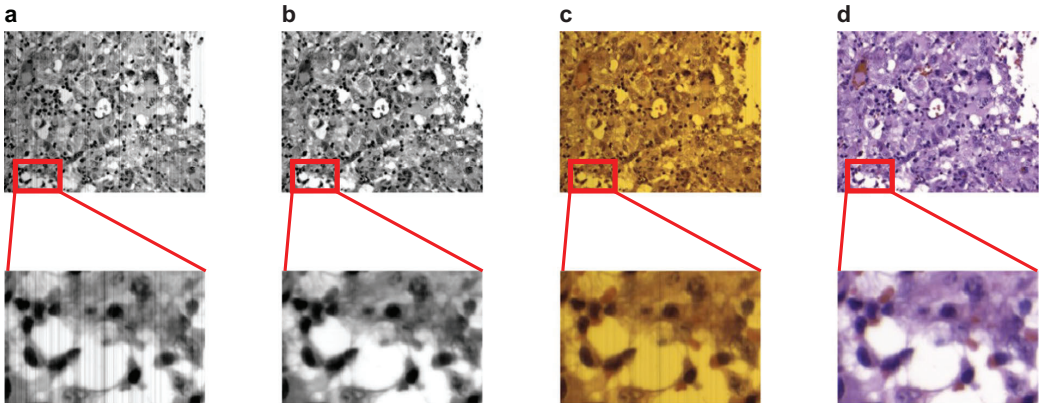


Fig. 5 Examples of the uncalibrated and calibrated HS images. (**a,b**) grayscale representation generated by averaging all spectral bands of the uncalibrated and calibrated HS images, respectively. (**c,d**) synthetic RGB image of the uncalibrated and calibrated HS images, respectively, generated using a model of human eye spectral response.

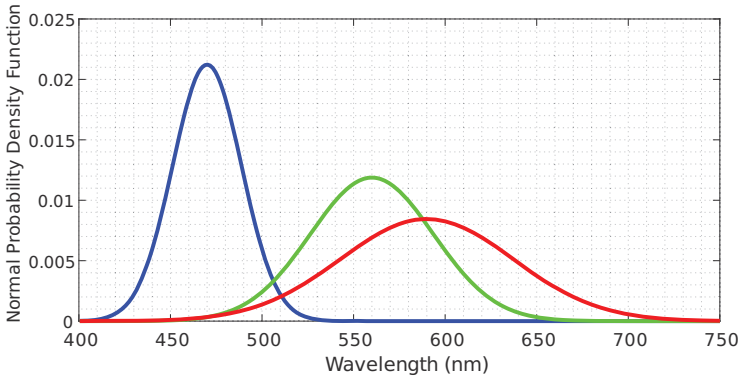


Fig. 6 Human eye spectral response to light where different colour line represents the normal probability distribution function modelling each channel.

Table 2. The HS cubes from the histological slides and the white and dark references are stored as ENVI files. The HistologyHSI-GB dataset comprises 469 images from 13 different patients, where 166 images are labelled as tumour, and 303 are labelled as non-tumour.

Technical Validation

A technical validation was accomplished to support the quality of the HistologyHSI-GB dataset. Linear sensor systems demonstrate analogous basis functions for both spectral sensitivity and responsivity decomposition[45]. Spectral responsivity refers to the effectiveness of light detection in relation to its frequency or wavelength. However, camera channels often exhibit varying sensitivity across different wavelengths due to the spectral responsivities of the detectors and the non-uniform output of diffractive or filtering elements[46]. Proper characterization is essential for ensuring the reliability and accuracy of HS data analysis and interpretation. HS data captured for noise quantification and spectral and spatial calibration which are used to perform the technical validation (Fig. 9) can be found in a published dataset[47].

Signal to noise ratio. In this section, we present the signal-to-noise ratio (SNR) measurements for our instrumentation. We obtained the signal (S) values by capturing images of the light without any sample, similar to the procedure used for recording the WR in flat-field calibration. For the noise (N) values, we recorded HS images in the absence of light. The SNR was calculated as the ratio between the mean value of S and the standard deviation of N. These recordings were taken over 100 push-broom frames under the same conditions as the image recordings. We calculated the SNR over the entire spectral range for the central pixel of the push-broom frame (Fig. 10a), which shows that the SNR exceeds 20 dB over the entire spectral range, peaking 42 dB at 655 nm.

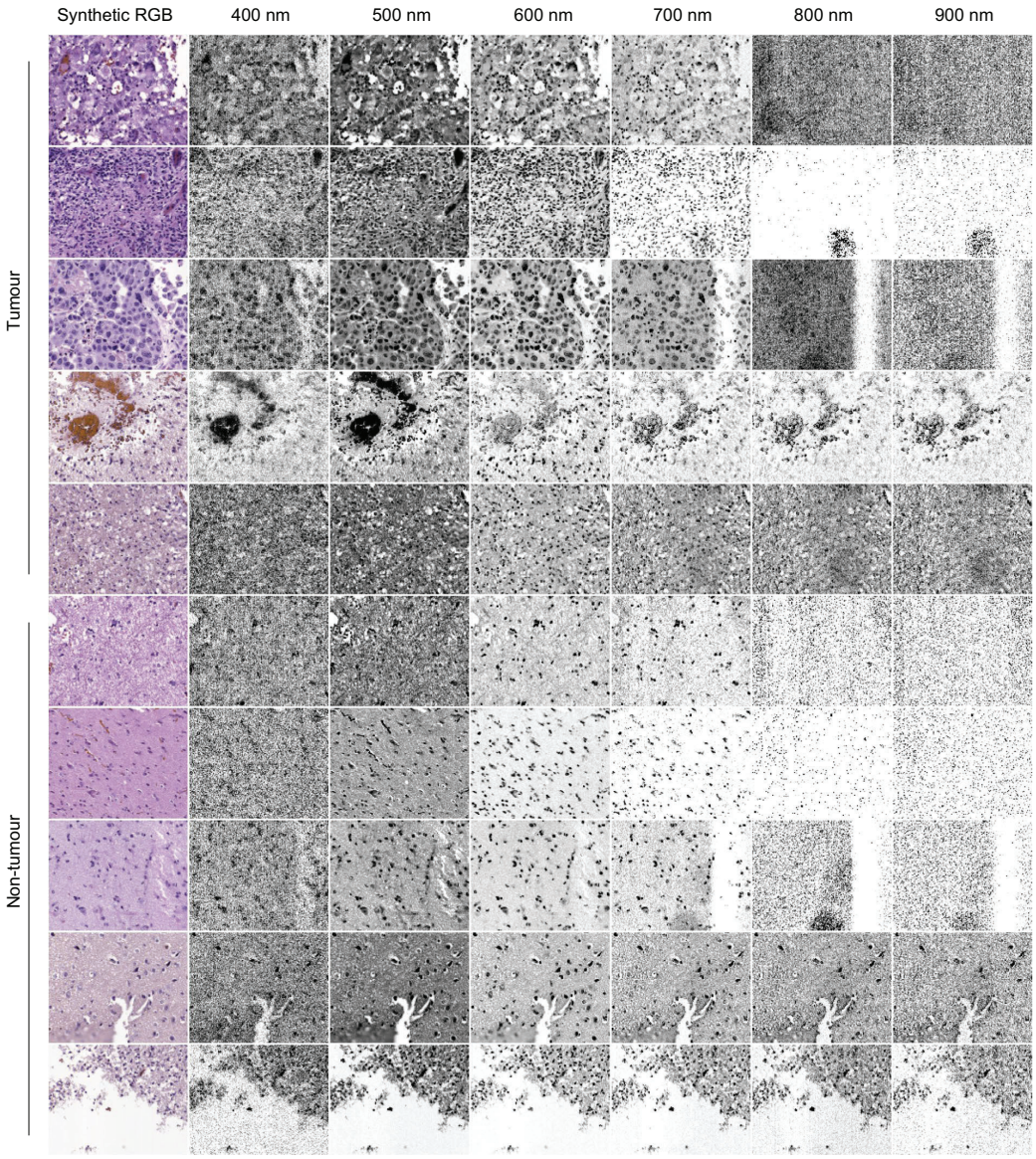


Fig. 7 Examples of HS images from the HistologyHSI-GB dataset showing the synthetic RGB images and different spectral bands after calibration for tumour and non-tumour samples.

Furthermore, the SNR remains above 30 dB for wavelengths ranging from 448 to 894 nm. The SNR spatial distribution was also calculated over the camera FOV for different spectral bands (Fig. 10b), showing that SNR is evenly distributed over the FOV for the different spectral bands, indicating a uniform spatial distribution.

Spectral characterization. The WCT-2065 polymer (Fig. 9a), a transmittance wavelength calibration standard from Avian Technologies (New London, USA), was employed to conduct the spectral characterization of the microscopic HS system. It represents an alternative designation for NIST (National Institute of Standards and Technology) SRM-2065 standard[48]. Its purpose lies in facilitating the calibration of spectrophotometers, covering the wavelength range of 400–2200 nm. The standard uses a glass filter material that incorporates a combination of rare earth oxides. This glass composition includes holmium oxide, samarium oxide, ytterbium oxide, and neodymium oxide, which are blended with lanthanum, boron, silicon, and zirconium oxides found in the base glass. The resulting combination of these oxides creates a filter material with specific optical properties suitable for calibration purposes.

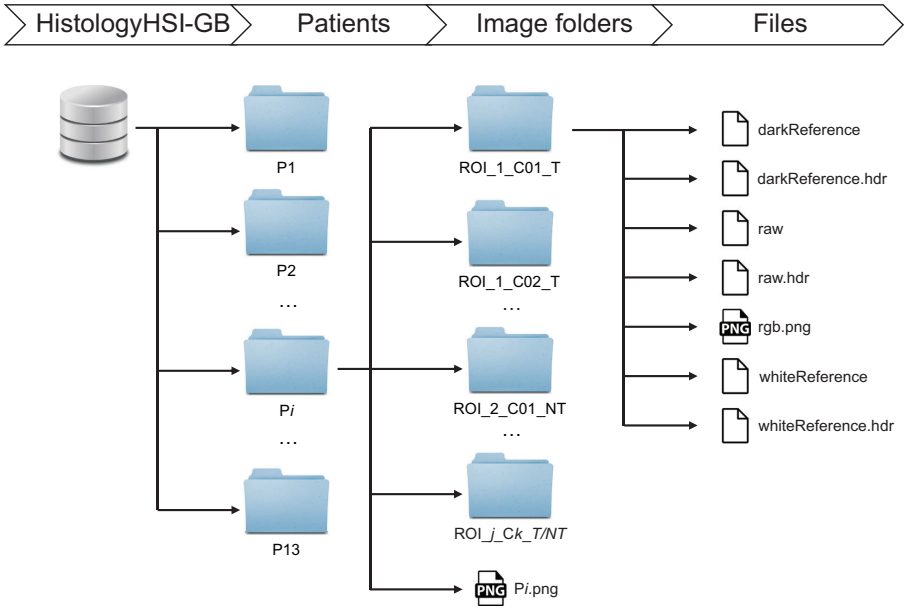


Fig. 8 Graphical representation of the HistologyHSI-GB dataset structure.

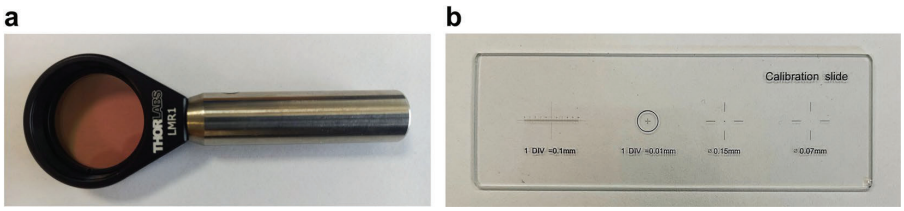


Fig. 9 Spectral and spatial calibration targets. (**a**) Certified WCT-2065 polymer. (**b**) 0.01 mm microscope slide reticule.

File name	Description
darkReference	ENVI binary file containing the dark reference used for calibration.
darkReference.hdr	ENVI header for the dark reference.
raw	ENVI binary file containing the histological HS data.
raw.hdr	ENVI header for the raw file.
rgb.png	Synthetic RGB image extracted from the HS cube.
whiteReference	ENVI binary file containing the white reference used for calibration.
whiteReference.hdr	ENVI header for the white reference.

Table 2. Brief description of the different files contained in each folder in the dataset.

An HS image of the WCT-2065 polymer was captured using the microscopic HS system and further pre-processed. This calibration standard can qualitatively validate the spectral quality of the employed instrumentation (Fig. 11). However, a systematic approach is required for a more accurate and thorough calibration. In order to perform the quantitative validation, the Pearson correlation coefficient (PCC) and root mean square error (RMSE) Eq. (3) were employed to measure the difference between two sets of data. PCC measure the degree of linear anti-correlation or correlation in the range $[-1, 1]$, where -1 indicates perfectly linearly anti-correlated data and 1 indicates perfectly linearly correlated data and it is computed following the Eq. (4). In addition, local maxima and minima were found to detect the most significant signal peaks. Thus, similar peaks were identified both in the captured image and the reference (represented by red and black crosses in Fig. 11a,

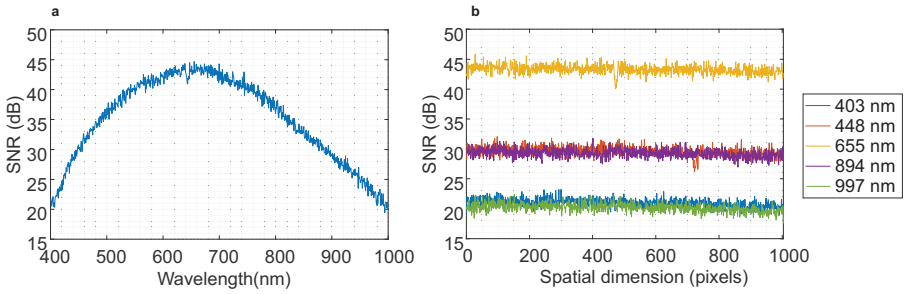


Fig. 10 SNR of the microscopic HS system: (**a**) over the spectral range for the central pixel of the push-broom frame and (**b**) its spatial distribution for different wavelengths (blue: 403 nm, orange: 448 nm, yellow: 655 nm, purple: 894 nm and green: 997 nm).

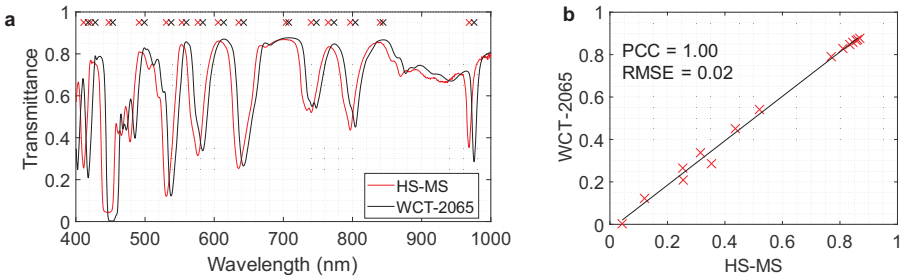


Fig. 11 Spectral characterization of the microscopic HS system. (**a**) Manufactured certified spectral signature of the WCT-2065 polymer (black line) and spectral signature captured by the microscopic HS system (red line). (**b**) Pearson Correlation Coefficient between WCT-2065 and the measured HS peak values.

respectively), resulting in a mean wavelength difference spectra shift of 6.60 *nm* between them. Furthermore, Fig. 11b shows that the PCC between WCT-2065 and the measured HS peak absorbance values provided good value, as well as, it RMSE ($PCC = 1$ and $RMSE = 0.02$). Thus, this NIST traceable standard allows accurate and reliable measurements of the spectral reliability of the HS acquisition system.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{3}$$

$$PCC = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{4}$$

Spatial characterization. Spatial resolution, the ability of a camera to capture fine details and distinguish between separate objects, is also a critical feature in imaging systems. It determines the smallest size of an object that can be recorded. This parameter is essential in applications like histological diagnosis, where identifying small details is essential. Accurate spatial resolution characterization enables improved system performance and precise analysis in various fields, including histopathology[49]. Firstly, the camera is manually aligned to capture the information properly[50]. Then, the spatial resolution of the microscopic HS system was evaluated both theoretically and empirically. The theoretical calculation of the FOV, shown in Eq. (5), considered factors such as pixel size (Ps), number of pixels (N), magnification (M_i), and sensor size (S_s).

$$FOV = \frac{Ps \cdot N}{M_i} = \frac{S_s}{M_i} \tag{5}$$

An empirical test using a micrometre ruler (Fig. 9b) provides further insight into the spatial resolution capabilities of the cameras. In order to perform this test, a $Y\lambda$ spatial profile of the ruler (Fig. 12a) and its first derivative (Fig. 12b) was analysed to determine the mean distance between peaks (local minima peaks signalled with red crosses and local maxima peaks with green crosses). The results of the theoretical calculation provide a pixel

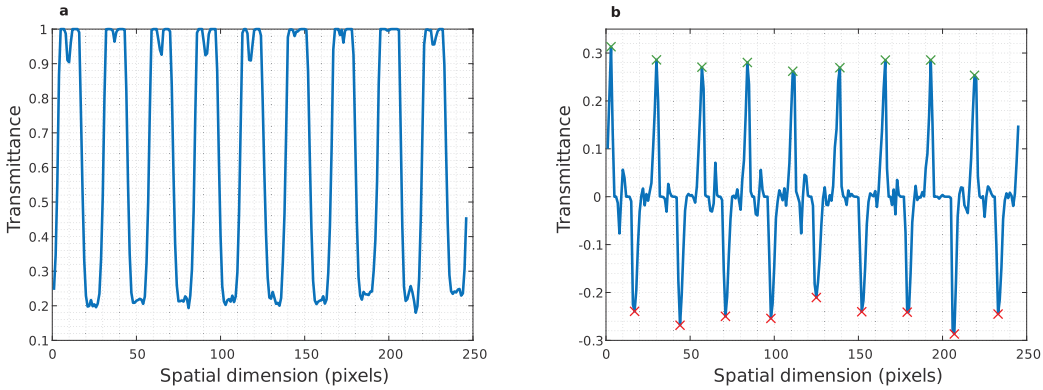


Fig. 12 Pixel size validation using a micrometre ruler. (**a**) Profile of $Y\lambda$ frame extracted from the $Y\lambda$ frame and (**b**) its first derivative where red crosses are local minima peaks and green crosses local maxima peaks.

size of 0.3700 μm and the empirical one is 0.3697 μm. This method confirmed that the spatial resolution of the microscopic HS system matches the theoretical pixel size with an average error of less than 0.0003 μm.

Once the pixel size has been calculated and the camera has been visually aligned, together with the information about the mechanical resolution and frame rate of the camera, the required motor rotation speed of the mechanical stage is determined. However, an additional analysis was conducted to further improve and verify the correct configuration of the scanning parameters. The entire HS acquisition system is considered as a whole, including the microscope, camera, and movement mechanism. For this evaluation, the goal is to capture an image of a circle printed in a calibration slide (*dot target*) and evaluate its appearance to identify camera misalignments and suboptimal movement speeds. The circle appears as a perfect rounded circle when captured at the correct speed but appears as an ellipse when the speed is too high or too low. While visual inspection provides a relatively good assessment, an automatic methodology[50] is needed for a more precise and rigorous calibration. First, principal component analysis (PCA) is employed to find the directions of the longest and shortest axes of the ellipse (ϕ_{min} and ϕ_{max}). Then, eccentricity can be calculated following Eq. (6), where a perfect circle would provide values close to zero. In our case, the dot target from the calibration slide was captured, and its eccentricity was computed providing accurate results ($e = 0.176$). Thus, the microscopic HS system is properly calibrated in the spatial domain, and it is possible to acquire HS images under satisfactory conditions.

$$e = \sqrt{1 - \frac{(\phi_{min})^2}{(\phi_{max})^2}} \tag{6}$$

Usage Notes

Recommended pre-processing. The pre-processing framework applied to each HS cube is based on standard calibration and spectral band reduction. First, HS images are transformed from radiance to normalized transmittance by calibration. As a result of the strong correlation of spectral information between adjacent spectral bands, we propose to reduce the spectral dimensionality of the original data. A spectrally reduced HS image is generated by averaging the spectral bands of adjacent neighbouring bands to perform this band reduction. Using a spectral window of three neighbours, this process reduces the original 826 bands to 275, while slightly decreasing the presence of white Gaussian noise. Furthermore, reducing the number of bands proves to be advantageous in terms of reducing the computational cost of subsequent image processing tasks. However, this band reduction is optional, depending on the further processing interest. Additionally, for image analysis involving the spectral analysis of the samples, it is recommended to perform a background sample segmentation, where the pixels corresponding to the tissue and the background light of the microscope are identified. Finally, to use classification methods, the label (tumour or non-tumour) of each HS cube should be extracted from the folder name.

Recommended data partition and data HS processing applications. To perform machine learning analysis, an unbiased data partition should be performed. The dataset used for this study poses three challenges. First, the dataset is limited in the number of patients (13 patients). Second, samples containing both classes (tumour and *non*-tumour) are only available for 8 patients. Hence, the non-tumour samples information is limited in terms of patients. Third, the dataset is unbalanced, with more images annotated as non-tumour. In previous works[36,38], a data partition based on 4 different folds was employed. Furthermore, spectral unmixing techniques could be performed as a preprocessing stage prior to classification[51], or they can be used to determine the abundances of known endmembers of the images, specifically identifying the proportions of the H&E stains in each pixel[52].

Limitations and future perspectives. The dataset has several limitations. As previously mentioned, the primary limitation is its relatively small cohort, consisting of data from only 13 patients. Furthermore, information for both classes of interest, tumour, and non-tumour, is available for only 8 of these patients. This leads to an imbalanced dataset, with a predominance of images classified as non-tumour. Such an imbalance could potentially introduce bias and affect the generalizability of the findings derived from this dataset.

Another limitation is related to the type of annotations available in this dataset. The macroscopic annotations of tumour and non-tumour regions on the pathological slides, leading to only image-level annotations for the HS images. A more sophisticated method for digitally annotating the images would allow to identify regions where tumour and non-tumour tissues are adjacent, making possible to capture regions comprising both classes in a single HS image. More detailed digital annotation would help in further validating the classification algorithms on a pixel-by-pixel basis and could also offer potential for other methods such as unsupervised learning or spectral unmixing. However, more detailed annotations would significantly increase the time and manual effort required to label each image.

Finally, this dataset is focused on images captured on a single magnification (20×). The motivation of using the higher magnification available for the instrumentation was driven by the need to capture detailed cell-level information from the histological slides. However, creating a dataset containing the same images at different magnifications could be of potential interest and benefit to the scientific community.

In summary, future datasets of HS histological samples will need to include a larger number of patients, ensure a balanced representation of the various classes of interest, incorporate more detailed annotations, and provide images at various magnification levels.

Code availability

A tutorial on how to read and display HS data is available in a public repository: https://github.com/HIRIS-Lab/HistologyHSI-GB. These tutorials include the use of custom MATLAB and Python functions and some of the most common toolbox/libraries.

Received: 28 January 2024; Accepted: 10 June 2024;
Published: 24 June 2024

References

1. Goetz, A. F. H., Vane, G., Solomon, J. E. & Rock, B. N. Imaging Spectrometry for Earth Remote Sensing. *Science (1979)* **228**, 1147–1153 (1985).
2. Khan, A., Vibhute, A. D., Mali, S. & Patil, C. H. A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. *Ecol Inform* **69**, 101678 (2022).
3. Sethy, P. K., Pandey, C., Sahu, Y. K. & Behera, S. K. Hyperspectral imagery applications for precision agriculture - a systemic survey. *Multimed Tools Appl* **81**, 3005–3038 (2022).
4. Saha, D. & Manickavasagan, A. Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. *Curr Res Food Sci* **4**, 28–44 (2021).
5. Ortega, S., Lindberg, S.-K., E. Anderssen, K. & Heia, K. Perspective Chapter: Hyperspectral Imaging for the Analysis of Seafood. in *Hyperspectral Imaging - A Perspective on Recent Advances and Applications* https://doi.org/10.5772/intechopen.108726 (IntechOpen, 2023).
6. Kang, Z. *et al.* Advances in Machine Learning and Hyperspectral Imaging in the Food Supply Chain. *Food Engineering Reviews* **14**, 596–616 (2022).
7. Zheng, Y., Bai, J., Xu, J., Li, X. & Zhang, Y. A discrimination model in waste plastics sorting using NIR hyperspectral imaging system. *Waste Management* **72**, 87–98 (2018).
8. Bonifazi, G., Capobianco, G. & Serranti, S. A hierarchical classification approach for recognition of low-density (LDPE) and high-density polyethylene (HDPE) in mixed plastic waste based on short-wave infrared (SWIR) hyperspectral imaging. *Spectrochim Acta A Mol Biomol Spectrosc* **198**, 115–122 (2018).
9. Cucci, C. *et al.* Reflectance hyperspectral data processing on a set of Picasso paintings: which algorithm provides what? A comparative analysis of multivariate, statistical and artificial intelligence methods. in *Optics for Arts, Architecture, and Archaeology VIII* (eds. Groves, R. & Liang, H.) 1. https://doi.org/10.1117/12.2593838 (SPIE, 2021).
10. Balas, C., Epitropou, G., Tsapras, A. & Hadjinicolaou, N. Hyperspectral imaging and spectral classification for pigment identification and mapping in paintings by El Greco and his workshop. *Multimed Tools Appl* **77**, 9737–9751 (2018).
11. Huang, S.-Y. *et al.* Recent Advances in Counterfeit Art, Document, Photo, Hologram, and Currency Detection Using Hyperspectral Imaging. *Sensors* **22**, 7308 (2022).
12. Koz, A. Ground-Based Hyperspectral Image Surveillance Systems for Explosive Detection: Part I—State of the Art and Challenges. *IEEE J Sel Top Appl Earth Obs Remote Sens* **12**, 4746–4753 (2019).
13. Karim, S., Qadir, A., Farooq, U., Shakir, M. & Laghari, A. A. Hyperspectral Imaging: A Review and Trends towards Medical Imaging. *Curr Med Imaging Rev* **19**, 417–427 (2022).
14. Fei, B. Hyperspectral imaging in medical applications. in 523–565. https://doi.org/10.1016/B978-0-444-63977-6.00021-3 (2019).
15. Barberio, M. *et al.* Intraoperative Guidance Using Hyperspectral Imaging: A Review for Surgeons. *Diagnostics* **11**, 2066 (2021).
16. Shapey, J. *et al.* Intraoperative multispectral and hyperspectral label-free imaging: A systematic review of *in vivo* clinical studies. *J Biophotonics* **12** (2019).
17. Mangotra, H., Srivastava, S., Jaiswal, G., Rani, R. & Sharma, A. Hyperspectral imaging for early diagnosis of diseases: A review. *Expert Syst* https://doi.org/10.1111/exsy.13311 (2023).
18. Johansen, T. H. *et al.* Recent advances in hyperspectral imaging for melanoma detection. *WIREs Computational Statistics* **12** (2020).
19. Reshef, E. R., Miller, J. B. & Vavvas, D. G. Hyperspectral Imaging of the Retina: A Review. *Int Ophthalmol Clin* **60**, 85–96 (2020).
20. Fouad Aref, M. H., Sharawi, A. A. R. & El-Sharkawy, Y. H. Delineation of the Arm Blood Vessels Utilizing Hyperspectral Imaging to Assist with Phlebotomy for Exploiting the Cutaneous Tissue Oxygen Concentration. *Photodiagnosis Photodyn Ther* **33**, 102190 (2021).
21. Tomanic, T. *et al.* Estimating quantitative physiological and morphological tissue parameters of murine tumor models using hyperspectral imaging and optical profilometry. *J Biophotonics* **16** (2023).
22. Köhler, H. *et al.* Evaluation of hyperspectral imaging (HSI) for the measurement of ischemic conditioning effects of the gastric conduit during esophagectomy. *Surg Endosc* **33**, 3775–3782 (2019).
23. Sucher, R. *et al.* Hyperspectral Imaging (HSI) of Human Kidney Allografts. *Ann Surg* **276**, e48–e55 (2022).

