Epithelial Ovarian Cancer Diagnosis of Second-Harmonic Generation Images: A Semiautomatic Collagen Fibers Quantification Protocol

Cancer Informatics
1–12

DOI: 10.1177/1176935117690162
SAGE

Angel A Zeitoune[1,2], Johana SJ Luna[3], Kynthia Sanchez Salas[3], Luciana Erbes[1,2], Carlos L Cesar[4,5], Liliana ALA Andrade[6], Hernades F Carvahlo[4,7], Fátima Bottcher-Luiz[4,8], Victor H Casco[2] and Javier Adur[1,2,3]

[1]Biofotónica y Procesamiento de Información Biológica (ByPIB), Centro de Investigación y Transferencia de Entre Ríos (CITER), CONICET-UNER, Entre Ríos, Argentina. [2]Microscopy Laboratory Applied to Molecular and Cellular Studies, Engineering School, National University of Entre Ríos, Entre Ríos, Argentina. [3]Laboratory Applied to Non-Ionizing Radiation, Engineering School, National University of Entre Ríos, Entre Ríos, Argentina. [4]National Institute of Science and Technology on Photonics Applied to Cell Biology (INFABiC), São Paulo, Brazil. [5]Department of Physics, Federal University of Ceará (UFC), Fortaleza, Brazil. [6]Department of Obstetrics and Gynecology, Faculty of Medical Sciences, State University of Campinas (UNICAMP), São Paulo, Brazil. [7]Department of Structural and Functional Biology, Biology Institute, State University of Campinas (UNICAMP), São Paulo, Brazil. [8]Department of Pathology of the Faculty of Medical Sciences, State University of Campinas (UNICAMP), São Paulo, Brazil.

ABSTRACT: A vast number of human pathologic conditions are directly or indirectly related to tissular collagen structure remodeling. The nonlinear optical microscopy second-harmonic generation has become a powerful tool for imaging biological tissues with anisotropic hyperpolarized structures, such as collagen. During the past years, several quantification methods to analyze and evaluate these images have been developed. However, automated or semiautomated solutions are necessary to ensure objectivity and reproducibility of such analysis. This work describes automation and improvement methods for calculating the anisotropy (using fast Fourier transform analysis and the gray-level co-occurrence matrix). These were applied to analyze biopsy samples of human ovarian epithelial cancer at different stages of malignancy (mucinous, serous, mixed, and endometrial subtypes). The semiautomation procedure enabled us to design a diagnostic protocol that recognizes between healthy and pathologic tissues, as well as between different tumor types.

KEYWORDS: FFT, GLCM, SHG microscopy, human cancer ovary, diagnostic protocol

RECEIVED: September 26, 2016. **ACCEPTED:** January 2, 2017.

PEER REVIEW: Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 1185 words, excluding any confidential comments to the academic editor.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by CEPOF (Optics and Photonics Research Center, FAPESP), INFABIC (National Institute of Photonics Applied to Cell Biology, FAPESP, and CNPq), and PIO CONICET-UNER (Res: 4337/15; No. 14620140100004 CO).

DECLARATION OF CONFLICTING INTEREST: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Javier Adur, Biofotónica y Procesamiento de Información Biológica (ByPIB), Centro de Investigación y Transferencia de Entre Ríos (CITER), CONICET-UNER, Ruta 11, Km 10, E3100 Oro Verde, Entre Ríos, Argentina. Email: jadur@ingenieria.uner.edu.ar

Introduction

Numerous human pathologic conditions are directly or indirectly related to the remodeling of the collagen structure. Quantitative studies have shown structural differences between the collagen fibers of normal and pathologic tissues. In pathologic samples, collagen fibers exhibit a disordered pattern, whereas normal extracellular matrixes have a well-defined order.[1] Therefore, a detailed characterization of this tissue compartment is of great importance for classifying the state of an organ.