24. Fuchs, T. J. & Buhmann, J. M. Computational pathology: Challenges and promises for tissue analysis. *Computerized Medical Imaging and Graphics* **35**, 515–530 (2011).
25. Louis, D. N. *et al*. Computational Pathology: A Path Ahead. *Arch Pathol Lab Med* **140**, 41–50 (2016).
26. Ortega, S., Halicek, M., Fabelo, H., Callico, G. M. & Fei, B. Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review [Invited. *Biomed Opt Express* **11**, 3195 (2020).
27. Sun, L. *et al*. Diagnosis of cholangiocarcinoma from microscopic hyperspectral pathological dataset by deep convolution neural networks. *Methods* **202**, 22–30 (2022).
28. Deng, Y. *et al*. ResNet-50 based Method for Cholangiocarcinoma Identification from Microscopic Hyperspectral Pathology Images. *J Phys Conf Ser* **1880**, 012019 (2021).
29. Ma, L. *et al*. Automatic detection of head and neck squamous cell carcinoma on histologic slides using hyperspectral microscopic imaging. *J Biomed Opt* **27** (2022).
30. Lv, M. *et al*. Membranous nephropathy classification using microscopic hyperspectral imaging and tensor patch-based discriminative linear regression. *Biomed Opt Express* **12**, 2968 (2021).
31. Wang, J. *et al*. PCA-U-Net based breast cancer nest segmentation from microarray hyperspectral images. *Fundamental Research* **1**, 631–640 (2021).
32. Duan, Y. *et al*. Leukocyte classification based on spatial and spectral features of microscopic hyperspectral images. *Opt Laser Technol* **112**, 530–538 (2019).
33. Wang, Q. *et al*. A 3D attention networks for classification of white blood cells from microscopy hyperspectral images. *Opt Laser Technol* **139**, 106931 (2021).
34. Ortega, S. *et al*. Information Extraction Techniques in Hyperspectral Imaging Biomedical Applications. in *Multimedia Information Retrieval*. https://doi.org/10.5772/intechopen.93960 (IntechOpen, 2021).
35. Zhang, Y., Wang, Y., Zhang, B. & Li, Q. A hyperspectral dataset of precancerous lesions in gastric cancer and benchmarks for pathological diagnosis. *J Biophotonics* **15** (2022).
36. Ortega, S. *et al*. Hyperspectral Imaging for the Detection of Glioblastoma Tumor Cells in H&E Slides Using Convolutional Neural Networks. *Sensors* **20**, 1911 (2020).
37. Ortega Sarmiento, S. Automatic classification of histological hyperspectral images: algorithms and instrumentation, Universidad de Las Palmas de Gran Canaria. https://hdl.handle.net/10553/107311 (2021).
38. Ortega, S. *et al*. Hyperspectral Superpixel-Wise Glioblastoma Tumor Detection in Histological Samples. *Applied Sciences* **10**, 4448 (2020).
39. Louis, D. N. *et al*. *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A Summary*. *Acta Neuropathologica* vol. 131 803–820 (Springer Berlin Heidelberg, 2016).
40. Gómez-Chova, L. *et al*. Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images. *Appl Opt* **47**, F46–60 (2008).
41. Wald, G. The Receptors of Human Color Vision. *Science (1979)* **145**, 1007–1016 (1964).
42. Ortega, S. *et al*. *Hyperspectral Histological Images for Diagnosis of Human Glioblastoma (HistologyHSI-GB)*. https://doi.org/10.7937/Z1K6-VD17 (2024).
43. Clark, K. *et al*. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J Digit Imaging* **26**, 1045–1057 (2013).
44. Triglav, J. Exelis Visual Information Solutions. *Geoinformatics, Emmeloord* **15**, 34–37 (2012).
45. *Physics-Based Vision: Principles and Practice*. https://doi.org/10.1201/9781439865880 (A K Peters/CRC Press, 1993).
46. 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).
47. Ortega, S. *et al*. Technical Validation data to support the quality of the HistologyHSI-GB dataset. *Figshare*, https://doi.org/10.6084/m9.figshare.23659170.v1 (2023).
48. Choquette, S. J., Duewer, D. L., Hanssen, L. M. & Early, E. A. Standard Reference Material 2036 Near-Infrared Reflection Wavelength Standard. *Appl Spectrosc* **59**, 496–504 (2005).
49. Intriligator, J. & Cavanagh, P. The Spatial Resolution of Visual Attention. *Cogn Psychol* **43**, 171–216 (2001).
50. Ortega, S. *et al*. Hyperspectral Push-Broom Microscope Development and Characterization. *IEEE Access* **7**, 122473–122491 (2019).
51. Cruz-Guerrero, I. A. *et al*. Hybrid Brain Tumor Classification Scheme of Histopathology Hyperspectral Images Using Linear Unmixing and Deep Learning. *SSRN Electronic Journal* https://doi.org/10.2139/ssrn.4292742 (2022).
52. Gibbs, S. L. *et al*. Near-Infrared Fluorescent Digital Pathology for the Automation of Disease Diagnosis and Biomarker Assessment. *Mol Imaging* **14**, 7290.2015.00005 (2015).

Acknowledgements

This work has been supported by Spanish Government and European Union (FEDER funds) in the context of TALENT-HExPERIA (HypErsPEctRal Imaging for Artificial intelligence applications) project (PID2020-116417RB-C42 AEI/10.13039/501100011033). Moreover, this work was completed while Laura Quintana-Quintana and Raquel Leon were beneficiary of the pre-doctoral grant given by the "Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI)" of the "Consejería de Economía, Conocimiento y Empleo", which is part-financed by the European Social Fund (FSE) (POC 2014–2020, Eje 3 Tema Prioritario 74 (85%)). Furthermore, Himar Fabelo was also beneficiary of the FJC2020-043474-I funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR".

Author contributions

S.O. designed and established the experimental system, conceived the study, generated the hyperspectral dataset, performed image processing and result analysis, and wrote the manuscript. L.Q.-Q. deposited the dataset, conducted the technical validation experiments, performed image processing and result analysis, and wrote the manuscript. R.L. deposited the dataset, supervised the technical validation, performed image processing and result analysis, and wrote the manuscript. H.F. Generated the hyperspectral dataset, supervised the study, and reviewed the manuscript. M.P. and R.C. prepared and diagnosed the histological samples and reviewed the manuscript. G.M.C. acquired the funding, supervised the study, and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.O.

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.

J5. Histological Hyperspectral Breast Cancer Recurrence Database (HistologyHSI-BC Recurrence)

Title:	Histological Hyperspectral Breast Cancer Recurrence Database (HistologyHSI-BC Recurrence)		
Authors:	**Laura Quintana-Quintana**[¥], Esther Sauras-Colón[¥], Alessio Fiorin, Javier Santana-Nunez, Samuel Ortega, Noèlia Gallardo-Borràs3,2, Alba Fischer-Carles, Tábata Sánchez-Alcántara, Himar Fabelo, Laia Adalid-Llansa, Daniel Mata-Cano, Ramon Bosch-Príncep, Marylène Lejeune, Gustavo M. Callico, Carlos López-Pablo		
Journal:	Scientific Data		
Publisher:	Nature Portfolio		
Date: 2025	**Vol.**: 12	**Pages**: 1886	
doi:	https://doi.org/10.1038/s41597-025-06157-4		
JIF Quartile:	Q1 in Multidisciplinary Sciences	**IF**:	5.8
Open Access, Open Code, Open Data, National Collaboration			

[¥]Equal Contribution

scientific data

OPEN

DATA DESCRIPTOR

Histological Hyperspectral Breast Cancer Recurrence Database (HistologyHSI-BC Recurrence)

Laura Quintana-Quintana[1,9], Esther Sauras-Colón[2,3,9] ✉, Alessio Fiorin[2,3,4], Javier Santana-Nunez[1,5,6], Samuel Ortega[1,7], Noèlia Gallardo-Borràs[2,3], Alba Fischer-Carles[2,3,8], Tábata Sánchez-Alcántara[2,3], Himar Fabelo[1,5,6], Laia Adalid-Llansa[2,3], Daniel Mata-Cano[2,3], Ramon Bosch-Príncep[2,3], Marylène Lejeune[2,3], Gustavo M. Callico[1] & Carlos López-Pablo[2,3]

Metastasis occurs in nearly 1 out of 3 breast cancer (BC) patients and significantly reduces survival rates, particularly in cases of distant metastases. As most distant metastases develop after diagnosis (i.e., recurrence) and remain incurable, there is a critical need for prognostic biomarkers to assess recurrence risk. Multimodal data analysis has emerged as a promising approach to integrate diverse information, offering a more comprehensive perspective. This study introduces the Histology HSI-BC (hyperspectral imaging - breast cancer) Recurrence Database, the first publicly accessible multimodal database designed to advance BC distant recurrence prediction. The database comprises 47 histopathological whole-slide images, 677 hyperspectral (HS) images, and clinical and demographic data from 47 BC patients, of whom 22 (47%) experienced distant recurrence over a 12-year follow-up. Histopathological slides were digitized using a whole-slide scanner and annotated by expert pathologists, while HS images were acquired with an HS camera coupled to a bright-field microscope. This database provides a promising resource for studying BC recurrence prediction and personalized treatment strategies by integrating the aforementioned multimodal data.

Background & Summary

In 2022, breast cancer (BC) was the most common type of cancer in women, with an incidence of 23.8%, and the leading cause of cancer-related death among women, accounting for 15.4% of all cancer-related deaths[1]. Cancer cells can spread from the primary tumor to other parts of the body, which is known as metastasis and is the main cause of death in most cancers[2,3]. Metastasis occurs in nearly 1 out of 3 patients diagnosed with BC and can appear in the axillary lymph nodes (regional metastasis) or in other organs (distant metastasis)[4]. Overall, the 5-year survival rate after diagnosis of BC is 91%. However, this rate is higher in patients with tumors located exclusively in the breast (99%) than in patients with regional metastasis (86%) or with distant metastasis (31%), where survival decreases dramatically[4]. Women with distant metastases may have either *de novo* distant spread, where distant metastases are already present at the time of diagnosis, or develop distant metastases after an initial diagnosis and treatment, which is known as *recurrence*[5]. While de novo cases account for approximately 25% of metastatic BC diagnoses, the majority result from recurrence[6]. Whether de novo or recurrent, distant metastases remain incurable[7,8].

[1]Institute for Applied Microelectronics, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain. [2]Department of Pathology, Hospital de Tortosa Verge de la Cinta, Institut Català de la Salut, Tortosa, Spain. [3]Oncological Pathology and Bioinformatics Research Group, Institut d'Investigació Sanitària Pere Virgili, Tortosa, Spain. [4]Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Tarragona, Spain. [5]Research Unit, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas de Gran Canaria, Spain, . [6]Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Las Palmas de Gran Canaria, Spain. [7]Norwegian Institute of Food, Fisheries and Aquaculture Research (Nofima), Tromsø, Norway. [8]BCN MedTech, Department of Engineering, Universitat Pompeu Fabra, Barcelona, Spain. [9]These authors contributed equally: Laura Quintana-Quintana, Esther Sauras-Colón. ✉e-mail: esauras.ebre.ics@gencat.cat

Certain classic prognostic factors are associated with the risk of developing distant metastasis, such as *age*, *tumor diameter*, *stage*, *tumor grade*, *tumor type* or *lymphovascular invasion* (LVI)[8]. Additionally, studies have identified other biomarkers with prognostic value in the disease that may be associated with metastasis, including genetic alterations, circulating tumor cells and circulating tumor DNA, biomarkers of response to immunotherapy and gene expression platforms to predict the risk of recurrence[9–11]. However, to date there is no consensus for the implementation of most of these biomarkers in routine clinical practice. Therefore, there continues to be a growing interest in identifying specific prognostic biomarkers that allow determining the probability of developing metastasis.

Cancer detection relies heavily on imaging methods like X-ray, ultrasound, and magnetic resonance imaging[12]. However, treatment decisions require a conclusive histopathological diagnosis, which is obtained from a tissue biopsy. BC can be broadly categorized into in situ carcinoma and invasive carcinoma. Among these, ductal carcinoma in situ (DCIS) represents the most prevalent subtype of in situ carcinoma, while invasive ductal carcinoma (IDC) is the most common subtype of invasive carcinoma. Nevertheless, given the heterogeneity of BC, the accurate identification of these subtypes among other histological subtypes requires extensive expertise and a deep understanding of breast pathology[13]. The rise of digital pathology, which leverages whole-slide images (WSIs), has revolutionized research and diagnosis in pathology, particularly in cancer, by enabling more efficient data sharing across institutions and promoting remote collaborations. WSIs are high-resolution digital images of traditional glass pathology slides, which can be viewed, analyzed, and shared on a computer screen[14,15]. The use of WSIs also paves the way for computational pathology, which started from the use of traditional image analysis methods to advanced machine learning (ML) and deep learning (DL) algorithms[16,17]. Remarkably, these novel approaches offer the potential to integrate multiple data modalities, extending beyond histopathology image analysis. This includes linking histopathological images with clinical factors, such as prognosis and genetic mutations, thereby enhancing BC diagnostics[18–20].

Beyond conventional methods, other imaging modalities show promising potential for improving the diagnosis and prognosis of BC patients. Among these, hyperspectral (HS) imaging (HSI), combines traditional imaging with spectroscopy to capture both spatial and spectral information. Each material interacts uniquely with emitted radiation, reflecting and absorbing it in a way that creates a distinct radiance vector, often named *spectral signature*. HSI sensors can capture these spectral signatures, acquiring significantly more data than standard RGB (Red, Green, Blue) cameras and extending imaging capabilities beyond human vision (e.g., near-infrared (NIR) HS sensors can capture wavelengths ranging from 900 to 1,700 nm)[21]. In recent years, the use of HSI in medicine has begun to achieve promising results regarding cancer detection by utilizing cutting-edge ML algorithms to process the high amount of HS data[22–24]. In the existing literature for medical histological applications, HSI has been used to identify pancreatic neoplasms with different prognoses[25], quantify Ki67 as a prognostic factor in lymphomas[26] and study the interactions between tumor cells and immune cells of the tumor microenvironment in response to immunotherapy in lung cancer[27], obtaining promising results.

Research in this area is still in its early stages, and the number of published studies remains limited. Regarding WSI and clinical and demographic databases, The Cancer Genome Atlas (TCGA) is one of the main publicly available sources for hematoxylin and eosin (H&E)-stained WSIs and associated clinical and demographic data[28]. The main challenge of this database is the lack of annotations, which makes the subsequent analysis of these WSIs difficult. The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) is another database that provides clinical, demographic, and molecular data of over 2,000 BC cases[29]. However, it primarily focuses on genomic and transcriptomic data and does not include WSIs. Several publicly available databases contain H&E-stained WSIs[30], but most include annotations focused on specific cell types in BC, such as tumor-infiltrating lymphocytes and inflammatory cells. We have identified one database - Breast Cancer Semantic Segmentation (BCSS) that provides specific annotations of tissue compartments, distinguishing between tumor and healthy tissue[31]. Regarding medical HS databases only two major sets were identified: in 2022, Zhang *et al*.[32] introduced a large-scale database for HS microscopic images of precancerous lesions in gastric cancer, and in 2024, Ortega *et al*. released the HistologyHSI-GB dataset[33], focused on HS glioblastoma histology. None of those databases included clinical or demographic data, this being one of the main challenges in this field. There is a limited availability of comprehensive, high-quality databases, which hinders the broader application of clinical and demographic data, WSIs and HSI in clinical practice and research. The lack of such databases makes it difficult to fully explore the potential of HSI together with conventional practices for diagnosing and predicting disease outcomes, such as recurrence in BC.

This paper presents a publicly accessible database designed to investigate specific prognostic biomarkers for predicting the likelihood of BC recurrence due to distant metastasis. The HistologyHSI-BC Recurrence Database includes clinical and demographic data from BC patients, along with WSIs and HS images obtained from their primary tumor samples. This database is intended to evaluate the ability to predict recurrence due to distant metastasis over a 12-year follow-up period. Biopsies from 47 patients diagnosed with BC were extracted, sliced and stained with H&E, 47 WSIs and 677 microscopic HS images were taken, and their clinical and demographic data were collected. Among these patients, 22 experienced distant recurrence. A schematic overview of the study workflow is presented in Fig. 1.

The HistologyHSI-BC Recurrence Database will benefit researchers by providing a comprehensive, multimodal database that integrates WSIs, HS images, and clinical and demographic data from BC patients. This resource enables the development and refinement of predictive models for BC recurrence due to distant metastasis, starting to fill a significant gap in available databases. Researchers can leverage this data to explore innovative ML approaches, enhance diagnostic accuracy, and identify novel biomarkers for BC recurrence. Additionally, the database promotes reproducibility, facilitates collaboration across institutions, and accelerates research in computational pathology, potentially improving personalized treatment strategies and benefiting broader cancer research efforts.

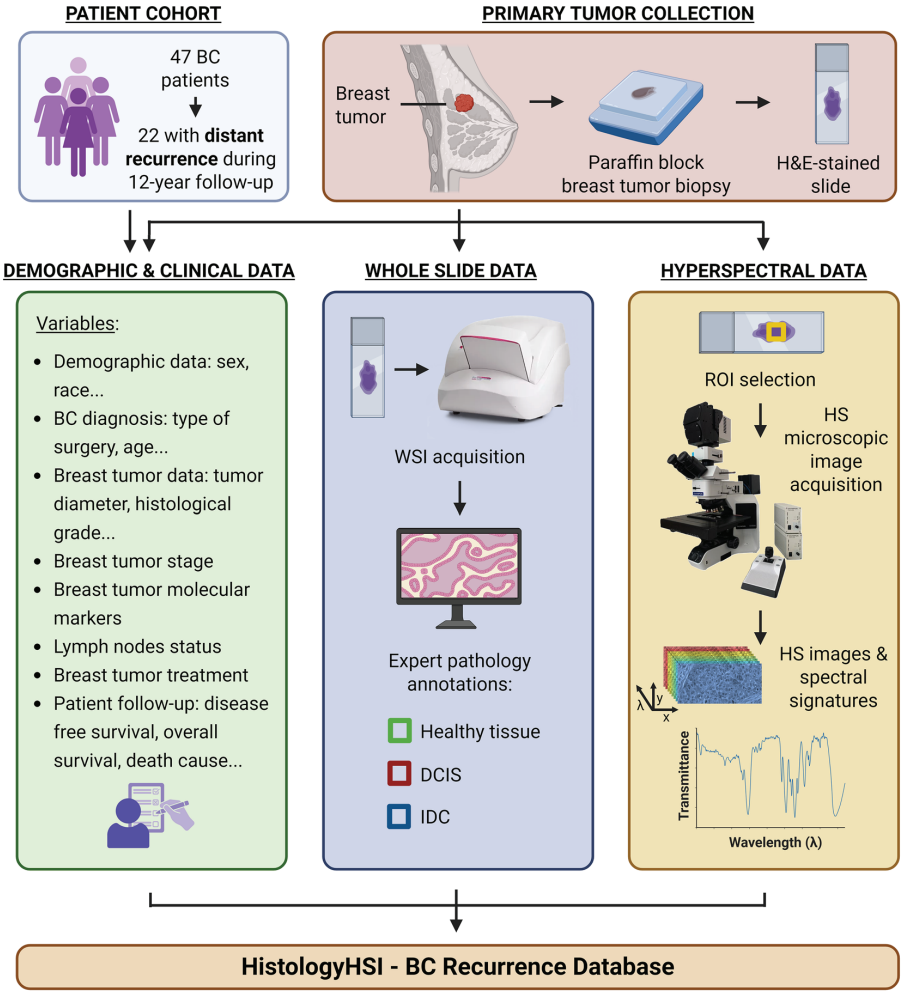


Fig. 1 Schematic overview of the study design or workflow. The patient cohort comprises 47 BC patients, of which 22 (47%) experienced distant recurrence over a 12-year follow-up period and 25 (53%) did not. All these patients had a breast tumor biopsy, which was collected, prepared, and stained with H&E. The H&E-stained slides were digitized to obtain WSIs and annotated to differentiate three tissue compartments: IDC, healthy tissue and DCIS. Within each annotated area, ROIs were selected, from which HS images were acquired to generate HS cubes and extract spectral signatures. Together with the clinical and demographic data, all this information constitutes the HistologyHSI - BC Recurrence Database.

Methods

Patients selection, eligibility criteria and ethics approval. This is a retrospective case-control study carried out on 47 BC patients diagnosed with IDC, now called invasive breast carcinoma of no special type[34], between 2006 and 2015, who met the eligibility criteria for inclusion (Table 1). Cases include 22 patients who experienced recurrence due to distant metastasis during the 12 years following diagnosis. The remaining 25 patients who did not experience recurrence during the 12 years of follow-up are included as control group.

The study was approved by the Drug Research Ethics Committee of the Institut d'Investigació Sanitària Pere Virgili (IISPV), Tarragona, Spain, under reference number 134/2022. The samples used in this study were obtained from Biobank IISPV-Node Tortosa, Tarragona, Spain, following the principles of ethical conduct and data protection. The Biobank has approved the open publication of the data associated with this work. All participants whose samples were stored in the biobank have previously signed an informed consent form, explicitly authorizing the collection, storage, and future use of their biological materials and associated data for research purposes. The processing, communication and transfer of personal data of all participants comply with the provisions of Organic Law 3/2018, of December 5, on the Protection of Personal Data and Guarantee of Digital Rights and with Regulation (EU) 2016/679 of the European Parliament and of the Council, of April 27, 2016, on

Inclusion	Exclusion
A diagnosis of IDC	Receipt of neoadjuvant treatment, as it is known to modify the tumor microenvironment
Representativeness of IDC tissue in surgical biopsy	Recurrence occurring in the breast rather than in distant organs
A clinical history with complete clinical and pathological data	Presence of distant metastases at the time of diagnosis
Patient's agreement to be included in the study	Failure to meet any of the inclusion criteria

Table 1. Eligibility criteria for patient inclusion.

the protection of natural persons with regard to the processing of personal data and the free circulation of these data, and repealing Directive 95/46/EC (General Data Protection Regulation). The data generated and collected during this study are anonymized to ensure the security of the information, safeguarding the confidentiality and privacy of the patients.

Data Collection

Clinical and demographic data. The data collection process involved extracting information from clinical records, including demographic and clinical data, which were following Table 2.

Histopathology WSIs. Paraffin blocks of primary tumor biopsies with sufficient representative IDC tissue were obtained from the Biobank IISPV-Node Tortosa, Tarragona, Spain. The samples were processed in the Pathology Department, where 2 µm-thick sections were prepared from each paraffin block and stained with H&E according to the instructions of the manufacturer. The slides were sealed with coverslips using dibutylphthalate polystyrene xylene (DPX) mounting medium for subsequent digitization and HS microscopic image acquisition.

The H&E-stained slides were digitized with the Pannoramic 250 Flash III WSI scanner (3DHISTECH Ltd., Budapest, Hungary) at 20 × magnification (0.2433 µm/pixel) using MRXS image format. WSIs were visualized using QuPath[35] (available at: https://qupath.github.io/) for technical validation and annotation, and anonymized using the SlideMaster software (3DHISTECH Ltd., Budapest, Hungary). The annotation process of each WSI was manually performed by pathologists using diverse colors to distinguish between IDC, healthy tissue, and DCIS. The annotations were made with the following color scheme: IDC was outlined in blue, healthy tissue in green, and DCIS in red (Fig. 2a).

Moreover, within each of these three tissue compartment types, different regions of interest (ROIs), surrounded by yellow line, were identified and annotated to subsequently acquire the HS image using the HS microscopic system (Fig. 2a). These ROIs were selected to ensure the inclusion of representative areas of each class (IDC, healthy and DCIS), capturing relevant spectral variability for further analysis.

HS images. The HS images were captured using a HS microscopic system (Fig. 2b). The system features the Hyperspec® VNIR (Visible and Near Infrared) A-Series camera (HeadWall Photonics, Fitchburg, MA, USA), a pushbroom HS camera that captures data by scanning the sample spatially. The camera is equipped with a charge-coupled device (CCD) sensor that covers a spectral range of 400–1,000 nm, capturing 826 spectral bands across 1,004 spatial pixels per line. It offers high spectral resolution with a slit image full width at half maximum (FWHM) of 2.5 nm and a pixel size of 7.4 µm. Data are acquired with a 12-bit ADC (Analog-to-digital Converter), and each HS line has a size of 1,004 × 826 pixels and requires 1.6 MB per line on disk for storage. The microscope used is the OLYMPUS BX-53 (Olympus, Tokyo, Japan), with LMPLN-IR (5 × , 10 ×) and LCPLN-IR (20 × , 50 ×) objective lenses optimized for infrared imaging. The system uses a 100 W TH4 halogen lamp (Olympus, Tokyo, Japan) as the light source, covering a wavelength range from 400 to 1,800 nm and supporting both transmittance and reflectance light paths. To acquire full HS images, the pushbroom camera requires spatial scanning, which is facilitated by a SCAN 130 × 85 scanning stage (Märzhäuser, Wetzlar, Germany). The stage ensures high precision (± 3 µm accuracy) as it moves the sample, keeping it aligned with the objective and light path. Furthermore, an RGB camera, the acA5472-17uc (Basler AG, Ahrensburg, Germany), provides real-time visualization of the sample to navigate it without the need of using the microscope binoculars. It is a 20 MP compact camera with a Sony IMX183 CMOS sensor (Tokyo, Japan), 5,496 × 3,672 resolution, and 17 fps. It features USB 3.0, a C-mount, and supports hardware/software triggers.

Calibration of the HS images is necessary to ensure the data accurately represents the sample's spectral signatures. The HS microscope captures spectral signatures for each pixel, but factors like the sensor's response, light transmission, and the light source can affect accuracy. The calibration process involves normalizing the pixel values of the HS image by adjusting them based on a white reference (WR) and a dark reference (DR). WR is obtained by focusing on an empty area of the slide at the same working distance. This ensures no sample material is present, allowing the frame to record the maximum signal the sensor can measure for each pixel and wavelength under the given conditions (e.g., exposure time, light intensity, and slide properties). Conversely, the DR is captured by completely blocking light transmission to the HS camera. This frame captures the minimum signal levels detectable by the sensor for each pixel and band, as well as dark current information from the CCD. Ideally, DR values approach zero; however, higher values may occur due to intrinsic sensor noise. To enhance the reliability of the calibration process, 100 frames are collected for both the WR and DR, ensuring that averaging reduces potential errors. Finally, the calibration of the HS image is achieved using Eq. (1), which relates the calibrated HS image (r_i) to the raw HS image (Raw_i).

	Attribute	Definition	Format
Demographic Data	**Sex**	Patients' gender	1: Female
	Race	Patients' race	1: White
	Ethnicity	Patients' ethnicity	1: Hispanic
	Menopausal status	Menopausal status of the patient	0: Premenopause and 1: Postmenopause
Diagnosis	**Dx surgery**	Type of surgery	0: Mastectomy and 1: Lumpectomy
	Dx age	Difference between diagnosis and birth dates	Years
Tumor Data	**Tumor diameter**	Maximum diameter of the irregular shaped tumor	Millimeters
	Tumor histologic grade	Degree of differentiation of tumor cells, reflecting how different they are from normal breast cells	1: Grade 1, 2: Grade 2 and 3: Grade 3
	LVI	Presence of tumor cells within lymphatic or blood vessels	0: Negative and 1: Positive
	PNI	Ability of cancer cells to proliferate around peripheral nerves and, eventually, invade them	0: Negative and 1: Positive
Tumor Stage	**T (tumor)**	Tumor size assessed by pathological evaluation	1: T1, 2: T2, 3: T3, and 4: T4
	N (node)	The cancer has spread to the LNs assessed by pathological evaluation	0: N0, 1: N1, 2: N2 and 3: N3
	M (metastasis)	Status of breast cancer spreading to a different part of the body	0: M0
Tumor Molecular Markers	**ER**	Status of ER	0: Negative and 1: Positive
	PR	Status of PR	0: Negative and 1: Positive
	HER2	Status of HER2	0: Negative and 1: Positive
	KI67	Index quantifying KI67 expression to measure how fast cancer cells are dividing in a tumor	0: KI67 index $<$ 20% and 1: KI67 index $\geq$ 20%
	Molecular subtype	Classification according to IHC status of ER, PR, HER2 and Ki67	0: Luminal A, 1: Luminal B HER2-, 2: Luminal B HER2+, 3: HER2+, and 4: Triple negative
LNs Status	**LN status**	Status of the spreading of tumor cells to the SLNs and non-SLNs	0: Negative, 1: ITC, 2: Micrometastasis and 3: Macrometastasis
	LN ITC number	LNs with ITC	Natural number
	LN MICRO number	LNs with micrometastasis	Natural number
	LN MACRO number	LNs with macrometastasis	Natural number
	LN number	LNs removed during SLN biopsy and/or LN dissection	Natural number
	SLN number	LNs removed during SLN biopsy	Natural number
	SLN status	Presence (or absence) of tumor cells in the SLNs	0: Negative, 1: ITC, 2: Micrometastasis and 3: Macrometastasis
Tumor Treatment	**Tx hormonal**	Patient received (or not) hormonal treatment	0: Not received and 1: Received
	Tx CT	Patient received (or not) adjuvant CT after the surgery	0: Not received and 1: Received
	Tx trastuzumab	Patient received (or not) trastuzumab	0: Not received and 1: Received
	Tx RT	Patient received (or not) RT	0: Not received and 1: Received
Follow-up	**Metastasis type**	Status of cancer spreading from the primary tumor to other organs during the follow-up period	0: No evidence of local or distant metastases, 1: Metastasized on nearby tissues or LNs, 2: Metastasized in distant organs from primary site and 3: Both local and distant metastases are present
	DFS	Time a patient survives without any signs or symptoms of cancer after finishing primary treatment. Difference between the relapse and diagnosis dates. If the patient did not relapse, the date of last follow-up is used instead.	Months
	Vital status		0: Alive and 1: Deceased
	Death cause		0: Other causes / Still alive and 1: Cancer
	OS	Time from the date of cancer diagnosis that patients remain alive. Difference between the death and diagnosis dates. If the patient did not die, the date of last follow-up is used instead.	Months

Table 2. Description of the study variables. CT, chemotherapy; DFS, disease-free survival; Dx, diagnosis; ER, estrogen receptors; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; ITC, isolated tumor cells; KI67, proliferation index; LN, lymph node; LVI, lymphovascular invasion; MACRO, macrometastasis; MICRO, micrometastasis; OS, overall survival; PNI, perineural invasion; PR, progesterone receptors; RT, radiotherapy; SLN, sentinel lymph node; Tx, treatment.

$$r_i = \frac{Raw_i - DR}{WR - DR} \quad (1)$$

In-house software was developed to serve multiple functions in the HSI acquisition process. It displays the RGB image to facilitate sample navigation under the microscope and ensures synchronization between the HS camera and the scanning platform by aligning their frame rate and platform movement. After capturing the HS image, the software removes the extreme bands from the raw HS image (reducing the spectral range from

400–1,000 nm to 400.5–938 nm), as these bands contain significant noise, and then saves the raw HS image. The calibrated HS image is then generated, using Eq. (1), and saved on memory as five-digit 16-bit unsigned integers (uint16), where the most significant digit represents the units, and the remaining digits correspond to the decimal places of the transmittance values. Therefore, to obtain true transmittance values, the calibrated HS image must be divided by 10^4. Additionally, the software creates a synthetic RGB image, following the methodology explained by Ortega *et al.*[33], to facilitate the visualization of the spatial characteristics of the HS image.

Prior to any HS image capture, magnification is selected, in this case the 10×. WR and DR reference images are collected. Then, to acquire the HS image, the associated WSI is examined in QuPath[35] to identify an ROI within a specific class, such as IDC (blue), healthy (green) or DCIS (red) tissue. The identified ROI is searched for in the HS microscopic system using the RGB camera and marked down on the histological image using a yellow rectangle (Fig. 2c). The ROI is then captured using the HS microscope, generating the raw HS image, the calibrated HS image, and the synthetic RGB image (Fig. 2d). The RGB image of the ROI is also captured (Fig. 2e) for future analysis. All data corresponding to one of these captures are saved using an identifier with its corresponding metadata, including the patient identifier, classification, and region (e.g., HSI_VNIR_15_IDC_x10_C01; see Data records section for more details).

Data Records

The HistologyHSI-BC Recurrence Database[36] has been deposited at The Cancer Imaging Archive (TCIA) repository[37] to make it publicly available, organized into a multilayer folder arrangement. The database is divided into three main components: clinical and demographic data, histological WSI and HS images (see Fig. 3a). The clinical and demographic data are stored at the *00_01_Clinical_Demographic_Data* file. This XLSX file documents patients' demographic status, breast tumor characteristics, treatment received, and their follow-up outcomes (detailed description on Table 2).

Furthermore, the histological data are structured in 3 folders. Firstly, *01_01_Histological_Images* folder contains the WSI for each patient, stored as MRXS files. Each WSI requires a corresponding metadata folder containing DAT and INI files for proper rendering. Moreover, *01_02_Tissue_Annotations* folder includes WSI histological annotations that classify tissue types, with boundaries of the regions outlined in blue (IDC), green (healthy), and red (DCIS), as shown in Fig. 2a. The third folder, *01_03_HSI_ROI_Annotations* contains the ROI for each HS image, with the boundaries of the region outlined in yellow (Fig. 2a). Both histological WSI and HS image ROI annotation files are provided in GeoJSON format. A summary of the areas of annotations per patient and tissue type is shown in Table 3. Lastly, *02_01_HSI_Images* folder contains the HS images of the histological slides, stored in ENVI format[38]. Each capture includes the raw HS image, WR and DR calibration files, and the calibrated HS image following the procedure described in Eq. (1). As the ENVI standard states, the HS image is saved as a flat-binary raster DAT (data) file with an accompanying HDR (header) file containing essential metadata to interpret it. Moreover, within each capture folder a synthetic RGB image and a view of the ROI captured by the RGB camera are stored. The HS image data are stored in folders named according to the regular expression *HSI_VNIR_{P}_{TT}_x10_C{CN}*, where {P} represents the patient ID, {TT} indicates the tissue type (IDC, healthy, or DCIS), and {CN} is the capture number (Fig. 3b).

Technical Validation

Clinical and demographic data statistic analysis. A preliminary statistical analysis was conducted to identify differences in the variables between patients with and without recurrence, as shown in Table 4. Statistical tests used for comparisons included the absolute frequency (percentage) for the Chi-square test or Fisher's exact test, and the median (interquartile range) for the Mann-Whitney U test. As expected, certain classic clinical and pathological variables were found to be associated with the risk of developing metastasis in the present cohort[8], including age at diagnosis, tumor diameter, and LVI.

Analysis of lymph nodes status revealed a significantly higher percentage of micrometastasis and macrometastasis in patients with recurrence compared to those without recurrence. This association remained significant when considering the number of affected lymph nodes. A similar trend was observed in sentinel lymph nodes status, where micrometastasis and macrometastasis were more prevalent in the recurrence group. However, no significant differences were found in lymph nodes containing isolated tumor cells, classified as negative lymph nodes[39]. These findings align with established knowledge that lymph node metastasis is associated with a higher risk of recurrence in BC patients during follow-up[5,40].

Regarding patient follow-up, we confirmed that all patients without recurrence show no evidence of local or distant metastases, whereas patients with recurrence do, with most of them having metastases only in distant organs and a smaller percentage presenting with both local and distant metastases. Among patients with recurrence, 90.9% died, with cancer being the cause of death in 77.3% of cases. In contrast, among the patients without recurrence who died, none died from cancer. As expected, the median disease-free survival (DFS) was significantly shorter in the recurrence group compared to the non-recurrence group, as was overall survival (OS).

Histopathology WSIs and annotation validation. Pathologists qualitatively verified the quality of histopathological slides after the sectioning, processing, and staining phases. They confirmed the absence of artifacts in the ROIs of the WSIs, ensuring that these were not introduced during the tissue preparation or digitization phases. Pathologists did not detect the presence of folds, broken tissues, tears, bubbles, scalpel marks, or bad staining on the ROIs due to the tissue preparation phase. Furthermore, they verified the quality of the digitized histopathological slides making sure there were no issues on the WSIs due to the scanning phase. They confirmed the absence of scanning artifacts like focus issues or white reference problems. The annotations on the WSIs (IDC,

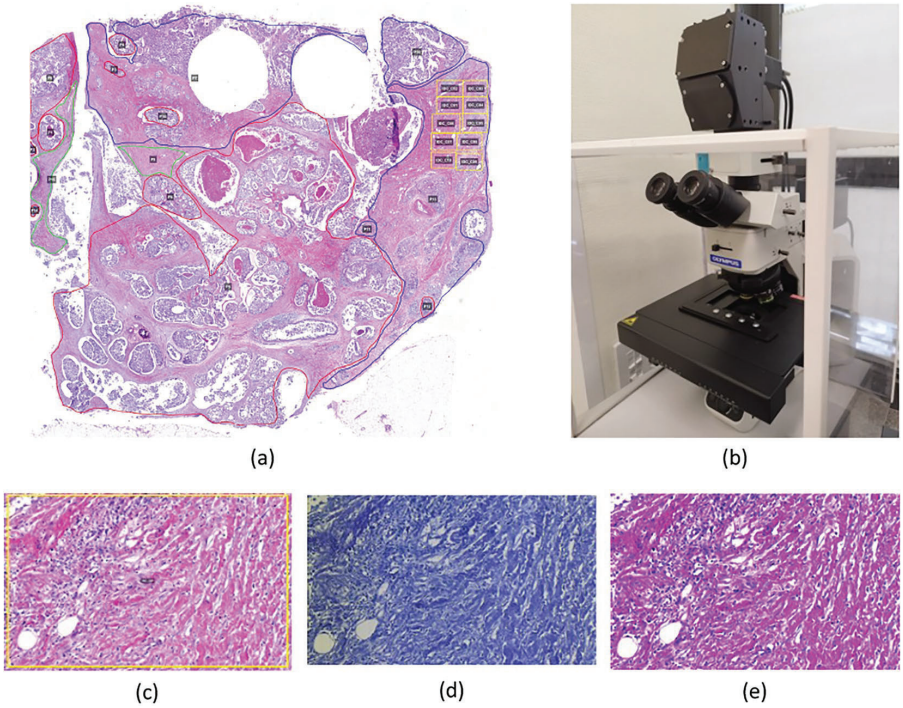


Fig. 2 Elements and outputs to capture an HS image. (**a**) Annotated WSI (IDC outlined in blue, healthy tissue in green, and DCIS in red) captured with the WSI scanner. (**b**) HS microscopic system used to acquire an HS image and its corresponding high spatial resolution RGB image of a selected ROI. (**c**) Zoom-in of one of the selected ROIs outlined by yellow rectangles in a). (**d**) Synthetic RGB image generated from the captured HS image. (**e**) High spatial resolution RGB image captured using the 20 MP RGB camera of the HS microscopic system.

healthy, and DCIS) were initially made by one pathologist, with the annotation process then subjected to validation by a second pathologist through a pairwise review. This validation phase played a crucial role in minimizing the inter-observability issue, ensuring consistency and reliability in the annotations. An example of the validated annotations is shown in Fig. 4.

HSI validation. A technical validation was performed to ensure the quality of the HistologyHSI-BC Recurrence Database. The HS microscope employed in this study has been thoroughly characterized in previous works[41], confirming its strong performance for spectral resolution-intensive applications. The system demonstrates a dynamic range of 65.3 ± 0.1 dB in transmittance mode, with a constant dark current of 20 digital numbers, which contributes to a reliable HS image capture. It is capable of capturing 826 spectral bands, providing detailed spectral information and accurately reflecting the spectral properties of the materials under analysis. This is evidenced by a spectral correlation measure of 0.88 ± 0.01 when capturing the WCT-2065 transmittance wavelength calibration standard (Avian Technologies, New London, USA) with a known spectral signature in transmittance mode. The system offers a spatial resolution of 0.739 ± 0.001 µm/pixel, along with a modulation transfer function (MTF) of 370 ± 10 line pairs/mm, ensuring sufficient detail for microscopic imaging. Spatial scanning accuracy is indicated by an eccentricity of 0.04 ± 0.04, and spatial repeatability is shown to have a relative difference of $14 \pm 8\%$ across consecutive captures. All values were measured at $10\times$ magnification, the same magnification used for the HS image capturing process in this work.

The characterization parameters obtained from the HS microscopic system demonstrate its capability to provide reliable and accurate HS data. The HS images captured from the 47 patients studied underwent a calibration. Afterwards, the database was evaluated to ensure the quality of the captured data. All HS images from each patient and tissue class (IDC, healthy and DCIS) were averaged for visualization purposes. Figure 5 groups the spectral signatures of patients with and without recurrence after 12 years. Interestingly, in the biopsies, patients without recurrence showed a greater similarity between healthy and DCIS tissues, while these tissue types were more distinctly separated in patients with recurrence. This finding raises the possibility that the closer resemblance of DCIS to healthy tissue could serve as an indicator of non-recurrence.

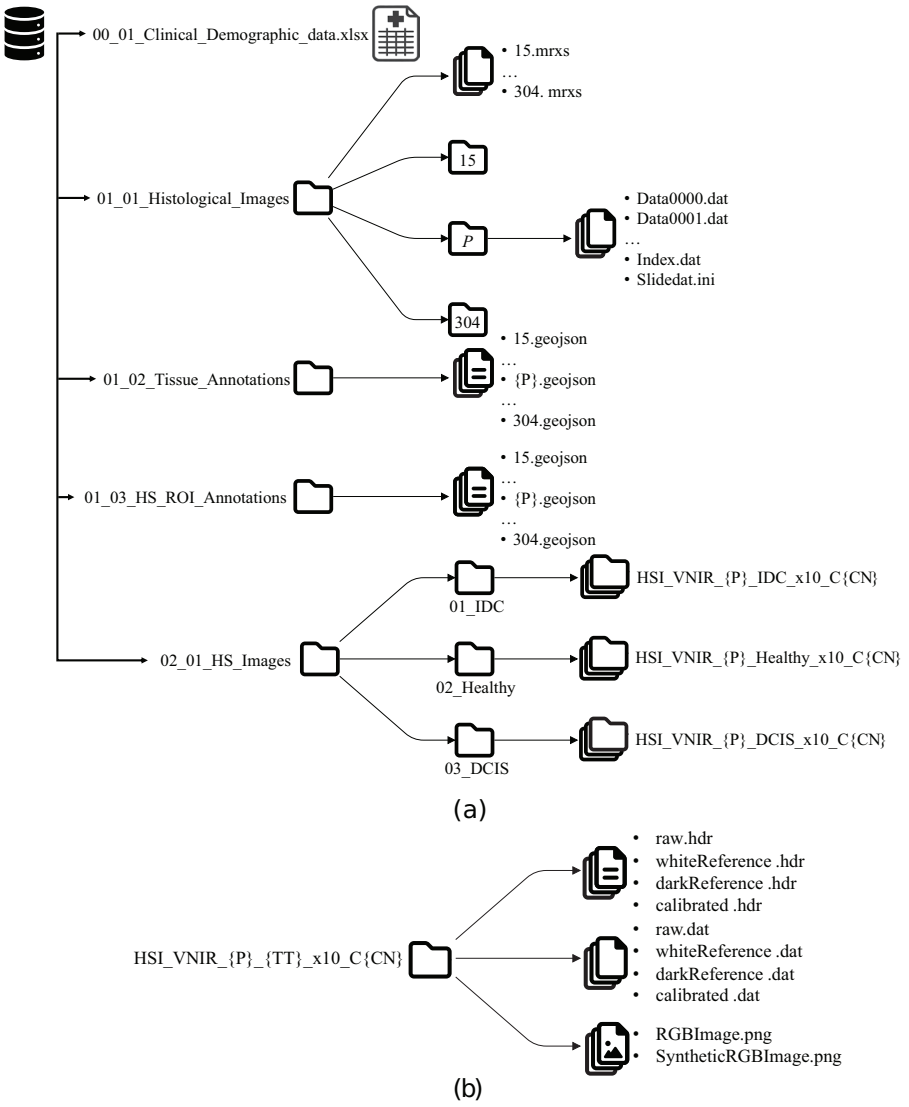


Fig. 3 Graphical representation of (**a**) the HistologyHSI-BC Recurrence Database structure and (**b**) the contents of each HS image capture.

Usage Notes

Visualizing histopathology WSIs. The authors recommend downloading and installing the QuPath software[35] to visualize the WSIs (MRXS format) and their related annotations (GeoJSON format) (Fig. 2a). Further image analysis can be performed using Python scripts (see sections Recommended histopathology WSI processing and Code availability). There are two ways to open a WSI in QuPath: drag and drop the MRXS file into QuPath or go on "File/Open" and select and open the MRXS file. There is also a tab on the left side of QuPath's user interface called "Image", in which it is possible to visualize the metadata of the histopathological image, such as width, height, magnification, and resolution. After opening the WSI on QuPath, the two available GeoJSON files containing annotations on the WSI should be imported. One includes the annotations related to the tissue compartments (IDC in red, healthy in green, and DCIS in blue). In contrast, the other defines the ROIs used for capturing the HS images, represented as yellow rectangles. These two files can be opened by dragging and dropping them into QuPath or clicking "File/Import objects from file" and selecting the GeoJSON files. The data from the GeoJSON files is visible by clicking on the tab "Annotations". If the annotation classes are not shown after clicking the "Annotations" tab, click on the button with the three vertical dots on the bottom right of the tab panel,

Patient ID	Tissue Annotations Area [mm²]			HS image ROI Annotations [Number of captures]			Recurrence Label
	IDC	Healthy	DCIS	IDC	Healthy	DCIS	
15	40.94	2.92	64.24	10	0	0	0
19	94.89	24.00	0.00	10	5	0	1
20	169.27	58.36	0.28	10	5	0	1
25	135.97	139.29	17.62	10	5	5	0
38	210.59	7.71	0.23	10	5	0	1
40	72.19	20.77	0.00	10	5	0	1
43	57.46	22.70	0.00	10	5	0	0
45	71.77	80.06	6.70	10	5	5	0
47	10.40	137.41	0.00	10	5	0	0
51	163.12	2.79	2.34	10	5	0	0
52	96.22	3.50	0.00	10	3	0	0
57	23.26	31.54	0.00	10	5	0	0
62	4.40	66.22	0.00	8	5	0	0
65	104.57	15.22	0.00	10	5	0	0
68	21.41	75.70	0.00	10	5	0	0
70	55.12	45.12	0.00	10	5	0	0
80	4.84	37.06	0.00	7	5	0	0
82	109.32	11.99	0.31	10	5	0	0
84	88.06	13.20	7.50	10	5	0	1
85	119.25	5.21	20.59	10	5	5	0
90	179.57	29.70	0.00	10	5	0	0
100	68.16	3.21	0.06	9	4	0	1
107	8.43	192.86	7.70	10	5	5	0
112	14.84	1.68	0.00	10	0	0	0
124	26.84	0.36	0.00	10	0	0	0
136	190.00	25.98	28.52	9	4	5	0
138	32.58	90.49	11.91	10	5	0	0
139	43.23	72.05	0.00	10	5	0	0
141	168.53	7.96	5.27	10	5	0	1
146	24.16	3.47	0.32	10	0	0	0
151	3.98	3.78	1.10	4	0	0	0
152	82.41	42.77	19.22	10	5	5	1
153	7.06	77.57	0.88	8	5	0	0
154	51.74	0.00	0.00	10	0	0	1
162	67.66	16.19	1.72	10	5	0	1
189	247.15	0.02	0.00	10	0	0	1
197	212.83	57.60	4.90	10	5	5	1
205	321.29	35.69	0.14	10	5	0	1
211	161.00	0.10	0.09	10	0	0	1
213	451.91	7.64	0.00	10	5	0	1
229	281.13	3.09	2.02	10	3	0	1
238	149.90	0.00	0.00	9	0	0	1
255	164.85	35.81	0.50	10	5	0	1
259	92.43	42.47	2.35	10	5	0	1
269	59.97	6.65	0.00	10	5	0	1
270	30.11	111.68	0.00	10	5	0	1
304	155.02	24.54	0.00	10	5	0	1

Table 3. Summary of histological area coverage and capture counts by tissue type per image.

then select "Populate from existing objects/All classes (including sub-classes)" and the class types along with the number of annotations for each will appear.