The nonlinear optical microscopy second-harmonic generation (SHG) is a powerful technique for imaging biological tissues containing anisotropic structures that are highly polarized, such as collagen. Second-harmonic generation images provide valuable morphologic and functional information about tissue because it possesses both good depth penetration and spatial resolution.[2,3] During the past years, SHG has already been largely used for imaging collagen-rich tissues such as cornea, tendon, and arteries.[4–7] Second-harmonic generation has been mainly used to selectively investigate collagen fiber orientation and their structural changes in human dermis, keloid, fibrosis, thermally treated samples, and also in tumor microenvironments.[7–11] Because collagen matrix remodeling is a relevant factor in carcinogenesis, preliminary results showed that analyses of samples under microscopic techniques, such as SHG, combined with different quantization methods could be a potential diagnosis tool.[12]

There are several quantification methods to analyze and evaluate SHG images. The most studied methods are as follows: tumor-associated collagen signatures (TACS), which measures the inclination of the collagen fibers from the boundary of the tumor; anisotropy study through fast Fourier transform (FFT) analysis; and the gray-level co-occurrence matrix (GLCM).[1,5–7,13–17]

The FFT is an important tool in image processing. Each point of the image in Fourier domain represents a frequency contained in the space domain image. Images formed by pure sinusoid have a 2-dimensional Fourier transform represented

by a point. This information is useful to characterize images with high alignment degree and with a predominant orientation as collagen fibers in normal samples. For regularly aligned fiber structures, the expected FFTs exhibit high values around the orthogonal axis to the fibers, acquiring an elliptical distribution. Conversely, randomly arranged fibers are expected to possess circular FFT images.[1,5,7,18,19]

Texture is an important feature used in the identification of objects or regions of interest (ROIs) in an image. Gray-level co-occurrence matrix is one of the most used methods to calculate the texture. It is based on second-order statistics,[1,5,7,15,18,20,21] determining the texture via measuring the gray level of pixels at different orientations from an individual pixel or group of pixels. The gray levels of these neighbors are recorded in a matrix, in which mathematical relationships can be used to describe that image parameter. Different textural measures (Haralick texture descriptors) are obtained based on GLCM calculation: contrast, energy, entropy, and correlation that provide information about different image textural characteristics.[22] The contrast estimates the amount of local variations in an image, whereas energy is the sum of squared elements in GLCM. Energy may also be referred as uniformity or the angular second moment. The entropy descriptor is a measurement of randomness in the image, being of higher entropy values, indication of greater texture complexity, or stochasticity. Finally, correlation allows the assessment of correspondence of neighbor pixels over the whole image.

These methods have been separately used in different studies with dissimilar implementation methods and mainly in animal cancer models.[1,20,21] Recently, we have used some of these methods to analyze human ovarian cancer SHG images.[5,18] In those previous studies, serous and mucinous ovarian tumors were manually analyzed, calculating TACS, FFT, and correlation texture features from the GLCM. Even though important differences were detected between tumor types, the quantification steps were tedious, time-consuming, and even prone to subjective bias. For example, in the previous studies, in the FFT procedure, the essential ellipse determination should be performed for each ROI, based on the operator criterion. In the GLCM calculation case, the step size (in pixels) and step direction (0°, 45°, 90°, and 135°) should be manually selected each time in the ImageJ plugin process. Thus, in the original plugin version, the same procedure must be successively executed for each ROI (from 1 to 12 pixels), in each step size and step direction (0°, 45°, 90°, and 135°). So, the plugin run each time and for each selection data must be saved in a table, which in turn must be copied to another file because in each run the table is overwritten.

Because of this, to ensure objectivity and reproducibility of the results are a need for automatic solutions, this work describes a semiautomation and improvement techniques for calculating the anisotropy using FFT and GLCM. They are applied to biopsy samples of human ovarian epithelial cancer at different degrees of malignancy (mucinous, serous, mixed, and endometrial). This procedure allows the design of a diagnostic protocol that discriminates between healthy and pathologic tissues, as well as between different tumor types.

Nevertheless, this preliminary approach requires to increase the biopsy number to be sure that it may be safely used for clinical diagnosis of ovarian malignancies.