Recommended histopathology WSI processing. This section provides guidelines for working with WSIs (MRXS format), for which the use of Python scripts is recommended (see the Code availability section for more details). Due to their high resolution, efficient processing techniques are necessary to optimize performance and memory usage. Processing high-resolution images can be time-consuming and memory intensive. The

	Attribute	Format	Recurrence	Non-Recurrence	*p*
Demographic Data	**Sex**	1: Female	22 (100.0)	25 (100.0)	—
	Race	1: White	22 (100.0)	25 (100.0)	—
	Ethnicity	1: Hispanic	22 (100.0)	25 (100.0)	—
	Menopausal status	0: Premenopause 1: Postmenopause	3 (13.6) 19 (86.4)	3 (12.0) 22 (88.0)	1.000*
Diagnosis	**Dx surgery**	0: Mastectomy 1: Lumpectomy	7 (31.8) 15 (68.2)	2 (8.0) 23 (92.0)	0.063*
	Dx age	Years	73.0 [22.0]	57.0 [14.0]	**0.017‡**
Tumor Data	**Tumor diameter**	Millimeters	26.5 [13.8]	15.0 [13.0]	**<0.001‡**
	Tumor histologic grade	1: Grade 1 2: Grade 2 3: Grade 3	1 (4.5) 11 (50.0) 10 (45.5)	4 (16.0) 14 (56.0) 7 (28.0)	0.285*
	LVI	0: Negative 1: Positive	8 (36.4) 14 (63.6)	18 (72.0) 7 (28.0)	**0.031***
	PNI	0: Negative 1: Positive	15 (68.2) 7 (31.8)	21 (84.0) 4 (16.0)	0.351*
Tumor Stage	**T (tumor)**	1: T1 2: T2 3: T3 4: T4	5 (22.7) 14 (63.6) 2 (9.1) 1 (4.5)	17 (68.0) 8 (32.0) 0 (0.0) 0 (0.0)	**0.012***
	N (node)	0: N0 1: N1 2: N2 3: N3	8 (36.4) 7 (31.8) 5 (22.7) 2 (9.1)	25 (100.0) 0 (0.0) 0 (0.0) 0 (0.0)	**<0.001***
	M (metastasis)	0: M0	22 (100.0)	25 (100.0)	—
Tumor Molecular Markers	**ER**	0: Negative 1: Positive	4 (18.2) 18 (81.8)	5 (20.0) 20 (80.0)	1.000*
	PR	0: Negative 1: Positive	6 (27.3) 16 (72.7)	9 (36.0) 16 (64.0)	0.744*
	HER2	0: Negative 1: Positive	15 (68.2) 7 (31.8)	22 (88.0) 3 (12.0)	0.154*
	KI67	0: KI67 index < 20% 1: KI67 index ≥ 20%	4 (18.2) 18 (81.8)	10 (40.0) 15 (60.0)	0.189*
	Molecular subtype	0: Luminal A 1: Luminal B HER2− 2: Luminal B HER2+ 3: HER2+ 4: Triple negative	4 (18.2) 10 (45.5) 4 (18.2) 3 (13.6) 1 (4.5)	6 (24.0) 13 (52.0) 2 (8.0) 1 (4.0) 3 (12.0)	0.512*
LNs Status	**LN status**	0: Negative 1: ITC 2: Micrometastasis 3: Macrometastasis	6 (27.3) 2 (9.1) 3 (13.6) 11 (50.0)	22 (88.0) 3 (12.0) 0 (0.0) 0 (0.0)	**<0.001***
	LN ITC number	Number of LNs with ITC	0.0 [0.0]	0.0 [0.0]	0.720‡
	LN MICRO number	Number of LNs with micrometastasis	0.0 [0.0]	0.0 [0.0]	**0.027‡**
	LN MACRO number	Number of LNs with macrometastasis	0.5 [7.0]	0.0 [0.0]	**<0.001‡**
	LN number	Total number of LNs removed during SLN biopsy and/or LN dissection	13.0 [15.0]	2.0 [2.0]	**0.001‡**
	SLN number	Number of LNs removed during SLN biopsy	0.5 [2.0]	2.0 [2.0]	**<0.001‡**
	SLN status	0: Negative 1: ITC 2: Micrometastasis 3: Macrometastasis	4 (36.4) 2 (18.2) 2 (18.2) 3 (27.3)	22 (88.0) 3 (12.0) 0 (0.0) 0 (0.0)	**0.002***
Tumor Treatment	**Tx hormonal**	0: Not received 1: Received	5 (22.7) 17 (77.3)	4 (16.0) 21 (84.0)	0.715*
	Tx CT	0: Not received 1: Received	11 (50.0) 11 (50.0)	14 (56.0) 11 (44.0)	0.906*
	Tx trastuzumab	0: Not received 1: Received	19 (86.4) 3 (13.6)	23 (92.0) 2 (8.0)	0.654*
	Tx RT	0: Not received 1: Received	3 (13.6) 19 (86.4)	3 (12.0) 22 (88.0)	1.000*
Follow-up	**Metastasis type**	0: No evidence of local or distant metastases 1: Metastasized on nearby tissues or LNs 2: Metastasized in distant organs from primary site 3: Both local and distant metastases are present	0 (0.0) 0 (0.0) 20 (90.9) 2 (9.1)	25 (100.0) 0 (0.0) 0 (0.0) 0 (0.0)	**<0.001***
	DFS	Months	39.0 [48.0]	150.0 [28.0]	**<0.001‡**
	Vital status	0: Alive 1: Deceased	2 (9.1) 20 (90.9)	21 (84.0) 4 (16.0)	**<0.001***
	Death cause	0: Other causes / Still alive 1: Cancer	5 (22.7) 17 (77.3)	25 (100.0) 0 (0.0)	**<0.001***
	OS	Months	66.5 [85.0]	150.0 [28.0]	**<0.001‡**

Table 4. Differences in the clinical and demographic variables in recurrence vs. non-recurrence groups. Data are expressed as absolute frequency (percentage) for qualitative variables, compared using the Chi-square test or Fisher's exact text*, and as median [interquartile range] for quantitative variables, analyzed using the Mann-Whitney U test‡. CT, chemotherapy; DFS, disease-free survival; Dx, diagnosis; ER, estrogen receptors; HER2, human epidermal growth factor receptor 2; ITC, isolated tumor cells; KI67, proliferation index; LN, lymph node; LVI, lymphovascular invasion; MACRO, macrometastasis; MICRO, micrometastasis; OS, overall survival; PNI, perineural invasion; PR, progesterone receptors; RT, radiotherapy; SLN, sentinel lymph node; Tx, treatment.

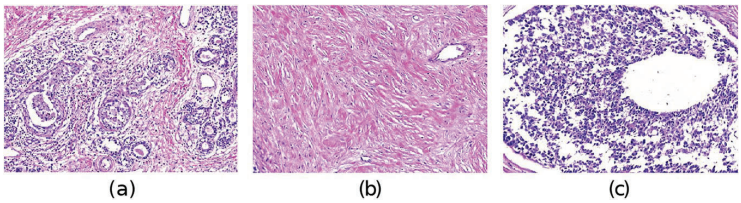


Fig. 4 Examples of (**a**) IDC, (**b**) healthy, and (**c**) DCIS tissue types on a WSI at 2x magnification.

highest available resolution of the selected slide image is approximately 85,000 × 202,000 pixels, making it significantly large. To optimize performance, a lower resolution (approximately 670 × 1,600) level should be selected for visualization. It is also important to downscale the annotations to match the selected lower resolution level.

Recommended HSI processing. After HS data capture, the calibration of HS images is a mandatory step; however, additional processing may be performed depending on the specific application of the data.

- Given the strong correlation between adjacent spectral bands, spectral dimensionality reduction can be beneficial in reducing intrinsic Gaussian noise and computational costs. This can be accomplished by averaging adjacent spectral bands to create a spectrally reduced HS image. For example, the data could be reduced from the original 826 bands to 275 using a spectral window that includes three neighboring bands.
- Normalization is also recommended when partial absorbance is less critical, but the specific absorption wavelengths are significant. This normalization can be performed to scale the data between 0 and 1 or to have a mean of 0 and a standard deviation of 1.
- For HS analysis of the samples, it is advised to remove the sample background by identifying areas with no absorbance, typically represented by the white background.
- The classification of the data can be based on recurrence status and/or tissue type (IDC, healthy, or DCIS). When using ML or DL, it is crucial to ensure that data from the same patient do not appear simultaneously in the training, testing, or validation sets.

All these processing steps can be implemented using Python (see the Code availability section for further details).

How to combine the different databases (clinical and demographic, WSIs, and HSI). Integrating the diverse databases within the Histology HSI-BC Recurrence Database can enhance the prediction of distant recurrence in BC by leveraging complementary information from multiple modalities. Histopathological WSIs provide morphological insights assessed by pathologists, HS images capture biochemical variations that may indicate early tumor progression, and clinical and demographic data offer critical patient-specific factors. By integrating these databases, researchers can develop more robust predictive models that go beyond traditional histological or clinical assessments, improving risk stratification and supporting personalized treatment decisions.

Researchers can employ various data fusion strategies to achieve this integration. *Early fusion* involves combining raw or preprocessed features from each modality before model training, allowing the model to learn directly from the integrated data[42]. *Intermediate fusion* entails extracting high-level features from each database separately and then merging them into a joint representation, capturing modality-specific patterns prior to integration[43]. *Late fusion* consists of training independent models for each modality and subsequently combining their outputs to improve overall prediction accuracy[19]. Implementing these fusion techniques requires meticulous preprocessing to ensure compatibility and maximize the value of each database. By effectively integrating these multimodal databases, researchers can uncover subtle patterns associated with BC recurrence, advancing precision oncology and personalized patient care.

Data availability

The HistologyHSI-BC Recurrence Database is publicly available on the TCIA repository at https://doi.org/10.7937/6KPY-YT49.

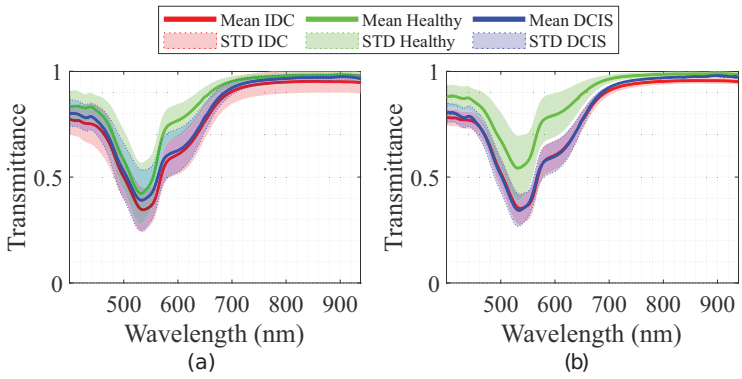


Fig. 5 Mean and standard deviation HS spectral signatures for different tissues (IDC, healthy and DCIS) for patients (**a**) without and (**b**) with recurrence.

Code availability

The Python scripts developed for this study, along with the required Python toolboxes and libraries, are available in the publicly accessible repository at https://github.com/HIRIS-Lab/HistologyHSI-BC-Recurrence. The repository includes a main script providing a basic example of how to load and perform preliminary preprocessing of hyperspectral data in ENVI format using Python, for which the spectral library is required. In addition, the repository contains a tutorial demonstrating how to manipulate WSI in MRXS format and annotation in GeoJSON format in Python. The script overlays tissue compartments (IDC, healthy, and DCIS) onto slide images using their corresponding colors: blue, green, and red. Working with WSIs in Python requires the installation of the OpenSlide library. In addition, the JSON library is used for loading annotations, and the PIL library is employed to prepare and save slide images.

Received: 5 June 2025; Accepted: 21 October 2025;
Published online: 28 November 2025

References

1. Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**, 229–263 (2024).
2. Chiang, A. C. & Massagué, J. Molecular Basis of Metastasis. *N Engl J Med* **359**, 2814 (2008).
3. Colzani, E. *et al.* Time-dependent risk of developing distant metastasis in breast cancer patients according to treatment, age and tumour characteristics. *Br J Cancer* **110**, 1378–1384 (2014).
4. Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J Clin* **74**, 12–49 (2024).
5. Riggio, A. I., Varley, K. E. & Welm, A. L. The lingering mysteries of metastatic recurrence in breast cancer. *Br J Cancer* **124**, 13–26 (2021).
6. Caswell-Jin, J. L. *et al.* Change in Survival in Metastatic Breast Cancer with Treatment Advances: Meta-Analysis and Systematic Review. *JNCI Cancer Spectr* **2** (2018).
7. Iwata, H. Future treatment strategies for metastatic breast cancer: curable or incurable? *Breast Cancer* **19**, 200–205 (2012).
8. Harbeck, N. *et al.* Breast cancer. *Nature Reviews Disease Primers* **5**, 1–31 (2019). *2019 5:1*.
9. Fasching, P. A. *et al.* Biomarkers in Patients with Metastatic Breast Cancer and the PRAEGNANT Study Network. *Geburtshilfe Frauenheilkd* **75**, 41 (2015).
10. Lee, K. K. *et al.* Prognostic Biomarkers for Breast Cancer Metastasis. *Cancer Metastasis*, https://doi.org/10.5772/INTECHOPEN.80576 (2018).
11. Ross, J. S. *et al.* Immunotherapy predictive biomarkers in metastatic breast cancer (MBC). *Journal of Clinical Oncology* **37**, 1023–1023 (2019).
12. Prasad, S. N. & Houserkova, D. The Role Of Various Modalities In Breast Imaging. *Biomedical Papers* **151**, 209–218 (2007).
13. *Breast Tumours. World Health Organization Classification of tumours* (IACR, Lyon, France, 2019).
14. Jain, E. *et al.* Whole Slide Imaging Technology and Its Applications: Current and Emerging Perspectives. *Int J Surg Pathol* **32**, 433–448 (2024).
15. Pantanowitz, L. *et al.* Twenty Years of Digital Pathology: An Overview of the Road Travelled, What is on the Horizon, and the Emergence of Vendor-Neutral Archives. *J Pathol Inform* **9**, 40 (2018).
16. Mezei, T., Kolcsár, M., Joó, A. & Gurzu, S. Image Analysis in Histopathology and Cytopathology: From Early Days to Current. *Perspectives. J Imaging* **10**, 252 (2024).
17. Soliman, A., Li, Z. & Parwani, A. V. Artificial intelligence's impact on breast cancer pathology: a literature review. *Diagn Pathol* **19**, 38 (2024).
18. Zhao, Y. *et al.* A review of cancer data fusion methods based on deep learning. *Information Fusion* **108**, 102361 (2024).
19. Llinas-Bertran, A., Butjosa-Espín, M., Barberi, V. & Seoane, J. A. Multimodal data integration in early-stage breast cancer. *The Breast* **80**, 103892 (2025).
20. Lipkova, J. *et al.* Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* **40**, 1095–1110 (2022).
21. Chang, C.-I. *Hyperspectral Imaging*, https://doi.org/10.1007/978-1-4419-9170-6 (Springer US, Boston, MA, 2003).
22. Fei, B. Hyperspectral imaging in medical applications. in 523–565, https://doi.org/10.1016/B978-0-444-63977-6.00021-3. (2019).
23. Karim, S., Qadir, A., Farooq, U., Shakir, M. & Laghari, A. A. Hyperspectral Imaging: A Review and Trends towards Medical Imaging. *Curr Med Imaging Rev* **19**, 417–427 (2022).
24. Lu, G. & Fei, B. Medical hyperspectral imaging: a review. *J Biomed Opt* **19**, 010901 (2014).

25. Saccomandi, P. *et al*. Estimation of anisotropy coefficient of swine pancreas, liver and muscle at 1064 nm based on goniometric technique. *J Biophotonics* **8**, 422–428 (2015).
26. Wang, Q., Wang, J., Zhou, M., Li, Q. & Wang, Y. Spectral-spatial feature-based neural network method for acute lymphoblastic leukemia cell identification via microscopic hyperspectral imaging technology. *Biomed Opt Express* **8**, 3017 (2017).
27. Enfield, K. S. S. *et al*. Hyperspectral cell sociology reveals spatial tumor-immune cell interactions associated with lung cancer recurrence. *J Immunother Cancer* **7**, 13 (2019).
28. Lingle, W. *et al*. The Cancer Genome Atlas Breast Invasive Carcinoma Collection (TCGA-BRCA) (Version 3) [Data set]. *The Cancer Imaging Archive*, https://doi.org/10.7937/K9/TCIA.2016.AB2NAZRP (2016).
29. Curtis, C. *et al*. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
30. Fiorin, A., López Pablo, C., Lejeune, M., Hamza Siraj, A. & Della Mea, V. Enhancing AI Research for Breast Cancer: A Comprehensive Review of Tumor-Infiltrating Lymphocyte Datasets. *Journal of Imaging Informatics in Medicine* **37**, 2996–3008 (2024).
31. Amgad, M. *et al*. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics* **35**, 3461–3467 (2019).
32. Zhang, Y., Wang, Y., Zhang, B. & Li, Q. A hyperspectral dataset of precancerous lesions in gastric cancer and benchmarks for pathological diagnosis. *J Biophotonics* **15** (2022).
33. Ortega, S. *et al*. Publisher Correction: Histological Hyperspectral Glioblastoma Dataset (HistologyHSI-GB). *Sci Data* **11**, 829 (2024).
34. *WHO Classification of Tumours of the Breast*. (IARC, Lyon, France, 2012).
35. Bankhead, P. *et al*. QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).
36. Quintana-Quintana, L. *et al*. Recurrent Breast Cancer: Histopathological and Hyperspectral Images Database (HistologyHSI-BC-Recurrence) (Version 1) [Data set]. *The Cancer Imaging Archive*, https://doi.org/10.7937/6KPY-YT49 (2025).
37. Clark, K. *et al*. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J Digit Imaging* **26**, 1045–1057 (2013).
38. ENVI Header Files. https://www.nv5geospatialsoftware.com/docs/ENVIHeaderFiles.html#The.
39. Naidoo, K. & Pinder, S. E. Micro- and macro-metastasis in the axillary lymph node: A review. *The Surgeon* **15**, 76–82 (2017).
40. Jatoi, I., Hilsenbeck, S. G., Clark, G. M. & Osborne, C. K. Significance of Axillary Lymph Node Metastasis in Primary Breast Cancer. *Journal of Clinical Oncology* **17**, 2334–2334 (1999).
41. Quintana-Quintana, L. *et al*. Roadmap for the Characterization and Validation of Hyperspectral Microscopic Systems [submitted]. *IEEE Trans Instrum Meas* 1–16, https://doi.org/10.1109/TIM.2024.3379090 (2025).
42. Wang, Z. *et al*. Deep learning-based multi-modal data integration enhancing breast cancer disease-free survival prediction. *Precis Clin Med* **7** (2024).
43. Nakach, F.-Z., Idri, A. & Goceri, E. A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification. *Artif Intell Rev* **57**, 327 (2024).

Acknowledgements

This work has been supported by the grant Dr. Ferran 2021 (FFPI21/BE01). Moreover, this work was completed while L.Q-Q. was beneficiary of the pre-doctoral grant given by the "Agencia Canaria de Investigación, Innovación y Sociedad de la Información (ACIISI)" of the "Consejería de Economía, Conocimiento y Empleo", which is part-financed by the European Social Fund (FSE) (POC 2014–2020, Eje 3 Tema Prioritario 74 (85%)). E.S-C. was beneficiary of a grant from the Generalitat de Catalunya (Doctorats Industrials, Pla DI, AGAUR, Grant No. 2022DI057). J.S-N. was supported by the 2024 predoctoral grant from Las Palmas provincial headquarters of the Scientific Foundation of the Spanish Association Against Cancer (PRDLP246561SANT). A.F., a member of the BosomShield project, received funding from Marie Sklodowska-Curie Doctoral Networks Actions (HORIZON-MSCA-2021-DN-01-01) under grant agreement 101073222. C.L-P. was the PI of the SCARLET, a project funded by Proyectos Estrategicos Orientados a la Transición Ecológica y a la Transición Digital, from the 2021 call of the Ministerio de Ciencia e Innovación, with grant number TED2021-130081B-C22 and funding from NextGenerationEU. Moreover, this work is part of the OASIS (Open Ai-driven Stack para plataformas hpec mejoradas en Sistemas Integrado) project (PID2023-148285OB-C43), supported by the Spanish Government and European Union (FEDER Funds). The samples used in this work were obtained from the Biobank IISPV-Node Tortosa which is part of the Xarxa de Bancs de Tumors de Catalunya (XBTC), financed by the Oncology Master Plan for Catalonia (Pla Director d'Oncologia de Catalunya). We would like to thank the Legal Unit and the Bioinformatics Unit of the IISPV for their support, as well as the collaboration of the Spanish Foundation for Science and Technology, F.S.P.-Ministry of Science and Innovation, within the framework of project MDG-23-11426.

Author contributions

L.Q.-Q. conceived the study, designed and established the experimental system, carried out a research stay at the Hospital de Tortosa Verge de la Cinta, generated the hyperspectral database, performed image processing and result analysis, and wrote the manuscript. E.S.-C. conceived the study, acquired the funding, collected and cleaned the clinical database, generated the histology database, conducted the technical validation experiments, and wrote the manuscript. A.F. generated the histology database and wrote the manuscript. J.S.-N. generated the hyperspectral database and wrote the manuscript. S.O. supervised the study and reviewed the manuscript. N.G.-B. collected and cleaned the clinical database and reviewed the manuscript. A.F.-C. conducted the technical validation experiments and reviewed the manuscript. T.S.-A. collected and cleaned the clinical database and reviewed the manuscript. H.F. supervised the study and reviewed the manuscript. L.A.-L., D.M.-C., and R.B.-P. prepared and diagnosed the histological samples, annotated the histopathology images, and reviewed the manuscript. M.L. supervised the study and reviewed the manuscript. G.M.C. and C.L.-P. acquired the funding, supervised the study, and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.S.-C.

Chapter 6. Conclusions & Future Lines

HS microscopy holds great promise for digital pathology, offering label-free, reproducible, and information-rich imaging with the potential to improve tissue characterization and enable the integration of additional information into ML/DL workflows. However, its clinical implementation is still hindered by several challenges that must be addressed to ensure robust, repeatable, and diagnostically reliable outcomes. This Ph.D. thesis aimed to address key challenges in the deployment of HSI at the microscopic scale for medical histology by tackling four major aspects: (1) the design and characterization of HSI microscopic systems, (2) the development and validation of image quality and focus assessment methods, (3) the optimization of sample preparation, specifically tissue thickness, to improve spectral contrast of HSI data and (4) the generation of standard, public available HS microscopic databases to support reproducible algorithm development and accelerate clinical translation of HS microscopy.

This final chapter presents a synthesis of the key findings obtained throughout the research and offers a critical reflection on their implications. The chapter is structured according to the three main thematic areas explored, followed by a section discussing the broader contributions of the Ph.D. thesis, its limitations, and suggested future directions.

6.1 Summary of Main Findings

Characterization and Design of HSI Microscopic Systems

The objective O1 of this Ph.D. thesis was to establish a standardized methodology for the characterization and performance

evaluation of HSI microscopic systems, which remains a critical yet underexplored area in the literature. To this end, Chapter 2 presented a comprehensive and sequential roadmap for system design and calibration, rooted in fundamental imaging science and adapted to the constraints of HS microscopy.

Key aspects addressed include assessing dynamic range, exposure time optimization, dark current evaluation, spatial resolution characterization (including Nyquist frequency analysis), scanning alignment, spatial repeatability, flat-field correction, tone transfer, and spectral sensitivity. Special attention was given to parameters often overlooked in biomedical imaging studies, such as spatial repeatability under low SNR conditions and the influence of wavelength-dependent spatial frequency response.

One of the novel contributions in Chapter 2 was the emphasis on wavelength-dependent resolution, informed by Rayleigh's criterion and verified empirically. This highlighted that shorter wavelengths provide inherently better spatial resolution, a fact that must be considered in the design and usage of HSI systems for resolving fine structures like cellular components. This finding laid the groundwork for Chapter 3, where focus and resolution variability across wavelengths drive the choice of blur metrics. Additionally, Chapter 2 presented practical guidelines for alignment verification using circular targets and eccentricity measures, helping to prevent distortions caused by misalignment in spatial scanning systems. These guidelines were not only instrumental during the initial system characterization but also became essential tools in ensuring data integrity in subsequent experiments (Chapters 3–5).

A round-robin validation study was performed involving HSI microscopic systems of two different institutions (UPM and ULPGC), thereby testing the robustness and generalizability of the proposed roadmap. The resulting evidence demonstrates the reproducibility of the methodology and supports its integration into a broader standardization effort. This is particularly relevant considering the current lack of a formal protocol for validating and comparing HSI microscopic devices. Following the example set by the 'Standard for Characterization and Calibration of Ultraviolet through Shortwave Infrared (250 nm to 2500 nm) Hyperspectral Imaging Devices' [102], this work may serve as a foundational

reference for the future development of a dedicated standard specifically tailored to HS microscopy systems.

In summary, the work presented in Chapter 2 provides a systematic and empirically validated reference for the biomedical HSI community, facilitating better interoperability between devices, improving result reproducibility, and supporting the integration of HSI into clinical workflows.

Focus Assessment on HS Microscopy

This Ph.D. thesis investigated another critical yet often overlooked aspect of HS microscopy: image quality, specifically regarding focus. Histological samples often exhibit surface irregularities due to the inherent softness of the tissue and the difficulty of achieving uniform thickness when sectioning at the micrometer scale. These irregularities cause the sample surface to lie at varying distances from the objective lens across the FOV. When combined with the shallow depth of field typical high-magnification optics, maintaining consistent focus across the entire image becomes especially challenging. Additionally, the thickness of the tissue can result in multiple focal planes coexisting within a single point of the sample. Additionally, as established in Chapter 2, spatial resolution is highly dependent on wavelength, meaning that the optimal focus position can also shift across the spectral range. These factors together make precise focusing a complex yet critical aspect of HS microscopy.

The study presented in Chapter 3 began with a detailed review of existing IQA algorithms, especially those that operate with no-reference (NR-IQA). These include methods, originally developed for RGB imaging, were subsequently extended to HS data. To facilitate algorithm evaluation, a benchmark dataset, HIDFA, was generated. It comprises 1,375 HS cubes acquired at 11 distinct working distances ranging from −0.5 mm to 0.5 mm in 0.1 mm increments, with the focused image corresponding to 0 mm. Each cube is paired with subjective quality ratings collected from 24 human participants using a Swiss tournament-style evaluation method. This allowed for the extraction of Mean Opinion Scores (MOS), forming a strong ground truth against which to assess algorithmic performance.

Objective methods were divided into learning-free and learning-based categories. Among the learning-free methods, transform-free techniques included edge-based (CPBD, EMBM, and PSI) and edge-free (ARISM and MLV) algorithms. Transform-based methods were further categorized into wavelet-based (FISHbb, LPC) and Fourier-like (BIBLE, S3) approaches. The only learning-based method tested was SPARISH, which represents a shallow learning, feature-based approach. These methods were applied to both synthetic (RGB-like) and monochromatic (mean of all bands) images derived from HS cubes, to assess their effectiveness in blur estimation. Results showed that MLV, FISHbb, LPC, and S3 yielded strong correlations with the MOS scores (with SRCC > 0.85 and RMSE < 1.8).

Subsequently, band-wise performance was evaluated using monochromatic images of each band, and algorithms were grouped into two main clusters: G1 (i.e., MLV, FISHbb, LPC) which yielded uniform scores across bands, and G2 (i.e., CPBD, EMBM, PSI, SPARISH) which showed spectral variability in their scores. Since G1 methods produced consistent blur scores across all spectral bands, they were initially considered more suitable for systems with a fixed working distance across the spectrum. However, further analysis of blur score behavior across four sample types (i.e., blood smears, leaves, rat histology, and stems) revealed that G1 methods more effectively maintained consistent focus scores across the spectral range. For example, while blood smears exhibit limited information in the green region and richer content in the red, G1 algorithms consistently reflected similar focus scores. This finding supports the selection of G1 algorithms for HS microscopy quality control, particularly because the entire HS cube is captured at a fixed working distance. Guidelines presented in Chapter 3 contributed to improving diagnostic reliability by ensuring that only high-quality, well-focused data are passed on to following analysis stages, either by guiding focus adjustment prior to HS image acquisition or by filtering out suboptimal HS data afterwards.

HS Microscopy Data Processing

Shifting focus from hardware and acquisition improvements to data refinement, Chapter 3 also explored how classical image processing techniques can enhance the quality and consistency of HS microscopic images, making them more robust for downstream ML

and DL classification tasks. Recognizing the lack of standardized processing pipelines in medical HS imaging, Chapter 3 proposed a generic, system-independent approach to HS data optimization. By incorporating a wavelength reflectance standard, the method enabled consistent evaluation of processing techniques across different HSI systems (i.e., VNIR colposcopic, VNIR microscopic, neurosurgical VNIR and NIR imaging systems) without being tied to a specific dataset or clinical application. The focus was on improving spectral data quality by systematically assessing and comparing noise-reduction and artifact-mitigation methods, including smoothing, normalization, and spectral derivative algorithms. A review of state-of-the-art techniques guided the selection of candidate algorithms, whose performance was quantitatively evaluated using the RMSE relative to the manufacturer-provided reference spectrum. Execution time was also measured to assess each method's feasibility for real-time or near-real-time HS applications.

Results showed that optimal processing varied across systems due to differences in spectral resolution and sampling of each HS sensor. For smoothing, GAUSSIAN, MOVMEDIAN and LOESS were the best performing algorithms. The SNV normalization method consistently yielded the best results across all systems, as it effectively corrected for scale and offset variations in the data. Combining smoothing and normalization improved spectral fidelity further, though results depended on the choice of smoothing window size, which was not optimized in this study. The evaluation of spectral derivatives revealed that first derivatives outperformed second derivatives in terms of RMSE, but both methods amplified noise and did not improve overall performance compared to smoothing and normalization alone. This underscores the importance of careful noise management, previously emphasized Chapter 2, when applying derivative-based processing in medical HSI.

Execution times for most algorithms were within 10^{-2} to 10^{-3} seconds, suggesting practical feasibility for many applications. However, more computationally intensive methods like RLOWESS and RLOESS had significantly longer runtimes (~40 s), limiting their suitability for time-sensitive environments. These findings

point to future work on hardware acceleration using GPUs or FPGAs to enable real-time processing even with slower algorithms.

Overall, Chapter 3 introduced a reproducible, system-independent framework for evaluating spectral processing in medical HSI. By shifting focus from application-specific datasets to a standardized reference, it enables consistent benchmarking across systems and offers actionable insights for developing more efficient, high-quality processing pipelines. This work lays the foundation for future integration of optimized processing into real-time HSI workflows for clinical diagnostics.

Influence of Tissue Thickness on HS Microscopy Data

Continuing in the path of standardization and quality improvement of HS microscopic systems, Chapter 4 explored the impact of tissue thickness on HS data, an often neglected or poorly reported parameter in the field. WSI relies on thin tissue slices (2–10 μm), to ensure light transmission, and enhances contrast through staining. However, HS microscopy could benefit from capturing the interaction of light with tissue intrinsic chromophores, generating unique high contrast spectral signatures. This shift necessitates rethinking conventional sample preparation protocols. A review of existing MS/HS microscopy literature revealed a pervasive lack of transparency regarding tissue thickness: 60% of the reviewed studies failed to report thickness, and the remainder mostly all of them followed conventional histological ranges (2–10 μm).

To address this gap, this Ph.D. thesis proposed a Monte Carlo (MC) light transport simulation framework designed to model the interaction of light with tissue samples of varying thicknesses, ranging from 5 μm to 2000 μm. The framework incorporated tissue-specific optical parameters, such as absorption, scattering, and refractive index, sourced from peer-reviewed literature, and covered four organs (breast, lung, liver, and colorectal) in both healthy and lesioned states. To better approximate experimental conditions and reflect the spectral variations discussed in Chapter 2 and Chapter 3, the simulated data were further enhanced through data augmentation techniques. These included modeling SNR degradation, illumination instability, and calibration errors to mimic common sources of variability in real-world HSI.

The simulated spectral signatures showed that thin samples (~5 μm) produced uniform transmission spectrum with minimal spectral contrast, confirming that the dominating contrast in standard histology arises from staining. Conversely, samples thicker than 1000 μm exhibited severe light attenuation, rendering them suboptimal for HS data capture. Samples with intermediate thicknesses (e.g., 200–500 μm) offered a trade-off between signal intensity and spectral discrimination power. To validate the findings, spectral similarity metrics (i.e., Euclidean distance, SAM, NS3, SID and, SIDSAM) were used to quantitatively compare the HS signatures of healthy and lesioned tissues. For each tissue type (i.e., breast, lung, colorectal and liver tissue), these metrics helped identify the thickness at which the contrast between healthy and lesion spectra was most pronounced. Although the spectral metrics produced largely consistent results, slight differences were observed. To resolve these discrepancies, a majority voting strategy was applied to determine the optimal thickness for each tissue. Based on this strategy, the best-performing thicknesses were: 1000 μm for breast tissue, 200 and 500 μm for lung and colorectal tissue, and 500 μm for liver tissue.

Summarizing, the results highlight that adhering strictly to traditional histological thicknesses may limit the diagnostic potential of HS microscopy. This work suggests revisiting histological protocols when HS is used, advocating for customized sample thickness selection based on the specific diagnostic goal, tissue type, and system capabilities. The MC framework introduced here can be adapted to other tissue types and imaging setups to help in this process.

Generating Benchmark HS Microscopy Databases

Chapter 5 focused on the development of benchmark HS microscopy datasets to support reproducible and clinically relevant research. Two comprehensive databases were developed, each addressing distinct clinical challenges and capturing real-world variability in tissue preparation and imaging conditions.

The first database, *HistologyHSI-GB*, offered a robust collection of glioblastoma HS images with expert annotations, supporting algorithm development in neuroncology. The second, *HistologyHSI-*

BC Recurrence, integrated multimodal HSI, histopathological, and clinical data to enable the study of breast cancer recurrence prediction. These datasets directly addressed the critical shortage of publicly available, well-annotated HS microscopic data, which had been identified as a major barrier to the objective evaluation and benchmarking of advanced image analysis methods.

By providing standardized and diverse examples of tissue morphology and spectral content, these databases facilitate reproducible assessments of classification, spectral unmixing, and biomarker discovery algorithms across various HS imaging systems and clinical use cases. Furthermore, they serve as essential resources not only for computational researchers but also for pathologists and clinicians assessing the clinical utility of HS imaging to improve diagnostic precision and support treatment decision-making. Summarizing, Chapter 5 underscored the pivotal role of curated, accessible HS microscopy datasets in bridging the gap between technological development and clinical application.

6.2 Other Contributions to HSI

Beyond the contributions carried out in line with this thesis, other contributions were also made to improve the development of HS microscopic systems. In a work (**C6**) titled "**In the use of artificial intelligence and hyperspectral imaging in digital pathology for breast cancer cell identification**" (*SPIE Medical Imaging 2022* held in San Diego), ML and DL algorithms were employed for breast tissue cells classification. Twelve breast cancer histology samples were digitized using whole-slide RGB imaging, annotated at the cellular level by pathologists, and then acquired with a HS microscopy system in the 400–1000 nm range at 20× magnification. The HS data was aligned with RGB scans, enabling precise transfer of annotations to extract individual cells for analysis. Two classification tasks were performed: distinguishing tumor cells from fibroblasts, and from lymphocytes. For tumor vs. fibroblast classification, the CNN achieved the highest sensitivity (96%) and AUC (0.91), while kNN provided the best balance between sensitivity and specificity (64%/52%). For tumor vs. lymphocyte classification, kNN again performed best in terms of balance (58.47%/ 58.86%),

with an F1-score of 74.12%. The SVM also yielded competitive results with an F1-score of 70.38%. These results demonstrate that combining HSI with ML/DL enables automated, accurate identification of cancerous cells, supporting faster and more efficient diagnosis in pathology. A similar work was carried out in the conference paper (**C8**) titled "**YOLOX-based framework for nuclei detection on whole-slide histopathological RGB and hyperspectral images**", presented in the same conference the next year (*SPIE Medical Imaging 2023* held in San Diego, USA). Building on the same dataset as previous work, this study introduces a novel framework for tumor cell detection in WSI using both RGB and HS imaging, with a focus on reducing annotation effort through training on partially labeled data. The proposed method achieved an F1-score of 66.2% on RGB images, closely matching the 68.7% obtained in fully labeled mitosis datasets [103], while maintaining consistent performance across entire tissue slides. When using PCA-reduced 3-band HS images, the framework achieved a 6.3% improvement in precision compared to synthetic RGB derived from HS cubes, indicating that even simplified HS data can enhance the performance of DL models in cancer detection tasks.

Further collaborative efforts supported the development of HSI-based systems in biomedical contexts. The knowledge obtain through the design and characterization of a HS microscopic system was later employed in a journal article **(J7**) titled "**Feasibility study of hyperspectral colposcopy as a novel tool for detecting precancerous cervical lesions**" published in *Scientific Reports*. This study introduced a custom HS colposcope designed to improve the precision of cervical lesion diagnosis, a critical global health challenge, particularly among younger women, where traditional methods often suffer from limited sensitivity, specificity, and operator dependence. The characterization of the acquisition system encompassed key parameters such as FOV, spatial frequency response, barrel distortion, spectral resolution, DOF, and optimal acquisition time. The comprehensive evaluation of both spectral and spatial characteristics offers valuable insights into the system's performance and limitations, highlighting its promising potential for advancing cervical cancer detection.

In the context of unmixing algorithms for HS microscopic data, collaborative efforts were presented in the *Optica Latin America Optics and Photonics Conference* 2024 held in Puerto Vallarta, México under the title (**C9**) "**Semi-Supervised Hyperspectral Unmixing: Integration of Fixed and Variable Endmembers**". This study introduced the ESSEAE (Extended Semi-Supervised Endmember and Abundance Extraction) methodology for spectral unmixing in cases where some spectral signatures are known in advance, while others must be estimated. Applied to histological VNIR images with known dye spectra (i.e., hematoxylin and eosin) but unknown biological components, the method shows potential for future use as a feature extraction step in HS image classification. And extension of this work was published in the journal *IEEE Access* (**J8**) under the title "**Robust and Unified Semi-Supervised Unmixing of Hyperspectral Imaging for Linear and Multilinear Models**". This paper introduced the ESSEAE and NESSEAE (Non-linear Extended Semi-Supervised Endmember and Abundance Extraction) methodologies as semi-supervised spectral unmixing approaches tailored for linear and multilinear mixing models, respectively, with applications in biomedical imaging and remote sensing. By incorporating partial prior knowledge of endmembers and addressing noise through a CCDO (Cyclic Coordinate Descent Optimization) based framework, these methods offer a flexible alternative to fully supervised or unsupervised techniques. Validation across synthetic and real-world datasets, including HS histological images, demonstrated their effectiveness in identifying both known and unknown materials. Future work will focus on automated hyperparameter tuning, broader biomedical applications, DL-based classification, and Python toolbox design encompassing various unmixing strategies.

6.3 Conclusions

This Ph.D. thesis makes significant contributions to advancing the clinical viability of HS microscopy in digital pathology by addressing key technical, methodological, and practical challenges in a systematic manner. The work spans four major domains:

- Standardization of HS Microscopy System Design and Characterization: A robust, empirically validated roadmap for the calibration and performance evaluation of HS microscopic systems was developed. This includes critical yet often neglected parameters like spatial repeatability and wavelength-dependent resolution. The introduction of this framework supports reproducibility, interoperability across devices, and lays the groundwork for future standardization efforts in biomedical HSI.
- Establishment of Objective, Spectrally Aware Focus Assessment Techniques: The Ph.D. thesis contributes to the evaluation of focus assessment algorithms tailored for HS microscopy, using a generated benchmark dataset (*HIDFA*). Findings emphasize the importance of accounting for wavelength-dependent spatial resolution and structural variability, guiding the selection of more informative, HS-specific image quality metrics. These insights advance quality assurance protocols and help ensure diagnostically reliable imaging.
- Optimization of Tissue Thickness to Maximize Spectral Discriminability: A Monte Carlo-based light transport framework was implemented to explore the influence of tissue thickness on spectral contrast. The work challenges conventional histological practices and suggests that optimal diagnostic performance requires customized thicknesses depending on tissue type and diagnostic goal. This contribution encourages a re-evaluation of sample preparation protocols for HS imaging.
- Framework for Standardized Spectral Processing in Medical HSI: By exploring system-agnostic processing evaluation methodologies, this Ph.D. thesis offers suggestions for enhancing spectral fidelity across different HS platforms, encouraging the look for consistent processing approaches (e.g., SNV normalization, appropriate smoothing filters) and encourages the development of real-time capable algorithms for clinical deployment.
- Development of Publicly Available, Clinically Oriented HS Microscopic Databases: this Ph.D. thesis presents two benchmark databases, *HistologyHSI-GB* and *HistologyHSI-*

BC Recurrence, which capture real-world variability in sample preparation and imaging conditions. These resources enable reproducible algorithm development in neurooncology and breast cancer recurrence prediction, promote collaboration between computational and clinical researchers, and lay the groundwork for advancing HS imaging toward clinical use.

As the main conclusion, this Ph.D. thesis explores the practical challenges of bringing HS microscopy closer to routine clinical use. It covers aspects such as data acquisition, system characterization, focus assessment, sample preparation, spectral processing, and the creation of public datasets to support its application in pathology. The work contributes meaningfully to open science and benchmarking by releasing resources such as the *HIDFA*, *HistologyHSI-GB* and *HistologyHSI-BC Recurrence* datasets, as well as establishing reproducible evaluation frameworks. These contributions enable the research community to validate, compare, and extend these contributions, supporting the design and development of automated diagnosis assistant tool.

6.4 Limitations and Future Work

While this Ph.D. thesis lays important groundwork for advancing HS microscopy toward clinical applicability, it is important to recognize the limitations that shape future research directions. These limitations span technical challenges, methodological constraints, and translational gaps, each of which informs how this work can be built upon to establish HS microscopy as a robust and reliable clinical tool.

One of the major contributions of this Ph.D. thesis is the development of a roadmap for evaluating HS microscopic systems. This framework provides a foundational step towards standardization, yet its impact depends on broader adoption. Future efforts must promote the routine reporting of system-level quality metrics in published datasets. Doing so will not only enhance reproducibility but also facilitate meaningful comparisons across different imaging platforms and experimental setups. Moreover, expanding the roadmap to include additional parameters, such as

polarization sensitivity, illumination geometry, and system-specific aberrations, could yield more comprehensive system characterization and further improve cross-study consistency.

Another contribution also identified effective NR-IQA tailored for focus assessment in HS microscopy, however, the current work aims primarily on static focus quality, leaving opportunities for extension into dynamic, multi-plane imaging. A logical next step is to apply these algorithms in z-stacking applications, where NR-IQA algorithms can help identify the most focused spectral bands and regions across depth. Such methods would enable the generation of all-in-focus HS cubes, offer clearer and more diagnostically useful 3D reconstructions, and mitigate the challenges posed by chromatic aberrations and limited depth of field.

Additionally, while the investigation of tissue thickness provided valuable insights into how it affects spectral contrast, it also presented certain limitations: the use of in vivo optical properties, even though histological analysis is performed on ex vivo samples that have lose some endogenous chromophores during preparation; unaccounted effects of formalin and paraffin fixation; missing instrumentation factors like optical power loss and glass slide interference; and the use of spatially homogeneous MC models that ignore tissue heterogeneity, optical aberrations, and focus variations. The methodology developed in this work can be adapted and extended to other systems and tissue types by generating more sophisticated MC models that simulate complex tissues, such as multilayer structures and spatial variations, enabling researchers to identify the optimal thickness for their specific applications. Also, while informative, MC results require empirical validation. Future work should involve acquiring HS microscopy data from real tissue samples of varying thicknesses and testing whether optimized thickness improves classification accuracy in tasks such as distinguishing between tissue types or identifying pathological states. Beyond validation, it will be critical to explore how modified sectioning protocols, such as preparing slightly thicker unstained tissue samples, can be adapted to clinical workflows without causing disruption. If proven beneficial, such adjustments could significantly enhance diagnostic performance in real-world settings.

In terms of HS data processing, smoothing, normalization and derivative methods where evaluated, but there remains room for refinement. Further optimization of smoothing window sizes is necessary to balance noise reduction with the preservation of key spectral features. Additionally, future work should investigate more sophisticated algorithms to address issues like spectral misalignment and calibration drift. Validating these methods across a wider range of datasets, and under operational constraints relevant to clinical settings, will be essential. Real-time implementation using hardware acceleration (e.g., GPUs or FPGAs) could also be explored to facilitate seamless integration into clinical imaging workflows.

Moreover, the acquisition and curation of well-annotated HS datasets are inherently time-consuming, resource-intensive and multidisciplinary processes. Each imaging session requires meticulous preparation, precise calibration, and expert annotation to ensure clinical relevance and data quality. Additionally, HS data files are typically large due to their high spectral and spatial resolution, which demands substantial disk storage capacity and poses challenges for data management, sharing, and long-term preservation. These practical constraints slow down dataset generation and limit the scale at which comprehensive, diverse databases can be built, underscoring the need for streamlined acquisition protocols, efficient data compression techniques, and collaborative frameworks that facilitate data sharing while maintaining quality standards.

Finally, while the potential of HS microscopy as a label-free, high-content imaging modality is clear, its clinical translation remains a long-term objective. Future research must design application-specific algorithms, mostly ML and DL models, for automatic signal processing and diagnosis assessment. These models should be trained using curated datasets that reflect real-world variability, including the effects of spectral noise, blur, and focus inconsistencies. In parallel, capturing HS microscopic data across an expanded wavelength range beyond the VNIR spectrum could further enhance diagnostic capabilities by enabling the detection of a broader array of biomolecules, such as DNA in the UV range and water in the IR. Once these algorithms are refined, subsequent studies should focus on deploying the proposed methodologies in

real-world clinical environments, including diverse patient populations and multicenter trials. Ensuring compatibility with existing histopathology infrastructure and navigating regulatory pathways will be crucial for adoption. In summary, the work presented here represents a significant step forward in the technical validation and standardization of HS microscopy. However, realizing its full potential in clinical practice will require continued interdisciplinary collaboration, iterative validation, and a strong emphasis on usability, reproducibility, and translational impact.

6.5 Impact of the Ph.D. Thesis

6.5.1 Research publications

The following acronyms are used to underline the distinct aspects of each listed publication: **OA**: Open Access; **OC**: Open Code; **OD**: Open Data; **NC**: National collaboration; **IC**: International collaboration; **CA**: Corresponding author; **IF**: Impact Factor; **FWCI:** Field-Weighted Citation Impact; ¥: Equal Contribution.

Journal publications

J1. **Laura Quintana-Quintana**, Gonzalo Rosa, Javier Santana-Nunez, Miguel Chavarrías, Samuel Ortega, Jaime Sancho, Himar Fabelo, Eduardo Juárez, and Gustavo M. Callico, "Roadmap for the Characterization and Validation of Hyperspectral Microscopic Systems", IEEE Transactions on Instrumentation and Measurement 74, 1-13, (**2025**). https://doi.org/10.1109/TIM.2025.3575989. [**IF: 5.9 - Q1 in Instruments & Instrumentation (2024)**] [Related with **Chapter 2**] [**OA, NC, CA**]

J2. **Laura Quintana-Quintana**, Samuel Ortega, Himar Fabelo, Francisco J. Balea-Fernández, and Gustavo M. Callico, "Blur-specific image quality assessment of microscopic hyperspectral images", Optics Express 31, 12261-12279 (**2023**). https://doi.org/10.1364/OE.476949. [**IF: 3.2 - Q2 in Optics**] [Related with **Chapter 3**] [**OA, OD**]

J3. **Laura Quintana-Quintana**, Mark Witteveen, Behdad Dashtbozorg, Samuel Ortega, Theo J.M. Ruers, Henricus J.C.M. Sterenborg, and Gustavo M. Callico, "Exploring the role of sample thickness for hyperspectral microscopy tissue discrimination through Monte Carlo simulations", Biomedical Optics Express 16, 4644-4661 (**2025**). https://doi.org/10.1364/BOE.563094. [**IF: 3.2 - Q1 in Radiology, Nuclear Medicine & Medical Imaging (2024)**] [Related with **Chapter 4**] [**OA, IC, CA**]

J4. Samuel Ortega[¥], **Laura Quintana-Quintana[¥]**, Raquel Leon[¥], Himar Fabelo, Maria de la Luz Plaza, Rafael Camacho and Gustavo M. Callico. "Histological Hyperspectral Glioblastoma Dataset (HistologyHSI-GB)", Scientific Data 11, 681 (**2024**). https://doi.org/10.1038/s41597-024-03510-x. [**IF: 6.9 - Q1 in Multidisciplinary Sciences**] [Related with **Chapter 5**] [**OA, OC, OD**]

J5. **Laura Quintana-Quintana[¥]**, Esther Sauras-Colón[¥], Alessio Fiorin, Javier Santana-Nunez, Samuel Ortega, Noèlia Gallardo-Borràs3,2, Alba Fischer-Carles, Tábata Sánchez-Alcántara, Himar Fabelo, Laia Adalid-Llansa, Daniel Mata-Cano, Ramon Bosch-Príncep, Marylène Lejeune, Gustavo M. Callico, Carlos López-Pablo. "Histological Hyperspectral Breast Cancer Recurrence Database (HistologyHSI-BC Recurrence)", Scientific Data Vol 12, 1886 (**2025**). https://doi.org/10.1038/s41597-025-06157-4. [**IF: 6.9 - Q1 in Multidisciplinary Sciences (2024)**] [Related with **Chapter 5**] [**OA, OC, OD, NC**]

J6. Carlos Urbina Ortega, Eduardo Quevedo Gutiérrez, **Laura Quintana**, Samuel Ortega, Himar Fabelo, Lucana Santos Falcón, and Gustavo Marrero Callico. 2023. "Towards Real-Time Hyperspectral Multi-Image Super-Resolution Reconstruction Applied to Histological Samples", Sensors 4, 1863 (**2023**). https://doi.org/10.3390/s23041863. [**IF: 3.4 – Q2 in Instruments & Instrumentation**] [**OA, OD, IC**]

J7. Carlos Vega, Norberto Medina, **Laura Quintana-Quintana**, Raquel Leon, Himar Fabelo, Jorge Rial, Alicia Martin and Gustavo M. Callico. "Feasibility study of hyperspectral colposcopy as a novel tool for detecting precancerous cervical lesions", Scientific Reports 15, 820 (**2025**).

https://doi.org/10.1038/s41598-024-84422-z. [**IF: 3.9 - FWCI: 3.47 - Q1 in Multidisciplinary Sciences (2024)**] [**OA, OD, NC**]

J8. Daniel Ulises Campos-Delgado, Juan Nicolás Mendoza-Chavarría, Omar Gutierrez-Navarro, **Laura Quintana-Quintana**, Raquel Leon, Samuel Ortega Sarmiento, Himar Fabelo, Carlos López, Marylène Lejeune, Gustavo M. Callico. "Robust and Unified Semi-Supervised Unmixing of Hyperspectral Imaging for Linear and Multilinear Models," in IEEE Access 13, 53140-53158 (**2025**). https://doi.org/10.1109/ACCESS.2025.3552439. [**IF: 3.6 - FWCI: 2.24 - Q2 in Engineering, Electrical & Electronic (2024)**] [**OA, IC, NC**]

Conference publications

C1. **Laura Quintana-Quintana**, Carlos Vega, Raquel Leon, Guillermo V. Socorro-Marrero, Samuel Ortega, and Gustavo M. Callico, "Assessing Processing Strategies on Data from Medical Hyperspectral Acquisition Systems," 2024 27th Euromicro Conference on Digital System Design (DSD), Paris, France, 464-471 (**2024**). https://doi.org/10.1109/DSD64264.2024.00068. [Related with **Chapter 3**] [**OA, CA**]

C2. Javier Santana-Nunez[¥], **Laura Quintana-Quintana[¥]**, Himar Fabelo, Samuel Ortega, Esther Sauras-Colón, Noèlia Gallardo-Borràs, Daniel Mata-Cano, Carlos López-Pablo, and Gustavo M. Callico "Analysing histology hyperspectral images: Does tissue thickness matter?", Proc. SPIE 13006, Photonics Europe 2024: Biomedical Spectroscopy, Microscopy, and Imaging III, 1300611 (**2024**). https://doi.org/10.1117/12.3017010. [Related with **Chapter 4**] [**OA, NC, CA**]

C3. **Laura Quintana**, Samuel Ortega, Raquel Leon, Himar Fabelo, Gustavo M. Callico, Carlos Lopez, Marylene Lejeune and, Ramon Bosch "Instrumentation Evaluation for Hyperspectral Microscopy Targeting Enhanced Medical Histology," 2021 XXXVI Conference on Design of Circuits and Integrated Systems (DCIS), Vila do Conde, Portugal, 1-6

(**2021**). https://doi.org/10.1109/DCIS53048.2021.9666188. [Related with **Chapter 2**] [**OA, NC, CA**]

C4. **Laura Quintana,** Samuel Ortega, Himar Fabelo, and Gustavo M. Callico, "Blur-Specific No-Reference Image Quality Assesment for Microscopic Hyperspectral Image Focus Quantification," 2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, Netherlands 1-5 (**2021**). https://doi.org/10.1109/WHISPERS52202.2021.9483992. [Related with **Chapter 3**] [IC]

C5. Raquel Leon, Sofia H. Gelado, Himar Fabelo, Samuel Ortega, **Laura Quintana**, Adam Szolna, Juan F. Piñeiro, Francisco Balea-Fernandez, Jesus Morera, Bernardino Clavo, and Gustavo M. Callico "Hyperspectral imaging for in-vivo/ex-vivo tissue analysis of human brain cancer", Proc. SPIE 12034, Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling, 1203429 (**2022**). https://doi.org/10.1117/12.2611420. [**FWCI: 3.17**] [**OA**]

C6. **Laura Quintana**, Samuel Ortega, Raquel Leon, Himar Fabelo, Francisco J. Balea-Fernández, Esther Sauras, Marylene Lejeune, Ramon Bosch, Carlos Lopez, and Gustavo M. Callico "In the use of artificial intelligence and hyperspectral imaging in digital pathology for breast cancer cell identification", Proc. SPIE 12039, Medical Imaging 2022: Digital and Computational Pathology, 120390E (**2022**). https://doi.org/10.1117/12.2611419. [**OA, NC, CA**]

C7. Carlos Urbina Ortega, Eduardo Gregorio Quevedo Gutiérrez, **Laura Quintana- Quintana**, Samuel Ortega, Lucana Santos Falcón, and Gustavo M. Callicó. "Towards real-time management of satellite microvibrations for on-board hyperspectral image quality enhancement." (**2023**). [**OA**]

C8. Carlos Vega, **Laura Quintana**, Samuel Ortega, Himar Fabelo, Esther Sauras, Noèlia Gallardo, Daniel Mata, Marylene Lejeune, Carlos Lopez, and Gustavo M. Callico "YOLOX-based framework for nuclei detection on whole-slide histopathological RGB and hyperspectral images", Proc. SPIE 12471, Medical Imaging 2023: Digital and Computational Pathology, 124711A (**2023**). https://doi.org/10.1117/12.2654036. [**OA**]

C9. Juan N. Mendoza Chavarría, **Laura Quintana-Quintana**, Samuel Ortega, Gustavo M. Callico, and Daniel U. Campos-Delgado, "Semi-Supervised Hyperspectral Unmixing: Integration of Fixed and Variable End-members," in Optica Latin America Optics and Photonics Conference (LAOP) 2024, Technical Digest Series (Optica Publishing Group), paper M3D.3 (**2024**). https://doi.org/10.1364/LAOP.2024.M3D.3 [**OA, IC**]

6.5.2 Research projects and groups

European projects

EP1. Ref. 101137416: **STRATUM** (3D Decision Support Tool for Brain Tumor Surgery – GA: 101137416). European Union's Horizon Europe Programme HORIZON-IA action. PI: Gustavo M. Callico. ULPGC. 01/12/2023 - 30/11/2028. 9,990,566.25 €. Task: Support in the coordination and management of the project, including handling multimodal data, evaluating acquisition systems, analyzing data with specialized software, and contributing to reports and publications.