Materials and Methods

Ovarian samples and images acquisition

The ovarian samples were provided by the Hospital of University Women (CAISM), Campinas, SP, Brazil. The research protocol was approved by the Ethics Committee (Institutional Ethics Committee of the Medical Sciences Faculty, Unicamp). All procedures were performed in accordance with the Declaration of Helsinki and the ethical principles of the medical community. Ovarian tumors were fixed in 10% formalin, dehydrated, embedded in paraffin, and sectioned to a thickness of 4 μm. Sections were dewaxed, rehydrated, and stained with hematoxylin-eosin (H&E). Subsequently, samples were examined using the SHG technique. Each H&E-stained section was classified by experienced pathologists on the basis of histologic criteria established by the World Health Organization.

The images were acquired with an inverted Axio Observer microscope Z.1 equipped with a confocal scanning head Zeiss LSM 780 NLO (Carl Zeiss AG, Jena, Germany). The excitation beam was provided by a pulsed laser Ti:Sapphire (Spectra-Physics, Irvine, CA, USA) under the same setup previously described.[5] A set of 96 images of normal tissues and subtypes of epithelial ovarian cancer (endometrial, serous, mucinous, and mixed) were used (see Table 1 and Figure 1 for representative images of H&E and SHG of normal ovarian epithelium and cancer tissues). The images of H&E were used as a criterion standard and for diagnostic purposes. Second-harmonic generation images were used for collagen identification and quantification over specific ROI.

Development and standardization of automatic quantification tools

Collagen anisotropy and geometric computations were performed over specific ROIs in SHG images. Over each image, 4 ROIs of 120 × 120 pixels (41.54 μm × 41.54 μm) were manually selected by the operator (Figure 1). These regions were located immediately below ovarian epithelium to make sure that mainly collagen network in the proximity of it was analyzed because this is the region responsible for stromal invasion.[5] The selection of ROIs (containing 13-41 fiber packages) was done to allow image sampling of complex epithelial surfaces. Image processing software Octave and ImageJ were used for the processing of available image bank.

Table 1. Ovary samples.

TISSUE TYPE		NUMBER OF PATIENTS	SAMPLES PER PATIENT	TOTAL
Normal		4	3	12
Endometrial		4	3	12
Mucinous	Adenoma	4	3	12
	Borderline	1	3	3
	Adenocarcinoma	4	3	12
Serous	Adenoma	4	3	12
	Borderline	3	3	9
	Adenocarcinoma	4	3	12
Mixed		4	3	12
			Total sample	96