EP2. Ref. 101017385: **WARIFA** (Artificial intelligence and the prevention of chronic conditions – GA: 101017385). European Union's Horizon 2020 research and innovation programme. PI: Gustavo M. Callico and Ana Wägner. ULPGC. 01/01/2021 – 31/12/2024. 6,726,468.75 €. Task: defining objectives, identifying diverse data sources, assessing and ensuring data quality and organizing and integrating data in the dermatology use case.

National projects

NP1. Ref. PID2023-148285OB-C43: **OASIS** (Open Ai-driven Stack para plataformas hpec mejoradas en Sistemas Integrado). Spanish Government and European Union (FEDER funds). PIs: Gustavo M. Callico and Pedro F. Pérez Carballo. ULPGC. 01/09/2024 - 31/12/2027 240,750.00 €. Task: Optimization and operation of the laboratory HS microscopic imaging system, including database management and the design and development of algorithms for data processing.

NP2. Ref. PID2020-116417RB-C42: **TALENT** (HypErsPEctRal Imaging for Artificial intelligence applications). Spanish Government and European Union (FEDER funds). PI: Gustavo M. Callico and Sebastian Lopez. ULPGC. 01/09/2021 - 01/09/2024 175,813.00 €. Task: Optimization and operation of the laboratory HS microscopic imaging system, including database management and the design and development of algorithms for data processing.

NP3. Ref. TEC2017-86722-C4-1-R: **PLATINO** (Plataforma HW/SW Distribuida para el Procesamiento Inteligente de Información Sensorial Heterogénea en Aplicaciones de Supervisión de Grandes Espacios Naturales). MINECO (Spain) Proyectos I+D+i – Retos Investigación. PIs: Sebastian Lopez and Gustavo M. Callico. ULPGC. 01/01/2018 - 31/12/2020. 228,690.00 €. Task: Design and implementation of data processing algorithms over HS databases.

Regional projects

RP1. Ref. ProID2017010164: **ITHACA** (IdenTificacion Hiperespectral de tumores CerebrAles). Gobierno de Canarias (Canary Islands) Programa de Apoyo a la Investigación María del Carmen Betancourt y Molina. PI: Gustavo M. Callico. ULPGC. 01/01/2018-30/09/2019. 69,914.45 €. Task: Processing of HS data.

RP2. Ref. 016/2019: **O3NPIQ** (Dolor por neuropatía periférica inducida por quimioterapia: Valor diagnóstico y predictivo de la imagen hiperespectral en pacientes del ensayo clínico) [(Clinical Trial: EudraCT 2019-000821-37) - (BF1-19-03) - (PI 19/00458) - (016/2019)]. Fundación DISA (Canary Islands) Premios Fundación DISA 2019 a la Investigación Biomédica. PI: Bernardino Clavo Varas. Hospital Univ. de Gran Canaria Dr. Negrín. 01/01/2020-01/01/2022. 15,600.00 €. Task: Development of an HS and thermal infrared acquisition system.

RP3. Ref. PIFUN44/17: **O3Cardio** (Effectiveness and cost-effectiveness of Ozone therapy in patients with ischemic heart disease refractory to medical and surgical treatment: Randomized, triple-blind clinical trial [EudraCT 2018-

000201-24]). FUNCANIS (Canary Islands) Ayudas para la financiación de proyectos de investigación, desarrollo einnovación en Biomedicina y Ciencias de la Salud. PI: Bernardino Clavo. Hospital Univ. de Gran Canaria Dr. Negrín. 01/01/2018-31/12/2019. 21,804.00 €. Task: Development of an HS and thermal infrared acquisition system.

RP4. Ref. FFPI21/BE01: Use of hyperspectral imaging as a prognostic factor for relapse due to distant metastasis in breast cancer. Fundació Doctor Ferran. PI: Esther Sauras. Hospital Verge de La Cinta. 01/01/2022-31/12/2022. 3,000 €. Task: Acquisition, processing, and analysis of a multimodal breast cancer dataset to evaluate indicators associated with tumor recurrence.

RP5. Ref. PIFIISC23/13: Valor diagnóstico y predictivo de la imagen hiperespectral en la neuropatíaperiférica inducida por quimioterapia. Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). PI: Bernardino Clavo Varas. Hospital Univ. de Gran Canaria Dr. Negrín. 01/12/2023 - 30/12/2025. 25.000 €. Task: Support in the acquisition of multimodal images, ensuring proper equipment setup, and adherence to acquisition protocols.

RP6. Ref. PIFIISC23/44: Evaluación con imagen hiperespectral del estrés oxidativo en pacientes con toxicidad por radioterapia o quimioterapia. Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC). PIs: Francisco Rodriguez Esparragon and Francisco Balea Fernandez. Hospital Univ. de Gran Canaria Dr. Negrín. 01/12/2023 - 30/12/2025. 12.000 €. Task: Capture of HS histology database of human blood and development of algorithms for the analysis of its oxidative stress.

Research groups

RG1. Member of the research group Design of Integrated Electronic Systems for Data Processing at Research Institute of Applied Microelectronics (IUMA-DSI), ULPGC. 01/09/2019 - *present*.

6.5.3 Grants and awards

Public Research contract

RC1. Researcher in the PLATINO project (**NP3**). 16/02/2021 - 15/05/2021.

Pre-doctoral Grant

G1. Pre-doctoral grant given by the "Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI)" of the "Consejería de Economía, Conocimiento y Empleo" of the "Gobierno de Canarias", which is part-financed by the European Social Fund (FSE) (POC 2014-2020, Eje 3 Tema Prioritario 74 (85%)). 01/07/2021 -01/06/2025.

Mobility Grants

M1. Research intership movility at the Hospital of Tortosa Verge de la Cinta, Tortosa, Spain. Financed by the "Consejo Social" of the "Universidad de Las Palmas de Gran Canaria". It took place from 16/05/2022 to 20/07/2022. Supervisor: Carlos López Pablo.

M2. Research intership movility at the Centro de Investigación en Tecnologías Software y Sistemas Multimedia (CITSEM) at the Universidad Politécnica de Madrid, Spain. Financed by the "Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI)" of the "Consejería de Universidades, Ciencia e Innovación y Cultura" of the "Gobierno de Canarias", which is part-financed by the European Social Fund Plus (FSE+) (BOC number 172 of 31/08/2023). It took place from 01/06/2023 to 30/06/2023. Supervisor: Eduardo Juarez Martínez.

M3. Research intership movility at the Image guided surgery Research Group at the Netherlands Cancer Institute, Netherlands. Financed by the "Agencia Canaria de Investigacion, Innovacion y Sociedad de la Información (ACIISI)" of the "Consejería de Universidades, Ciencia e Innovación y Cultura" of the "Gobierno de Canarias", which is part-financed by the European Social Fund Plus (FSE+) (BOC number 172 of 31/08/2023). It took place from 15/09/2023 - 30/11/2023. Supervisor: Behdad Dasht Bozorg.

M4. 5th International MCX Training Workshop at the Boston campus of Northeastern University, USA. Financed by the TALENT (HypErsPEctRal Imaging for Artificial intelligence applications) project (PID2020-116417RB-C41 and PID2020-116417RB-C42 AEI/10.13039/501100011033), supported by the Spanish Government and European Union (FEDER Funds). It took place from 29/07/2024 -01/07/2024. Professor: Qianqian Fang.

Awards

A1. Best Paper Award at "XXXVI Conference on Design of Circuits and Integrated Systems". DCIS2021 General Chairs. Tool place: 26/11/2021 in Vila do Conde, Portugal.

Bibliography

[1] F. Bray *et al.*, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J Clin*, vol. 74, no. 3, pp. 229–263, May 2024, https://doi.org/10.3322/caac.21834.

[2] "Global cancer burden growing, amidst mounting need for services." Accessed: Jul. 13, 2025. [Online]. Available: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services

[3] "Global Cancer Observatory." Accessed: Jul. 13, 2025. [Online]. Available: https://gco.iarc.who.int/en

[4] "Cancer deaths rose to 10 million worldwide in 2019 | ScienceDaily." Accessed: Jul. 13, 2025. [Online]. Available: https://www.sciencedaily.com/releases/2021/12/211230130926.htm

[5] L. Fass, "Imaging and cancer: A review," *Mol Oncol*, vol. 2, no. 2, pp. 115–152, Aug. 2008, https://doi.org/10.1016/j.molonc.2008.04.001.

[6] M. Titford, "A Short History of Histopathology Technique," *J Histotechnol*, vol. 29, no. 2, pp. 99–110, Jun. 2006, https://doi.org/10.1179/his.2006.29.2.99.

[7] D. B. McMillan and R. James Harris, *An Atlas of Comparative Vertebrate Histology*. Elsevier, 2018. https://doi.org/10.1016/C2012-0-06909-8.

[8] J. L. Burton and J. Underwood, "Clinical, educational, and epidemiological value of autopsy," *The Lancet*, vol. 369, no. 9571, pp. 1471–1480, Apr. 2007, https://doi.org/10.1016/S0140-6736(07)60376-6.

[9] M. J. Thrall, "Manual of Surgical Pathology, 2nd edition: Author: Susan C. Lester, Illustrators: Christopher A. French and Glenn Curtis, Elsevier Churchill Livingstone, Philadelphia, 2006," *Diagn Cytopathol*, vol. 34, no. 9, pp. 659–659, Sep. 2006, https://doi.org/10.1002/dc.20482.

[10] K. Jensen, "Theory and Practice of Histological Techniques, 6th Edition," *J Neuropathol Exp Neurol*, vol. 67, no. 6, pp. 633.2-633, Jun. 2008, https://doi.org/10.1097/NEN.0b013e31817e2933.

[11] S. S. Cross, "Grading and scoring in histopathology," *Histopathology*, vol. 33, no. 2, pp. 99–106, Aug. 1998, https://doi.org/10.1046/j.1365-2559.1998.00495.x.

[12] E. Jain *et al.*, "Whole Slide Imaging Technology and Its Applications: Current and Emerging Perspectives," *Int J Surg Pathol*, vol. 32, no. 3, pp. 433–448, May 2024, https://doi.org/10.1177/10668969231185089.

[13] L. Pantanowitz, A. Sharma, A. B. Carter, T. Kurc, A. Sussman, and J. Saltz, "Twenty Years of Digital Pathology: An Overview of the Road Travelled, What is on the Horizon, and the Emergence of Vendor-Neutral Archives," *J*

Pathol Inform, vol. 9, no. 1, p. 40, Jan. 2018, https://doi.org/10.4103/jpi.jpi_69_18.

[14] T. Mezei, M. Kolcsár, A. Joó, and S. Gurzu, "Image Analysis in Histopathology and Cytopathology: From Early Days to Current Perspectives," *J Imaging*, vol. 10, no. 10, p. 252, Oct. 2024, https://doi.org/10.3390/jimaging10100252.

[15] A. Soliman, Z. Li, and A. V. Parwani, "Artificial intelligence's impact on breast cancer pathology: a literature review," *Diagn Pathol*, vol. 19, no. 1, p. 38, Feb. 2024, https://doi.org/10.1186/s13000-024-01453-w.

[16] Y. Zhao *et al.*, "A review of cancer data fusion methods based on deep learning," *Information Fusion*, vol. 108, p. 102361, Aug. 2024, https://doi.org/10.1016/j.inffus.2024.102361.

[17] A. Llinas-Bertran, M. Butjosa-Espín, V. Barberi, and J. A. Seoane, "Multimodal data integration in early-stage breast cancer," *The Breast*, vol. 80, p. 103892, Apr. 2025, https://doi.org/10.1016/j.breast.2025.103892.

[18] J. Lipkova *et al.*, "Artificial intelligence for multimodal data integration in oncology," *Cancer Cell*, vol. 40, no. 10, pp. 1095–1110, Oct. 2022, https://doi.org/10.1016/j.ccell.2022.09.012.

[19] G. ElMasry and D.-W. Sun, "Principles of Hyperspectral Imaging Technology," in *Hyperspectral Imaging for Food Quality Analysis and Control*, Elsevier, 2010, pp. 3–43. https://doi.org/10.1016/B978-0-12-374753-2.10001-2.

[20] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*, vol. 1. Boston, MA: Springer Science & Business Media, 2003. https://doi.org/10.1007/978-1-4419-9170-6.

[21] F. Mohammadimanesh, M. Mahdianpari, A. Radman, D. Varon, M. Hemati, and M. Marjani, "Advancements in satellite-based methane point source monitoring: A systematic review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 224, pp. 94–112, Jun. 2025, https://doi.org/10.1016/j.isprsjprs.2025.03.020.

[22] B. Lu, P. Dao, J. Liu, Y. He, and J. Shang, "Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture," *Remote Sens (Basel)*, vol. 12, no. 16, p. 2659, Aug. 2020, https://doi.org/10.3390/rs12162659.

[23] S. Peyghambari and Y. Zhang, "Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review," *J Appl Remote Sens*, vol. 15, no. 03, Jul. 2021, https://doi.org/10.1117/1.JRS.15.031501.

[24] B. Fei, "Hyperspectral imaging in medical applications," in *Data handling in science and technology*, vol. 32, Elsevier, Ed., 2019, pp. 523–565. https://doi.org/10.1016/B978-0-444-63977-6.00021-3.

[25] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *J Biomed Opt*, vol. 19, no. 1, p. 010901, Jan. 2014, https://doi.org/10.1117/1.JBO.19.1.010901.

[26] S. Ortega, M. Halicek, H. Fabelo, G. M. Callico, and B. Fei, "Hyperspectral and multispectral imaging in digital and computational pathology: a systematic review," *Biomed Opt Express*, vol. 11, no. 6, p. 3195, Jun. 2020, https://doi.org/10.1364/BOE.386338.

[27] M. Al Ktash, M. Knoblich, M. Eberle, F. Wackenhut, and M. Brecht, "UV Hyperspectral Imaging with Xenon and Deuterium Light Sources:

Integrating PCA and Neural Networks for Analysis of Different Raw Cotton Types," *J Imaging*, vol. 10, no. 12, p. 310, Dec. 2024, https://doi.org/10.3390/jimaging10120310.

[28] U. Simoncic and M. Milanic, "Hyperspectral Imaging with Active Illumination: A Theoretical Study on the Use of Incandescent Lamp and Variable Filament Temperature," *Sensors*, vol. 23, no. 23, p. 9326, Nov. 2023, https://doi.org/10.3390/s23239326.

[29] T. Adão *et al.*, "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sens (Basel)*, vol. 9, no. 11, 2017, https://doi.org/10.3390/rs9111110.

[30] M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic range improvement through multiple exposures," in *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, IEEE, pp. 159–163. https://doi.org/10.1109/ICIP.1999.817091.

[31] H.-T. Lim and V. M. Murukeshan, "Spatial-scanning hyperspectral imaging probe for bio-imaging applications," *Review of Scientific Instruments*, vol. 87, no. 3, p. 033707, Mar. 2016, https://doi.org/10.1063/1.4943968.

[32] S. Ortega, "Automatic classification of histological hyperspectral images: algorithms and instrumentation," 2021.

[33] X. Cao *et al.*, "Computational Snapshot Multispectral Cameras: Toward dynamic capture of the spectral world," *IEEE Signal Process Mag*, vol. 33, no. 5, pp. 95–108, Sep. 2016, https://doi.org/10.1109/MSP.2016.2582378.

[34] C. Loebich, D. Wueller, B. Klingen, and A. Jaeger, "Digital camera resolution measurement using sinusoidal Siemens stars," in *Digital Photography III*, R. A. Martin, J. M. DiCarlo, and N. Sampat, Eds., Feb. 2007, pp. 214–224. https://doi.org/10.1117/12.703817.

[35] S. Ortega *et al.*, "Hyperspectral Push-Broom Microscope Development and Characterization," *IEEE Access*, vol. 7, pp. 122473–122491, 2019, https://doi.org/10.1109/ACCESS.2019.2937729.

[36] S. l. Tan *et al.*, "Detectivity Optimization of InGaAs Photon Emission Microscope Systems," in *13th International Symposium on the Physical and Failure Analysis of Integrated Circuits*, IEEE, Jul. 2006, pp. 315–319. https://doi.org/10.1109/IPFA.2006.251053.

[37] S. L. Jacques, "Optical properties of biological tissues: a review," *Phys Med Biol*, vol. 58, no. 11, pp. R37–R61, Jun. 2013, https://doi.org/10.1088/0031-9155/58/11/R37.

[38] W. M. Star, "Diffusion Theory of Light Transport," in *Optical-Thermal Response of Laser-Irradiated Tissue*, Boston, MA: Springer US, 1995, pp. 131–206. https://doi.org/10.1007/978-1-4757-6092-7_6.

[39] H. Suzaki *et al.*, "Noninvasive measurement of total hemoglobin and hemoglobin derivatives using multiwavelength pulse spectrophotometry - In vitro study with a mock circulatory system," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, Aug. 2006, pp. 799–802. https://doi.org/10.1109/IEMBS.2006.260273.

[40] G. M. Hale and M. R. Querry, "Optical Constants of Water in the 200-nm to 200-µm Wavelength Region," *Appl Opt*, vol. 12, no. 3, p. 555, Mar. 1973, https://doi.org/10.1364/AO.12.000555.

[41] R. L. P. van Veen, H. J. C. M. Sterenborg, A. Pifferi, A. Torricelli, E. Chikoidze, and R. Cubeddu, "Determination of visible near-IR absorption

coefficients of mammalian fat using time- and spatially resolved diffuse reflectance and transmission spectroscopy," *J Biomed Opt*, vol. 10, no. 5, p. 054004, 2005, https://doi.org/10.1117/1.2085149.

[42] S. L. Jacques and D. J. McAuliffe, "The Melanosome: Threshold Temperature for Explosive Vaporization and Internal Absorption Coefficient during Pulsed Laser Irradiation," *Photochem Photobiol*, vol. 53, no. 6, pp. 769–775, Jun. 1991, https://doi.org/10.1111/j.1751-1097.1991.tb09891.x.

[43] R. R. Anderson and J. A. Parrish, "Selective Photothermolysis: Precise Microsurgery by Selective Absorption of Pulsed Radiation," *Science (1979)*, vol. 220, no. 4596, pp. 524–527, Apr. 1983, https://doi.org/10.1126/science.6836297.

[44] S. L. Jacques, "Origins of Tissue Optical Properties in the UVA, Visible, and NIR Regions," in *Advances in Optical Imaging and Photon Migration*, Washington, D.C.: Optica Publishing Group, 1996, p. OPC364. https://doi.org/10.1364/AOIPM.1996.OPC364.

[45] K. J. Parker, "Shapes and distributions of soft tissue scatterers," *Phys Med Biol*, vol. 64, no. 17, p. 175022, Sep. 2019, https://doi.org/10.1088/1361-6560/ab2485.

[46] V. V. Tuchin, *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*. Society of Photo-Optical Instrumentation Engineers (SPIE), 2015. https://doi.org/10.1117/3.1003040.

[47] A. R. Young, "Chromophores in human skin," *Phys Med Biol*, vol. 42, no. 5, pp. 789–802, May 1997, https://doi.org/10.1088/0031-9155/42/5/004.

[48] "Assorted Spectra." Accessed: Jul. 23, 2025. [Online]. Available: https://omlc.org/spectra/index.html

[49] C. Boudoux, *Fundamentals of biomedical optics*. Blurb, 2017.

[50] S. Ortega, M. Halicek, H. Fabelo, E. Quevedo, B. Fei, and G. Marrero Callico, "Information Extraction Techniques in Hyperspectral Imaging Biomedical Applications," in *Multimedia Information Retrieval*, IntechOpen, 2021. https://doi.org/10.5772/intechopen.93960.

[51] A. Bekina *et al.*, "Multispectral assessment of skin malformations using a modified video-microscope," *LatJP*, vol. 49, no. 5, pp. 4–8, Oct. 2012, https://doi.org/10.2478/V10047-012-0024-2.

[52] L. Rey-Barroso *et al.*, "Visible and Extended Near-Infrared Multispectral Imaging for Skin Cancer Diagnosis," *Sensors*, vol. 18, no. 5, p. 1441, May 2018, https://doi.org/10.3390/s18051441.

[53] B. Fei *et al.*, "Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients," *J Biomed Opt*, vol. 22, no. 08, p. 1, Aug. 2017, https://doi.org/10.1117/1.JBO.22.8.086009.

[54] E. Kho *et al.*, "Hyperspectral Imaging for Resection Margin Assessment during Cancer Surgery," *Clinical Cancer Research*, vol. 25, no. 12, pp. 3572–3580, Jun. 2019, https://doi.org/10.1158/1078-0432.CCR-18-2089.

[55] G. M. Palmer, C. Zhu, T. M. Breslin, F. Xu, K. W. Gilchrist, and N. Ramanujam, "Monte Carlo-based inverse model for calculating tissue optical properties Part II: Application to breast cancer diagnosis," *Appl Opt*, vol. 45, no. 5, p. 1072, Feb. 2006, https://doi.org/10.1364/AO.45.001072.

[56] C. Zhu, G. M. Palmer, T. M. Breslin, J. Harter, and N. Ramanujam, "Diagnosis of breast cancer using fluorescence and diffuse reflectance spectroscopy: a Monte-Carlo-model-based approach," *J Biomed Opt*, vol. 13, no. 3, p. 034015, 2008, https://doi.org/10.1117/1.2931078.

[57] P. Saccomandi *et al.*, "Estimation of anisotropy coefficient of swine pancreas, liver and muscle at 1064 nm based on goniometric technique," *J Biophotonics*, vol. 8, no. 5, pp. 422–428, May 2015, https://doi.org/10.1002/jbio.201400057.

[58] Q. Wang, J. Wang, M. Zhou, Q. Li, and Y. Wang, "Spectral-spatial feature-based neural network method for acute lymphoblastic leukemia cell identification via microscopic hyperspectral imaging technology," *Biomed Opt Express*, vol. 8, no. 6, p. 3017, Jun. 2017, https://doi.org/10.1364/BOE.8.003017.

[59] K. S. S. Enfield *et al.*, "Hyperspectral cell sociology reveals spatial tumor-immune cell interactions associated with lung cancer recurrence," *J Immunother Cancer*, vol. 7, no. 1, p. 13, Jan. 2019, https://doi.org/10.1186/S40425-018-0488-6.

[60] Y. Khouj, J. Dawson, J. Coad, and L. Vona-Davis, "Hyperspectral Imaging and K-Means Classification for Histologic Evaluation of Ductal Carcinoma In Situ," *Front Oncol*, vol. 8, Feb. 2018, https://doi.org/10.3389/fonc.2018.00017.

[61] Q. Wang *et al.*, "Identification of Melanoma From Hyperspectral Pathology Image Using 3D Convolutional Networks," *IEEE Trans Med Imaging*, vol. 40, no. 1, pp. 218–227, Jan. 2021, https://doi.org/10.1109/TMI.2020.3024923.

[62] L. Ma, J. V. Little, A. Y. Chen, L. Myers, B. D. Sumer, and B. Fei, "Automatic detection of head and neck squamous cell carcinoma on histologic slides using hyperspectral microscopic imaging," *J Biomed Opt*, vol. 27, no. 04, Apr. 2022, https://doi.org/10.1117/1.JBO.27.4.046501.

[63] K. S. Banu, M. Lerma, S. U. Ahmed, and J. L. Gardea-Torresdey, "Hyperspectral microscopy- applications of hyperspectral imaging techniques in different fields of science: a review of recent advances," *Appl Spectrosc Rev*, pp. 1–24, Oct. 2023, https://doi.org/10.1080/05704928.2023.2270035.

[64] J. Stergar, R. Hren, and M. Milanič, "Design and Validation of a Custom-Made Hyperspectral Microscope Imaging System for Biomedical Applications," *Sensors*, vol. 23, no. 5, p. 2374, Feb. 2023, https://doi.org/10.3390/s23052374.

[65] L. Quintana-Quintana *et al.*, "Roadmap for the Characterization and Validation of Hyperspectral Microscopic Systems," *IEEE Trans Instrum Meas*, 2025, https://doi.org/10.1109/TIM.2025.3575989.

[66] "Encapsulated Gray Scale Standards - Avian Technologies." Accessed: Jun. 03, 2021. [Online]. Available: https://aviantechnologies.com/product/encapsulated-gray-scale-standards/

[67] "Reflectance Wavelength Calibration Standards - Avian Technologies." Accessed: Jun. 02, 2021. [Online]. Available: https://aviantechnologies.com/product/reflectance-wavelength-calibration-standards/

[68] B. Boldrini, W. Kessler, K. Rebner, and R. W. Kessler, “Hyperspectral Imaging: A Review of Best Practice, Performance and Pitfalls for in-line and on-line Applications,” *J Near Infrared Spectrosc*, vol. 20, no. 5, pp. 483–508, Oct. 2012, https://doi.org/10.1255/jnirs.1003.

[69] P. Geladi, J. Burger, and T. Lestander, “Hyperspectral imaging: calibration problems and solutions,” *Chemometrics and Intelligent Laboratory Systems*, vol. 72, no. 2, pp. 209–217, Jul. 2004, https://doi.org/10.1016/j.chemolab.2004.01.023.

[70] D. Nouri, Y. Lucas, and S. Treuillet, “Calibration and test of a hyperspectral imaging prototype for intra-operative surgical assistance,” M. N. Gurcan and A. Madabhushi, Eds., Mar. 2013, p. 86760P. https://doi.org/10.1117/12.2006620.

[71] A. Noviyanto and W. H. Abdulla, “Segmentation and calibration of hyperspectral imaging for honey analysis,” *Comput Electron Agric*, vol. 159, pp. 129–139, Apr. 2019, https://doi.org/10.1016/j.compag.2019.02.006.

[72] G. Yang *et al.*, “The DOM Generation and Precise Radiometric Calibration of a UAV-Mounted Miniature Snapshot Hyperspectral Imager,” *Remote Sens (Basel)*, vol. 9, no. 7, p. 642, Jun. 2017, https://doi.org/10.3390/rs9070642.

[73] D. Cozzolino, P. J. Williams, and L. C. Hoffman, “An overview of pre-processing methods available for hyperspectral imaging applications,” *Microchemical Journal*, vol. 193, p. 109129, Oct. 2023, https://doi.org/10.1016/j.microc.2023.109129.

[74] J. Ezenarro, D. Schorn-García, O. Busto, and R. Boqué, “ProSpecTool: A MATLAB toolbox for spectral preprocessing selection,” *Chemometrics and Intelligent Laboratory Systems*, vol. 247, p. 105096, Apr. 2024, https://doi.org/10.1016/j.chemolab.2024.105096.

[75] S. Singh and S. S. Kasana, “A Pre-processing framework for spectral classification of hyperspectral images,” *Multimed Tools Appl*, vol. 80, no. 1, pp. 243–261, Jan. 2021, https://doi.org/10.1007/s11042-020-09180-2.

[76] M. Witteveen, H. J. C. M. Sterenborg, T. G. van Leeuwen, M. C. G. Aalders, T. J. M. Ruers, and A. L. Post, “Comparison of preprocessing techniques to reduce nontissue-related variations in hyperspectral reflectance imaging,” *J Biomed Opt*, vol. 27, no. 10, Oct. 2022, https://doi.org/10.1117/1.JBO.27.10.106003.

[77] S. Ortega *et al.*, “Hyperspectral Imaging for the Detection of Glioblastoma Tumor Cells in H&E Slides Using Convolutional Neural Networks,” *Sensors*, vol. 20, no. 7, p. 1911, Mar. 2020, https://doi.org/10.3390/s20071911.

[78] L. Ma, M. Halicek, X. Zhou, J. D. Dormer, and B. Fei, “Hyperspectral microscopic imaging for automatic detection of head and neck squamous cell carcinoma using histologic image and machine learning,” in *Medical Imaging 2020: Digital Pathology*, J. E. Tomaszewski and A. D. Ward, Eds., SPIE, Mar. 2020, p. 31. https://doi.org/10.1117/12.2549369.

[79] L. Fass, “Imaging and cancer: A review,” *Mol Oncol*, vol. 2, no. 2, pp. 115–152, Aug. 2008, https://doi.org/10.1016/j.molonc.2008.04.001.

[80] B. Hu, J. Du, Z. Zhang, and Q. Wang, “Tumor tissue classification based on micro-hyperspectral technology and deep learning,” *Biomed Opt Express*, vol. 10, no. 12, p. 6370, Dec. 2019, https://doi.org/10.1364/BOE.10.006370.

[81] S. Ortega *et al.*, "Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images," in *Medical Imaging 2020: Digital Pathology*, J. E. Tomaszewski and A. D. Ward, Eds., SPIE, Mar. 2020, p. 30. https://doi.org/10.1117/12.2548609.

[82] S. Ortega, H. Fabelo, R. Camacho, M. de la Luz Plaza, G. M. Callicó, and R. Sarmiento, "Detecting brain tumor in pathological slides using hyperspectral imaging," *Biomed Opt Express*, vol. 9, no. 2, p. 818, Feb. 2018, https://doi.org/10.1364/BOE.9.000818.

[83] D. Pertzborn *et al.*, "Intraoperative Assessment of Tumor Margins in Tissue Sections with Hyperspectral Imaging and Machine Learning," *Cancers (Basel)*, vol. 15, no. 1, p. 213, Dec. 2022, https://doi.org/10.3390/cancers15010213.

[84] H. Pu, L. Lin, and D. Sun, "Principles of Hyperspectral Microscope Imaging Techniques and Their Applications in Food Quality and Safety Detection: A Review," *Compr Rev Food Sci Food Saf*, vol. 18, no. 4, pp. 853–866, Jul. 2019, https://doi.org/10.1111/1541-4337.12432.

[85] "Characterization and Calibration of Hyperspectral Imaging Devices Working Group (P4001) - GRSS-IEEE." Accessed: Nov. 30, 2025. [Online]. Available: https://www.grss-ieee.org/technical-committees/standards-for-earth-observations/working-group-standards-for-earth-observations/characterization-and-calibration-of-hyperspectral-imaging-devices-working-group-p4001/

[86] A. V. Paterova, S. M. Maniam, H. Yang, G. Grenci, and L. A. Krivitsky, "Hyperspectral infrared microscopy with visible light," *Sci Adv*, vol. 6, no. 44, Oct. 2020, https://doi.org/10.1126/sciadv.abd0460.

[87] O. Pekkala, T. Pulli, A. Kokka, and E. Ikonen, "Setup for characterising the spectral responsivity of Fabry–Pérot-interferometer-based hyperspectral cameras," *Metrologia*, vol. 56, no. 6, p. 065005, Dec. 2019, https://doi.org/10.1088/1681-7575/ab3fd1.

[88] M. S. Shaikh, K. Jaferzadeh, and B. Thörnberg, "Extending Effective Dynamic Range of Hyperspectral Line Cameras for Short Wave Infrared Imaging," *Sensors*, vol. 22, no. 5, p. 1817, Feb. 2022, https://doi.org/10.3390/s22051817.

[89] N. A. Épshtein, "Validation of HPLC Techniques for Pharmaceutical Analysis," *Pharm Chem J*, vol. 38, no. 4, pp. 212–228, Apr. 2004, https://doi.org/10.1023/B:PHAC.0000038422.27193.6c.

[90] L. Palmieri, G. Scrofani, N. Incardona, G. Saavedra, M. Martínez-Corral, and R. Koch, "Robust Depth Estimation for Light Field Microscopy," *Sensors*, vol. 19, no. 3, p. 500, Jan. 2019, https://doi.org/10.3390/s19030500.

[91] G. Wang and N. Fang, "Detecting and Tracking Nonfluorescent Nanoparticle Probes in Live Cells," 2012, pp. 83–108. https://doi.org/10.1016/B978-0-12-391857-4.00004-5.

[92] E. A. El-Gabry, A. V. Parwani, and L. Pantanowitz, "Whole-slide imaging: widening the scope of cytopathology," *Diagn Histopathol*, vol. 20, no. 12, pp. 456–461, Dec. 2014, https://doi.org/10.1016/j.mpdhp.2014.10.006.

[93] C. Meng, P. An, X. Huang, C. Yang, L. Shen, and B. Wang, "Objective Quality Assessment of Lenslet Light Field Image Based on Focus Stack," *IEEE Trans Multimedia*, pp. 1–1, 2021, https://doi.org/10.1109/TMM.2021.3096071.

[94] D. Li and T. Jiang, “Blur-Specific No-Reference Image Quality Assessment: A Classification and Review of Representative Methods,” 2019, pp. 45–68. https://doi.org/10.1007/978-3-319-91659-0_4.

[95] J. Engel *et al.*, “Breaking with trends in pre-processing?,” *TrAC Trends in Analytical Chemistry*, vol. 50, pp. 96–106, Oct. 2013, https://doi.org/10.1016/j.trac.2013.04.015.

[96] R. Koprowski, “Hyperspectral imaging in medicine: image pre-processing problems and solutions in Matlab,” *J Biophotonics*, vol. 8, no. 11–12, pp. 935–943, Nov. 2015, https://doi.org/10.1002/jbio.201400133.

[97] Åsmund Rinnan, F. van den Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra,” *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, Nov. 2009, https://doi.org/10.1016/j.trac.2009.07.007.

[98] I. Ruisánchez *et al.*, “Preliminary results of an interlaboratory study of chemometric software and methods on NIR data. Predicting the content of crude protein and water in forages,” *Chemometrics and Intelligent Laboratory Systems*, vol. 63, no. 2, pp. 93–105, Sep. 2002, https://doi.org/10.1016/S0169-7439(02)00039-4.

[99] L. Quintana *et al.*, “Instrumentation Evaluation for Hyperspectral Microscopy Targeting Enhanced Medical Histology,” in *2021 XXXVI Conference on Design of Circuits and Integrated Systems (DCIS)*, IEEE, Nov. 2021, pp. 1–6. https://doi.org/10.1109/DCIS53048.2021.9666188.

[100] Q. Zhang *et al.*, “An autofocus algorithm considering wavelength changes for large scale microscopic hyperspectral pathological imaging system,” *J Biophotonics*, Jan. 2022, https://doi.org/10.1002/jbio.202100366.

[101] L. Quintana, S. Ortega, H. Fabelo, and G. M. Callico, “Blur-Specific No-Reference Image Quality Assesment for Microscopic Hyperspectral Image Focus Quantification,” in *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, IEEE, Mar. 2021, pp. 1–5. https://doi.org/10.1109/WHISPERS52202.2021.9483992.

[102] J. R. Gilchrist, T. Skauli, and C. Durell, “Developing the IEEE P4001 Standard for Characterisation and Calibration of Hyperspectral Imaging Devices,” in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Jul. 2022, pp. 4679–4682. https://doi.org/10.1109/IGARSS46834.2022.9883166.

[103] S. Çayır *et al.*, “MITNET: a novel dataset and a two-stage deep learning approach for mitosis recognition in whole slide images of breast cancer tissue,” *Neural Comput Appl*, vol. 34, no. 20, pp. 17837–17851, Oct. 2022, https://doi.org/10.1007/s00521-022-07441-9.

Annex A: Resumen en Español

A.1. Motivación

El cáncer continúa siendo uno de los principales desafíos para la salud pública a nivel mundial, con cerca de 20 millones de nuevos casos y 9,7 millones de muertes registradas en 2022 [1]. Entre ellos, el cáncer de pulmón se mantiene como el más diagnosticado y el de mayor mortalidad, seguido por los cánceres de mama y colorrectal [2]. Aunque menos frecuentes, los tumores cerebrales y del sistema nervioso central presentan una elevada letalidad, atribuida en gran parte a su agresividad y complejidad [3]. Si bien los avances en la detección precoz y en los tratamientos han contribuido a reducir la mortalidad, el diagnóstico definitivo del cáncer sigue dependiendo del análisis histopatológico de biopsias, un proceso complejo y sujeto a variabilidad debido a la alta heterogeneidad de los tejidos. En los últimos años, la digitalización de la patología y la integración de técnicas de inteligencia artificial han mejorado la precisión diagnóstica y abierto nuevas oportunidades para la medicina personalizada [12], [13]. Sin embargo, las técnicas convencionales de microscopía se limitan al espectro visible (400–750 nm), lo que restringe su capacidad para identificar características bioquímicas relevantes en los tejidos, fundamentales para un diagnóstico más preciso.

La microscopía multiespectral e hiperespectral (MSI/HSI) emerge como una tecnología con gran potencial en el diagnóstico biomédico, especialmente en el ámbito oncológico [26], [63]. A diferencia de la microscopía convencional de campo claro, limitada al espectro visible y frecuentemente insuficiente para detectar características bioquímicas clave, la MSI/HSI adquiere datos en un amplio rango de longitudes de onda, lo que proporciona una visión

molecular más profunda y precisa de los tejidos. Esta información espectral ampliada mejora la reproducibilidad, la objetividad y la precisión diagnóstica de las evaluaciones histopatológicas, situando a estas tecnologías como herramientas transformadoras en patología digital. En el contexto del cáncer, la HSI ha demostrado un notable potencial diagnóstico, permitiendo diferenciar tejidos sanos y patológicos en función de sus firmas espectrales específicas. Estudios recientes han mostrado su eficacia en la identificación de tumores pancreáticos con diferentes perfiles pronósticos [57], la cuantificación de la expresión de Ki67 como marcador pronóstico en linfomas [58] y la evaluación de interacciones entre células tumorales e inmunitarias en pacientes con cáncer de pulmón tratados con inmunoterapia [59]. Asimismo, se han reportado altas precisiones diagnósticas en cánceres de mama [60], melanoma [61] y tumores de cabeza y cuello [62]. En conjunto, estos avances posicionan a la microscopía MSI/HSI como una tecnología no destructiva y complementaria a los métodos histopatológicos tradicionales, con un gran potencial para mejorar el diagnóstico oncológico. Sin embargo, a pesar de estas ventajas, la adopción clínica de la microscopía HSI sigue siendo incipiente y mayoritariamente confinada a entornos experimentales, debido a desafíos técnicos, metodológicos y normativos. Entre los principales obstáculos se encuentran la falta de estandarización en la configuración de los sistemas, los protocolos de adquisición y los procedimientos de preparación de muestras, así como la disponibilidad de bases de datos públicas que sirvan como referencia. Estos puntos limitan su uso eficaz, reproducible e integrado en la práctica clínica [26].

Esta tesis doctoral tiene como propósito adaptar de manera efectiva la tecnología HSI a la escala microscópica, con el fin de posibilitar el análisis cuantitativo, no invasivo y sin marcadores de tejidos biológicos. Para abordar los retos asociados a su implementación, se ha desarrollado un marco conceptual integral que guíe el diseño de una herramienta diagnóstica robusta, como se muestra en la Figura A1. Este marco consta de cinco etapas interrelacionadas que abordan los principales desafíos técnicos y metodológicos. (1) *Instrumentación y caracterización del sistema*, centrada en garantizar una adquisición de datos coherente y reproducible mediante un hardware fiable y calibrado (Figura A1a).

(2) *Evaluación y optimización del enfoque*, orientada a maximizar la calidad de imagen mediante el ajuste de parámetros ópticos (Figura A1b). (3) *Preparación de muestras y estandarización de protocolos*, para reducir la variabilidad en el manejo de tejidos biológicos (Figura A1c). (4) *Disponibilidad y accesibilidad de bases de datos públicas*, actualmente escasas, pero esenciales para validar y comparar metodologías (Figura A1e). (5) *Flujos de preprocesamiento específicos para HSI*, necesarios para limpiar, normalizar y preparar los datos espectrales antes del análisis (Figura A1d). Este enfoque estructurado busca mejorar la reproducibilidad de los resultados, facilitar la comparabilidad entre laboratorios y promover la adopción clínica de la microscopía HSI en el diagnóstico biomédico.

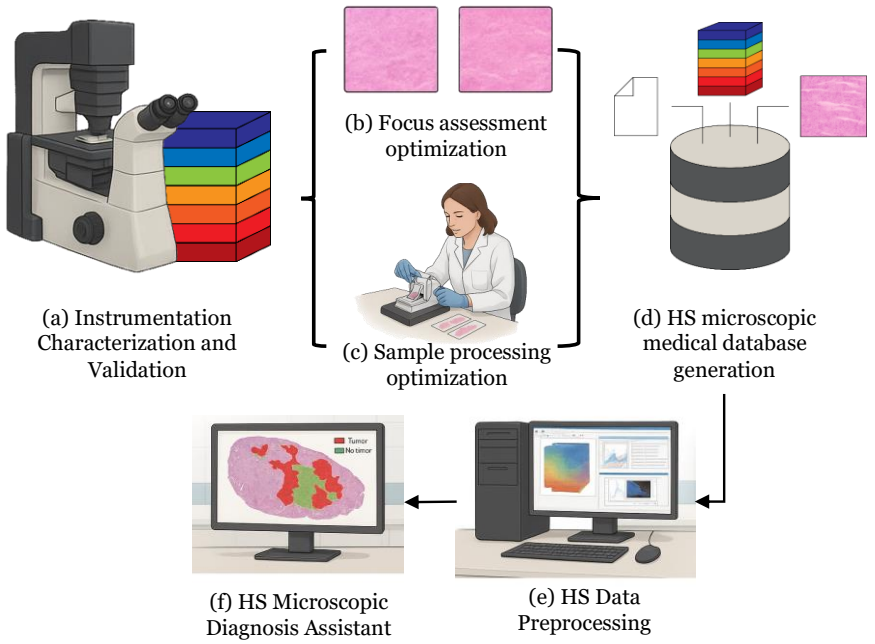


Figura A1. Marco de desarrollo de herramientas de diagnóstico microscópico HSI: (a) instrumentación, caracterización y validación, (b) optimización de la evaluación del enfoque, (c) optimización del procesamiento de muestras, (d) preprocesamiento de datos HS y (e) generación de base de datos médicas de microscopía HS.

A.2. Sistema de Microscopía Hiperespectral

HSI es una potente técnica de imagen capaz de capturar información tanto espacial como espectral, lo que permite la caracterización no invasiva de materiales, tejidos y otros elementos. Aunque HSI se utiliza ampliamente a nivel macroscópico, su aplicación en el ámbito microscópico sigue siendo limitada debido a varios retos técnicos. Una de las principales limitaciones es la ausencia de metodologías estandarizadas para caracterizar los sistemas microscópicos HS, que son esenciales para garantizar la adquisición adecuada de datos y la reproducibilidad experimental en diferentes instituciones de investigación.

El primer capítulo de esta tesis doctoral contribuye al análisis y la evaluación de los sistemas de adquisición microscópica HS adaptados a la obtención de imágenes de muestras biológicas. En concreto, su objetivo es identificar los parámetros esenciales que afectan al rendimiento del sistema, como la resolución espacial y espectral, el comportamiento del ruido y la fidelidad geométrica, y definir las configuraciones óptimas para obtener imágenes eficaces de muestras histológicas. Para alcanzar estos objetivos, se propone una hoja de ruta de caracterización exhaustiva. Esta hoja de ruta reúne los parámetros clave identificados en el estado actual de la técnica e incluye una lista detallada de los materiales necesarios para la evaluación del sistema. La metodología evalúa una amplia gama de métricas de rendimiento, entre las que se incluyen el rango dinámico, la cuantificación del ruido, el tamaño de los píxeles, la respuesta de frecuencia espacial, la precisión del escaneo espacial, la repetibilidad espacial, la corrección de campo plano, la transferencia de tonos y la sensibilidad espectral. Estas métricas se integran en un marco unificado, estandarizado y replicable que ha sido validado mediante una prueba Round Robin utilizando dos sistemas microscópicos HSI independientes (Figura A2), lo que demuestra su generalizabilidad. En consecuencia, la hoja de ruta no solo promueve la coherencia y la fiabilidad en el desarrollo y la validación de los sistemas microscópicos HSI, sino que también proporciona una orientación clara para que las futuras investigaciones informen públicamente sobre las métricas de calidad, promoviendo la

coherencia y la fiabilidad en el desarrollo y la validación de los sistemas microscópicos HSI.

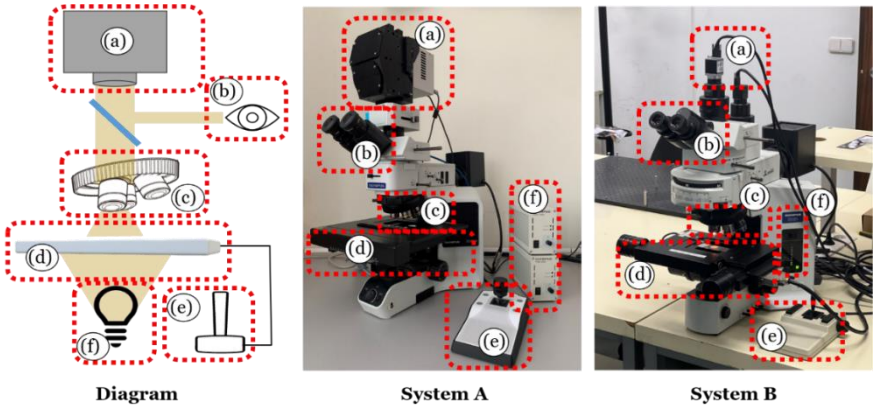


Figura A2. Diagrama de los sistemas microscópicos HS (izquierda) y ejemplo de dos sistemas microscópicos HS (sistema A en el centro y sistema B a la derecha): (a) cámara HS, (b) prismáticos, (c) lentes objetivo, (d) platina portaobjetos, (e) joystick para el escaneo espacial y (f) fuente de luz halógena.

Los resultados preliminares se presentaron en el artículo de la conferencia (**C1**) titulado «**Instrumentation Evaluation for Hyperspectral Microscopy Targeting Enhanced Medical Histology**» (Evaluación de la instrumentación para la microscopía hiperespectral destinada a mejorar la histología médica) en la *XXXVI Conferencia sobre Diseño de Circuitos y Sistemas Integrados (DCIS)*, celebrada en Portugal en 2021. Este trabajo presentó los principales retos asociados a la instrumentación microscópica HS. Posteriormente, la metodología completa se desarrolló y probó mediante un estudio colaborativo con el Centro de Investigación en Tecnologías de Software y Sistemas Multimedia de la Universidad Politécnica de Madrid, durante la estancia de investigación de la doctoranda (**M1**). La hoja de ruta propuesta se validó en dos sistemas microscópicos HS representativos, uno ubicado en la Universidad de Las Palmas de Gran Canaria y otro en la Universidad Politécnica de Madrid (Figura A1). Ambos sistemas integran un sensor HSI en un microscopio de campo claro, una configuración comúnmente utilizada en la obtención de imágenes biomédicas. Los resultados de este trabajo se publicaron en el artículo de revista (**0**) titulado «**Roadmap for the Design and Characterization of Microscopic Hyperspectral Systems**»

(Hoja de ruta para el diseño y la caracterización de sistemas microscópicos hiperespectrales), en *IEEE Transactions on Instrumentation and Measurement*. En general, la hoja de ruta de caracterización propuesta proporciona directrices prácticas y detalladas para la comunidad científica que desea desarrollar y evaluar sistemas microscópicos HS fiables, eficientes y precisos.

A.3. Optimización de la Calidad de la Imagen HS

Una vez que los sistemas microscópicos HS se han desarrollado y caracterizado adecuadamente, la atención debe centrarse en la calidad y la fidelidad de los datos que producen. En este contexto, surgen dos áreas críticas: la optimización de la calidad de los datos durante la adquisición y a través del procesamiento posterior. Durante la captura de imágenes, el operador suele ajustar el enfoque manualmente utilizando únicamente la inspección visual. Este enfoque a menudo da lugar a una nitidez de imagen no uniforme, especialmente dadas las condiciones de poca luz inherentes a la microscopía HSI y la naturaleza heterogénea de las muestras biológicas. Tras la adquisición, es esencial utilizar técnicas de procesamiento específicas para mitigar el ruido y los artefactos, mejorando así la calidad de los datos y garantizando un análisis espectral robusto y preciso.

El segundo capítulo de esta tesis doctoral se centra en el análisis de los factores que afectan al enfoque de la imagen, la iluminación y la calidad de la señal antes de la adquisición de datos, desarrollando estrategias y protocolos que permitan una captura de imágenes consistente y reproducible. Uno de los principales desafíos abordados es el desenfoque en imágenes microscópicas HS, estrechamente relacionado con la profundidad de campo (DOF) del sistema, definida como la distancia entre los puntos más cercanos y más lejanos en los que un objeto puede registrarse con una nitidez aceptable (Figura A3). En este contexto, se lleva a cabo una evaluación exhaustiva de las limitaciones del enfoque manual, así como del impacto de la DOF, y se exploran métodos avanzados orientados a mejorar la uniformidad y fiabilidad del enfoque en todo

el campo de visión. Esto incluye la evaluación de la calidad del enfoque mediante algoritmos de evaluación de la calidad de la imagen sin referencia (NR-IQA) adaptados de datos RGB a HS. Estos algoritmos son evaluados en una base de datos generada, HIDFA (Hyperspectral Imaging Dataset for Focus Assessment), que contiene 125 fotogramas de referencia capturados a 11 distancias de trabajo (1375 cubos HSI en total). Los resultados indicaron que, mientras que algunos algoritmos NR-IQA mostraban respuestas fluctuantes en diferentes bandas espectrales (MLV, FISHbb, LPC), otros proporcionaban puntuaciones consistentes independientemente de las variaciones de longitud de onda (CPBD, EMBM, PSI, SPARISH). Esta consistencia es especialmente importante en la HSI microscópica, donde todas las bandas espectrales corresponden al mismo plano focal del microscopio. Por lo tanto, se identificaron como los más robustos aquellos algoritmos con un rendimiento estable en todo el espectro. Estos hallazgos sientan las bases para futuras aplicaciones en la corrección del enfoque, como la selección automatizada de las regiones más nítidas en los procedimientos de apilamiento z. Estos hallazgos subrayan tanto la relevancia como la complejidad de adaptar los métodos NR-IQA a las imágenes HS.

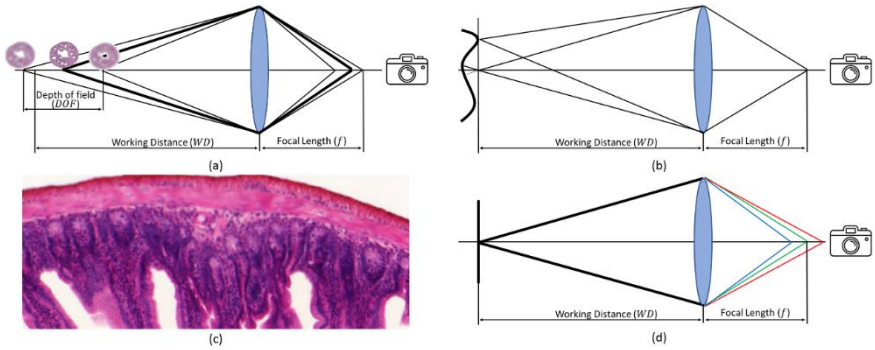


Figura A3. Representación gráfica de (a) DOF, (b) desplazamiento WD, (c) Ejemplo de un intestino de rata desenfocado (20×) y (d) problemas de aberración cromática.