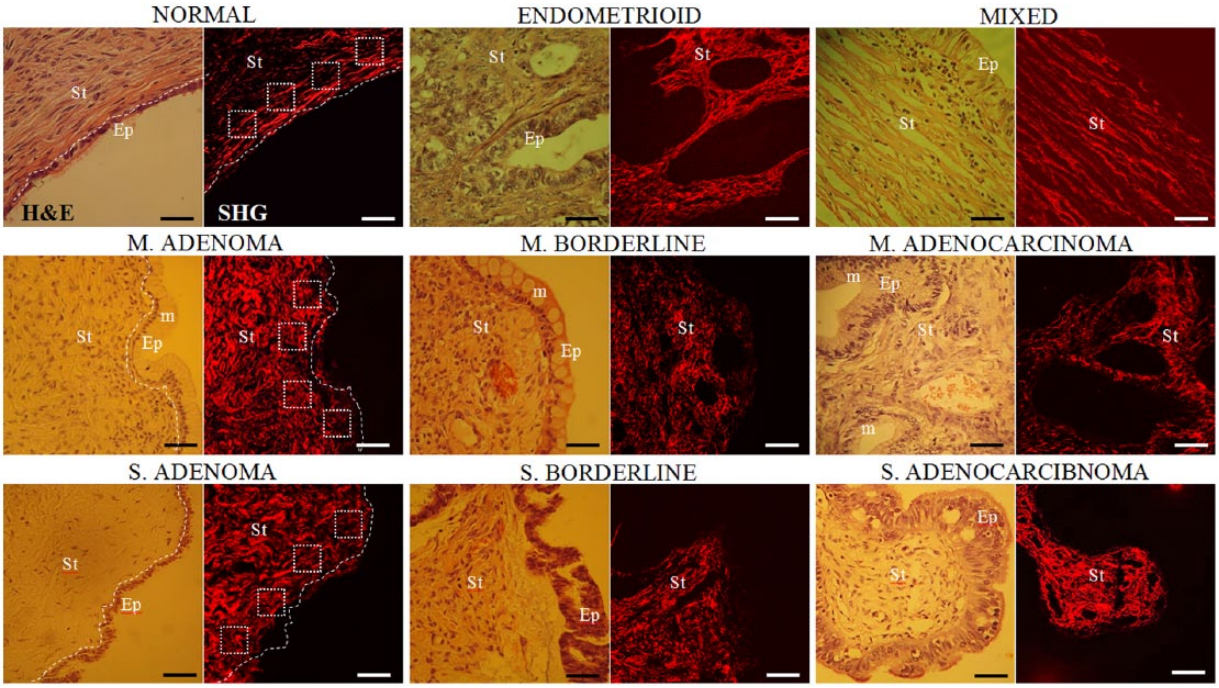

Figure 1. Representative bright-field hematoxylin-eosin (H&E) and second-harmonic generation (SHG) images (1024 × 1024 pixels) at a 2-photon excitation wavelength of 940 nm for normal and pathologic human ovary tissues. Boundary epithelium (white outline) and 4 different regions of interest (ROIs) (120 × 120 pixels, white squared) in which fast Fourier transform analysis and gray-level co-occurrence matrix were performed are shown. For simplicity, boundary and ROIs are only represented in the first column images. Ep indicates epithelium; m, mucin; M, mucinous; S, serous; St, stroma. All scale bars = 40 µm. 40X/1.30 NA oil immersion objective (EC Plan-Neofluar Carl Zeiss AG, Jena, Germany), was used.

Collagen anisotropy quantification using FFT. The anisotropy of images was calculated using the aspect ratio (AR) between the major and minor axes of the ellipse resulting from the 2-dimensional Fourier transforms. The transformed image from each ROI was binarized through threshold, which was dynamically established to a value of 125% by the Otsu method,[23] obtaining values between 0.5 and 0.6. Despite having a certain remaining noise, a binary image with a significantly elliptical arrangement was obtained. Several steps were followed for noise removal. First, to obtain a homogeneous white ellipse, black pixels were filled inside it. Then, external white pixels having a radius greater than 4 pixels were closed using a morphological operator, forming a single structure. Finally, to soften the edges and remove speckle noise, 2 averaging filters (16 × 16 pixels) were successively applied. The resulting lengths of the major and minor axes of the ellipse were used to automatically calculate the AR. The tilt angle of the major axis was also obtained. In summary, for each ROI, FFT is performed, and the image obtained is processed by eliminating noise to finally obtain the ellipse figure (see Figure 2,