Los resultados preliminares de este trabajo se presentaron en el artículo de conferencia (**C4**) titulado «**Blur-Specific No-Reference Image Quality Assesment for Microscopic Hyperspectral Image Focus Quantification**» (Evaluación de la calidad de imagen sin referencia específica para el desenfoque para

la cuantificación del enfoque de imágenes hiperespectrales microscópicas), presentado en el *11.º Taller sobre Imágenes Hiperespectrales y Procesamiento de Señales: Evolución en la Teledetección (WHISPERS)*, celebrado en Ámsterdam en 2021. En él se presentaron los retos relacionados con la nitidez de la imagen y el preprocesamiento de datos en microscopía HSI. Seguidamente, se llevó a cabo una investigación más detallada que se publicó en el artículo de revista (**J2**) titulado «**Blur-specific Image Quality Assessment of Microscopic Hyperspectral Images**» (Evaluación de la calidad de imagen específica para el desenfoque de imágenes hiperespectrales microscópicas) y publicado en *Optics Express*. Junto con este trabajo se publicó una base de datos pública, HIDFA, para la cuantificación del enfoque, de modo que otros investigadores puedan beneficiarse de ella.

Otro objetivo que también se aborda en este capítulo es mejorar la calidad de los datos microscópicos HS tras su adquisición, a través de varias técnicas de procesamiento de imágenes. En concreto, se centra en evaluar métodos eficaces para la reducción del ruido, la corrección de artefactos y la mejora espectral, mejorando así la fidelidad espectral y facilitando un análisis preciso. Se evaluaron sistemáticamente una serie de estrategias de preprocesamiento, incluidos los filtros de suavizado, la normalización de la variable normal estándar (SNV) y los cálculos de derivadas espectrales, utilizando el error cuadrático medio (RMSE) frente a un estándar de referencia de longitud de onda conocido. También se tuvo en cuenta la eficiencia del procesamiento para equilibrar la mejora de la calidad con la aplicabilidad práctica. Los resultados demostraron que métodos individuales como el suavizado de media móvil, la SNV y las primeras derivadas espectrales mejoraban significativamente la calidad de los datos, y que su aplicación combinada producía mejoras aún mayores. Las contribuciones relevantes se presentaron en el artículo de la conferencia (C1) titulado «**Assessing Processing Strategies on Data from Medical Hyperspectral Acquisition Systems**» (Evaluación de estrategias de procesamiento de datos de sistemas de adquisición hiperespectral médica), presentado en la *27.ª Conferencia Euromicro sobre Diseño de Sistemas Digitales (DSD)*, celebrada en París en 2024.

En conjunto, este capítulo proporciona un marco completo para mejorar la fidelidad de los datos en la obtención de imágenes microscópicas HS. Al abordar los retos tanto del momento de la adquisición como de la post-adquisición, contribuye al desarrollo de metodologías robustas que promueven resultados más precisos, consistentes e interpretables en la obtención de imágenes biomédicas y más allá.

A.4. Optimización de la Calidad de la Muestra Histológica

Si bien los avances recientes en microscopía HS se han centrado principalmente en el desarrollo de sensores y sistemas, otro factor crítico que influye en la calidad de la imagen es la preparación de las propias muestras biológicas. El análisis histológico de muestras de biopsia es actualmente el método de referencia para la detección y clasificación de tumores, y se basa en imágenes de alta resolución para distinguir las estructuras celulares. Para lograr una nitidez óptima de la imagen en los microscopios WSI, las secciones de tejido se cortan tradicionalmente en láminas finas, normalmente de 3 a 5 μm de grosor, y se tiñen con colorantes que mejoran el contraste (como la hematoxilina y eosina, H&E). Sin embargo, en la microscopía HS, la obtención de una transmitancia difusa de una muestra, basada en los principios de la interacción entre el tejido y la luz (Figura A4), puede requerir diferentes técnicas de preparación de muestras para adquirir información espectral valiosa de los tejidos biológicos.

El cuarto capítulo de esta tesis se centra en investigar cómo la preparación de muestras biológicas, concretamente el grosor del tejido afecta a las características espaciales y espectrales de las imágenes microscópicas HS, y en proponer directrices que mejoren la calidad de la imagen y la fiabilidad del diagnóstico. Para cumplir estos objetivos se siguió un enfoque multifásico. En primer lugar, se realizaron imágenes experimentales en secciones de tejido mamario teñidas con H&E y preparadas con un grosor de 2 μm y 3 μm. Basándose en los picos de las firmas espectrales de la hematoxilina y la eosina, las imágenes HS se segmentaron en regiones de estroma

(teñidas con eosina), núcleos (teñidos con hematoxilina) y fondo (sin teñir). Los resultados mostraron que las muestras de 3 µm no solo proporcionaban un mayor número de células captadas en imagen, sino que también presentaban un mejor contraste espectral, lo que sugiere una mayor interacción entre el tejido y la luz y una mejor discriminación de la estructura en los datos HS en comparación con las secciones más finas.

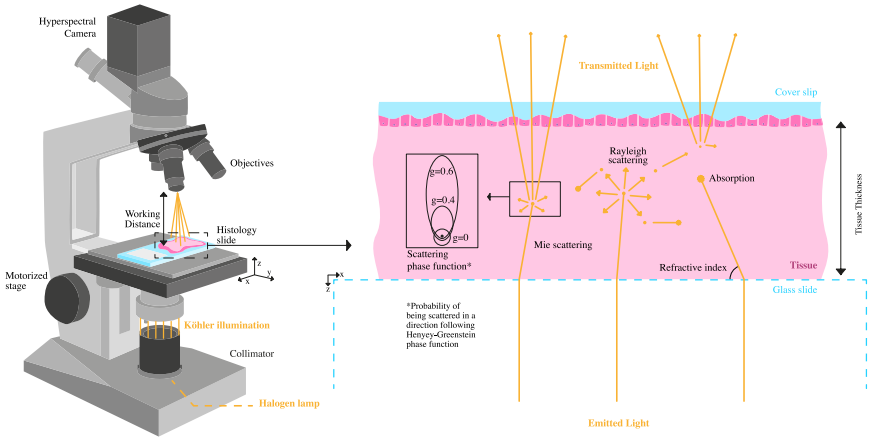


Figura A4. Esquema del sistema microscópico HS capturando una lámina histológica en modo de transmisión (izquierda) y las interacciones entre la luz y el tejido dentro del tejido que se está analizando (derecha).

En una segunda fase, se realizó un estudio para explorar el grosor de las muestras, más allá de los protocolos histológicos convencionales, mediante análisis computacional. Una revisión de la literatura reveló que aproximadamente el 40 % de los estudios de microscopía HS no informan del grosor del tejido y, de los que lo hacen, el 98 % utiliza cortes de entre 2 y 10 µm. Para abordar esta laguna, se llevó a cabo un conjunto de simulaciones de transporte de luz Monte Carlo en varios tipos de tejidos (mama, hígado, pulmón y colorrectal) para evaluar la influencia del grosor del tejido no teñido en la calidad de las imágenes HS. Las simulaciones demostraron que las muestras delgadas producen una discriminación espectral deficiente debido a la insuficiente interacción entre la luz y el tejido, mientras que las muestras más gruesas, que se acercan a los 500 µm, permiten una mejor clasificación del tejido a expensas de la intensidad de la luz.

El primer enfoque se presentó en el artículo de la conferencia (**C2**) titulado «**Analysing Histology Hyperspectral Images, Does Tissue Thickness Matter**» (Análisis de imágenes hiperespectrales histológicas, ¿importa el grosor del tejido?), presentado en la conferencia *SPIE Photonics Europe* celebrada en Estrasburgo en 2024. En él se introdujo el impacto del grosor de la muestra en la calidad de la información espectral y espacial. El segundo enfoque se desarrolló posteriormente durante la estancia de investigación de la doctoranda en el Departamento de Cirugía Guiada por Imágenes (**M3**) del Instituto Neerlandés del Cáncer, en colaboración con expertos en óptica biomédica e histología del cáncer. Los resultados completos se publicaron en el artículo (**J3**) titulado «**Exploring the Role of Sample Thickness for Hyperspectral Microscopy Tissue Discrimination through Monte Carlo Simulations**» (Exploración del papel del grosor de la muestra en la discriminación de tejidos mediante microscopía hiperespectral a través de simulaciones de Monte Carlo), en *Biomedical Optics Express*.

Este capítulo subraya la importancia de la investigación interdisciplinaria en el desarrollo de datos microscópicos HS fiables, demostrando que los avances técnicos en sensores y algoritmos deben ir acompañados de una preparación optimizada de las muestras biológicas. Al abordar la variable del grosor del tejido, a menudo pasada por alto, este trabajo contribuye al establecimiento de protocolos mejorados que promueven imágenes HS precisas, reproducibles y clínicamente relevantes para el diagnóstico médico.

A.5. Bases de datos médicas microscópicas HS

A pesar del creciente interés y del rápido progreso tecnológico en microscopía HS, sigue existiendo un reto importante: la escasez de conjuntos de datos microscópicos HS disponibles públicamente y bien anotados. Esta carencia dificulta considerablemente la capacidad de los investigadores para evaluar y comparar objetivamente el rendimiento de los métodos de procesamiento avanzados, como la clasificación de imágenes HS, el *spectral*

unmixing (separación espectral) y la compresión de datos. Sin conjuntos de datos públicos y estandarizados, la evaluación comparativa se vuelve inconsistente, lo que dificulta la reproducibilidad y complica el progreso colectivo en este campo. El capítulo 5 de esta tesis doctoral está dedicado a la adquisición de imágenes HS de muestras histológicas en diferentes condiciones de preparación y obtención de imágenes. El objetivo es crear conjuntos de datos representativos que reflejen las prácticas reales de los laboratorios y permitan una investigación reproducible y comparable. Con este fin, se desarrollaron dos bases de datos completas, cada una de ellas adaptada para satisfacer diferentes necesidades clínicas y de investigación en el campo de la microscopía HS.

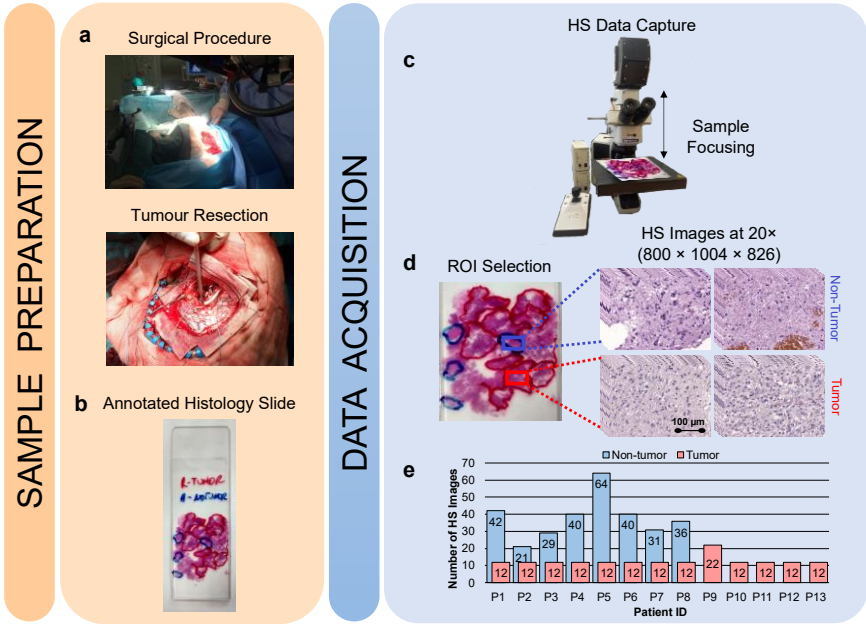


Figura A5. Resumen gráfico de la metodología seguida. (a) Procedimiento de resección. (b) Anotaciones macroscópicas de las ubicaciones de la GB. (c) Captura de datos HS utilizando un sistema microscópico HS. (d) Selección de ROI. (e) Resumen del conjunto de datos.

La primera base de datos (**J3**), «**Histological Hyperspectral Glioblastoma Dataset (HistologyHSI-GB)**», se creó en la Universidad de Las Palmas de Gran Canaria y se publicó en el repositorio *The Cancer Imaging Archive* y en la revista *Scientific*

Data. Este conjunto de datos contiene 469 imágenes hiperespectrales obtenidas de 13 pacientes diagnosticados con glioblastoma, un tumor cerebral muy agresivo. Las secciones de tejido teñidas con H&E se capturaron con un aumento de 20× utilizando un sistema de microscopía HS personalizado con una sensibilidad espectral que abarca de 400 a 1000 nm. Patólogos expertos proporcionaron anotaciones a nivel de imagen, lo que garantizó la relevancia clínica del conjunto de datos. La Figura A5 muestra un resumen gráfico del mismo. Al capturar las características espectrales y espaciales del tejido, este conjunto de datos ofrece un rico recurso para estudiar la heterogeneidad de tumores de cerebro y desarrollar métodos de análisis de imágenes HS con aplicación en neurooncología.

El segundo conjunto de datos se creó gracias a la colaboración con el Departamento de Patología del Hospital de Tortosa Verge de la Cinta, en Tortosa (España). Durante su estancia de investigación (**M1**), la doctoranda contribuyó a la adquisición y curación de este conjunto de datos de tejido mamario, centrado en casos de recurrencia. Esta base de datos (**J5**), «**Histological Hyperspectral Breast Cancer Recurrence Dataset (HistologyHSI-BC Recurrence)**» (Conjunto de datos histológicos hiperespectrales sobre la recurrencia del cáncer de mama (HistologyHSI-BC Recurrence), que se está publicando actualmente en el repositorio *The Cancer Imaging Archive* y en la revista *Scientific Data*, ofrece valiosos datos de imágenes para estudiar los retos diagnósticos asociados a los tumores recurrentes de mama. Se diseñó específicamente para avanzar en la predicción de la recurrencia del cáncer de mama a distancia. Integra 677 imágenes HS, 47 imágenes histopatológicas de láminas completas y datos clínicos y demográficos detallados de 47 pacientes con cáncer de mama, de las cuales casi la mitad desarrollaron recurrencia durante un período de 12 años. Las láminas histológicas se digitalizaron mediante escaneo de láminas completas y se anotaron por expertos patólogos, mientras que las imágenes HS se adquirieron utilizando un microscopio HS. Se puede consultar el resumen gráfico de todo el conjunto de datos en la Figura A6. Este enfoque multimodal ofrece oportunidades sin precedentes para identificar biomarcadores pronósticos mediante la combinación de

datos cínicos con imágenes histopatológicas y HS, lo que respalda las estrategias de tratamiento personalizadas.

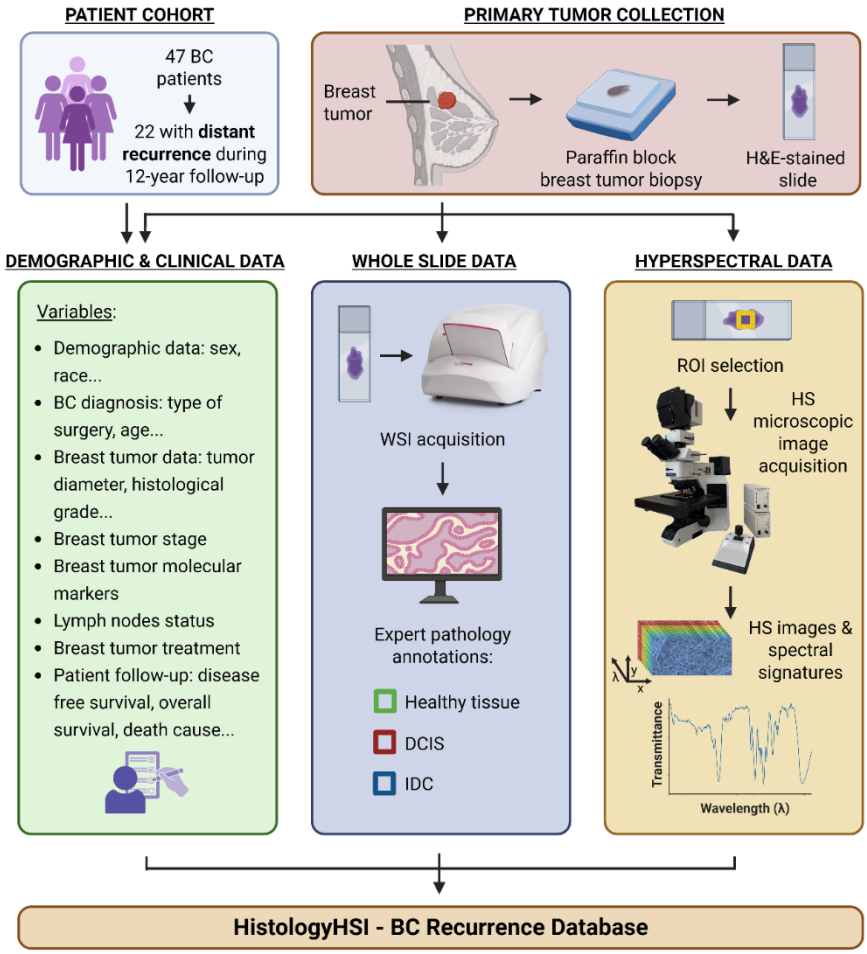


Figura A6. Se obtuvieron biopsias de tumor mamario de 47 pacientes, que fueron procesadas y teñidas con H&E. Las muestras se digitalizaron (WSI), se anotaron y se capturaron imágenes hiperespectrales. Junto con los datos clínicos y demográficos, se creó la base de datos *HistologyHSI - BC Recurrence Database*.

El contenido de este capítulo ofrece importantes contribuciones a la comunidad científica al publicar dos conjuntos de datos microscópicos HS disponibles públicamente, lo que conducirá a avances tanto en la investigación como en el uso clínico de las imágenes microscópicas HS. Las bases de datos *HistologyHSI-GB* e *HistologyHSI-BC Recurrence* proporcionan recursos valiosos y clínicamente relevantes que respaldan el desarrollo de algoritmos y

la traducción clínica. Al abordar la variabilidad de los datos y promover la estandarización, este trabajo ayuda a acelerar la reproducibilidad, la colaboración y la adopción de la microscopía HS en la práctica médica habitual.

A.6. Conclusiones, Limitaciones y Trabajo Futuro

Esta tesis doctoral contribuye de manera significativa al avance de la viabilidad clínica de la microscopía HS en patología digital al abordar de manera sistemática sus principales retos técnicos, metodológicos y prácticos. El trabajo abarca cinco ámbitos principales:

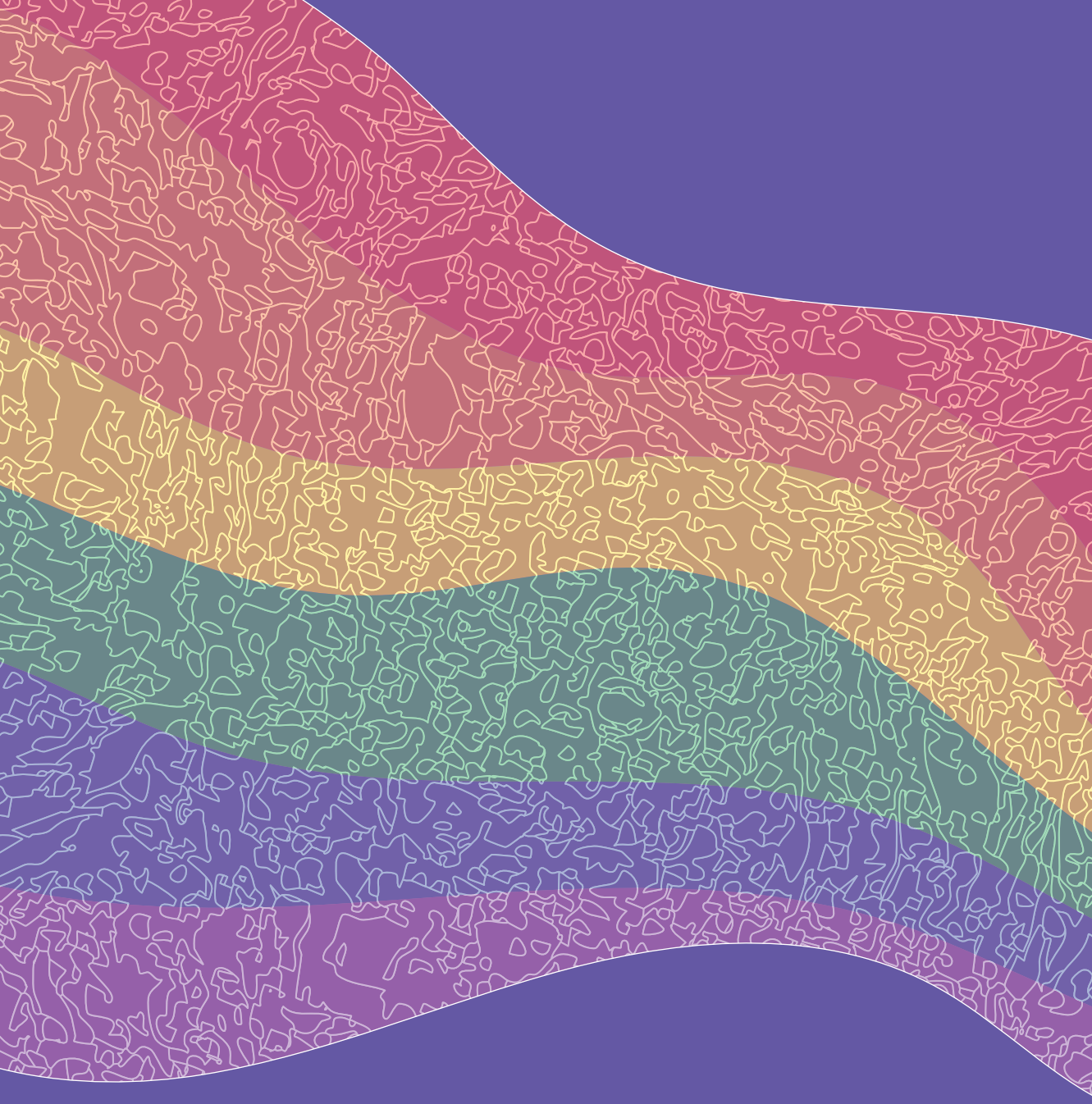
- Estandarización del diseño y la caracterización de los sistemas de microscopía HS: se ha desarrollado una hoja de ruta sólida y validada empíricamente para la calibración y la evaluación del rendimiento de los sistemas microscópicos HS. Esto incluye parámetros críticos, pero a menudo descuidados, como la repetibilidad espacial y la resolución dependiente de la longitud de onda. La introducción de este marco favorece la reproducibilidad y la interoperabilidad entre dispositivos, y sienta las bases para futuros esfuerzos de estandarización en HSI biomédica.
- Establecimiento de técnicas objetivas de evaluación del enfoque sensibles al espectro: la tesis presenta un conjunto de datos de referencia único (*HIDFA*) y evalúa una serie de algoritmos de evaluación del enfoque adaptados a la microscopía HS. Los resultados enfatizan la importancia de tener en cuenta la resolución espacial dependiente de la longitud de onda y la variabilidad estructural, lo que orienta la selección de métricas de calidad de imagen más informativas y específicas para HS. Estos conocimientos mejoran los protocolos de garantía de calidad y ayudan a garantizar una imagen diagnóstica fiable.
- Optimización del grosor del tejido para maximizar la discriminabilidad espectral: Se implementó un novedoso modelo de simulación del transporte de luz a través de tejido

biológico basado en Monte Carlo para explorar la influencia del grosor del tejido en el contraste espectral. El trabajo desafía las prácticas histológicas convencionales y demuestra que un rendimiento diagnóstico óptimo requiere grosores personalizados en función del tipo de tejido y el objetivo diagnóstico. Esta contribución fomenta una reevaluación de los protocolos de preparación de muestras para la obtención de imágenes HS.

- Marco para el preprocesamiento espectral estandarizado en HSI médica: mediante la introducción de un proceso de evaluación del preprocesamiento independiente del sistema, la tesis proporciona directrices prácticas para mejorar la fidelidad espectral en diferentes plataformas HS. Promueve la adopción de estrategias de preprocesamiento coherentes (por ejemplo, normalización SNV, filtros de suavizado adecuados) y fomenta el desarrollo de algoritmos con capacidad en tiempo real para su implementación clínica.
- Desarrollo de bases de datos microscópicas HS de acceso público y orientación clínica: se presentan dos bases de datos de referencia, *HistologyHSI-GB* e *HistologyHSI-BC Recurrence*, que recogen la variabilidad real en la preparación de muestras y las condiciones de obtención de imágenes. Estos recursos permiten el desarrollo de algoritmos reproducibles en neurooncología y en la predicción de la recurrencia del cáncer de mama, promueven la colaboración entre investigadores computacionales y clínicos, y sientan las bases para avanzar en el uso clínico de las imágenes HS.

Hacia la traducción clínica de la microscopía HS, esta tesis doctoral ha abordado retos clave en el proceso de adquisición y análisis de imágenes hiperespectrales microscópicas aplicadas a la patología. Desde la caracterización de sistemas y evaluación del enfoque, hasta la preparación de muestras, el preprocesamiento espectral y la creación de bases de datos públicas, este trabajo sienta unas bases metodológicas sólidas para avanzar hacia aplicaciones clínicas. Sin embargo, aún se identifican desafíos que orientan futuras líneas de investigación. La hoja de ruta propuesta para evaluar sistemas HS debe extenderse y adoptarse de forma más generalizada, integrando parámetros adicionales como la polarización o las aberraciones ópticas. Las métricas de calidad de

enfoque sin referencia han de evolucionar hacia escenarios dinámicos, como el apilamiento en profundidad, con el objetivo de generar cubos espectrales todo-en-foco que resulten clínicamente útiles. La simulación del efecto del grosor tisular ofrece un marco conceptual valioso, pero demanda validación empírica con muestras reales y adaptación de protocolos histológicos a la práctica clínica. En cuanto al procesamiento de datos, se requiere el perfeccionamiento de técnicas de suavizado y corrección espectral, así como su implementación en tiempo real mediante soluciones aceleradas por hardware. La inteligencia artificial aplicada a estos datos debe adaptarse a la variabilidad inherente de los tejidos y condiciones de adquisición, promoviendo estrategias de aumento de datos y marcos de evaluación comparativa robustos. Además, la escasez de bases de datos bien anotadas continúa siendo una barrera crítica, lo que subraya la necesidad de establecer protocolos de anotación eficientes, estrategias de compresión y plataformas colaborativas para su creación y compartición. En definitiva, la validación de estas metodologías en entornos clínicos reales será el paso decisivo para facilitar la adopción de la microscopía HS como herramienta de diagnóstico de próxima generación.



ULPGC
Universidad de
Las Palmas de
Gran Canaria