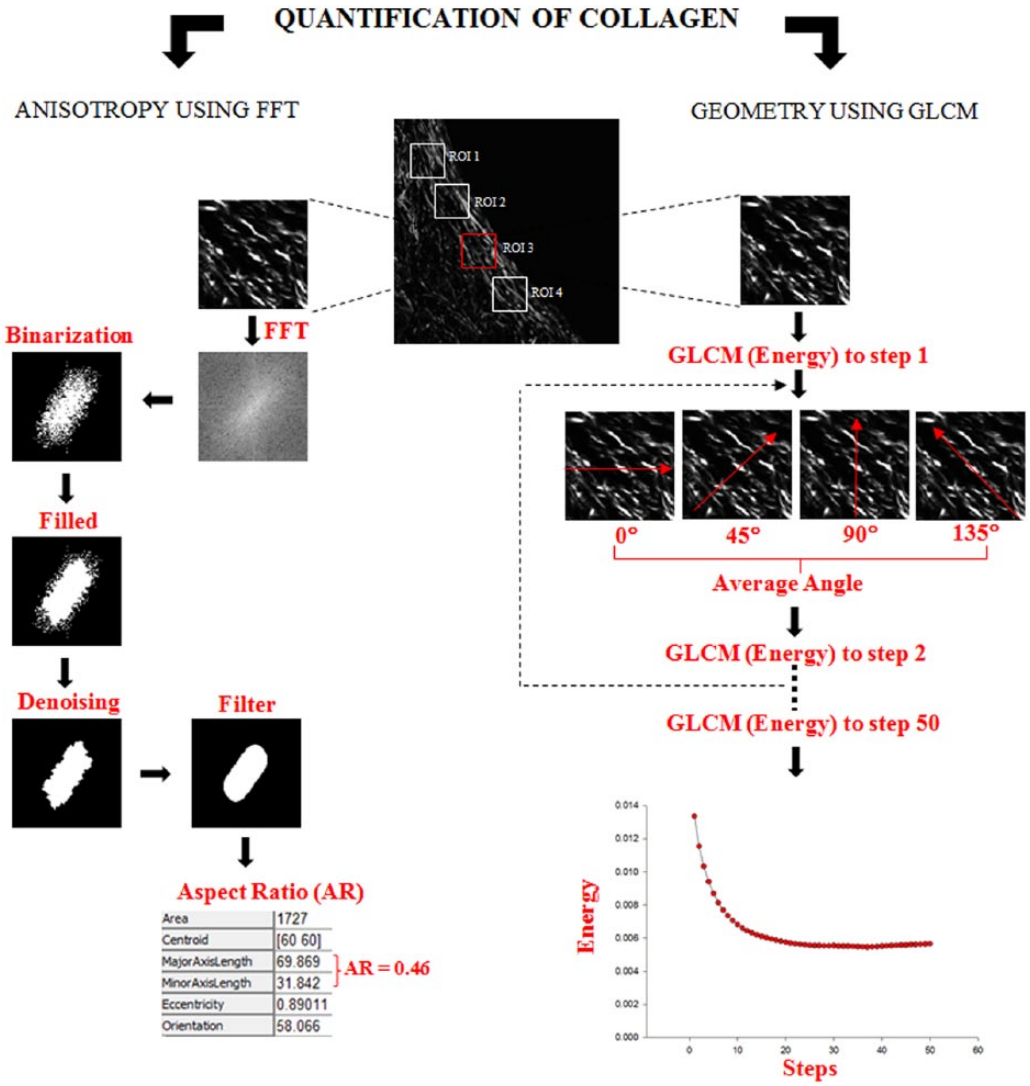

Figure 2. Diagram of the steps involved in the measurements of collagen in SHG signals using FFT (left) and GLCM (right) parameters. From the SHG image, 4 different regions adjacent to the epithelial/stromal interface were used. For simplicity, just the steps for 1 ROI are shown. GLCM is exemplified using a range of 50 pixels and 4 directions (0°, 45°, 90°, and 135°). Similar procedure is applied to the remaining variables (contrast, entropy, and correlation). FFT indicates fast Fourier transform; GLCM; gray-level co-occurrence matrix; ROI, region of interest; SHG, second-harmonic generation.

left panel). The AR Website (https://github.com/Crandelz/Fiber-Organization) contains the most up-to-date version of the Octave plugin.

Collagen geometry quantification using GLCM. Textural measures were obtained using the ImageJ software with a customized version of the TextureToo plugin. Briefly, in the original version, after the ROI choice, variables must be selected and final outcomes are presented in an output table. After calculus, it is necessary to change the step and repeat the procedure from the beginning, with the results overwritten on the same table. This procedure must be followed in each direction change, for the remaining 3 angles. The procedure becomes tedious if many steps are required for a complete quantification. The plugin was modified with the aim to optimize this process and reduce the computational cost, allowing the user to enter the initial and final steps (pixels) of a given angle. The entire range, for all texture estimation, is calculated with 1-stop increments in a single program execution. Next, new step direction is selected and the entire range is calculated. In addition, results are exported in a file instead of being overwritten as in the original plugin. Finally, for each variable, an average is performed between the 4 directions (see Figure 2, right panel, where the whole procedure is shown). Final calculation was performed in 4 ROIs, and the curve does represent the mean ± standard deviation for each pixel. The Texture-Analysis Website (https://github.com/Crandelz/Texture-Analysis) contains the most up-to-date version of the ImageJ plugin.

Phantoms

Facing different fiber arrangements, to understand the meaning and behavior of the obtained variables, artificial images (phantoms) simulating those arrangements were made (Figure 3). Ovary collagen fiber package widths were measured using the *line* and the *Scarbar plus* commands of the ImageJ

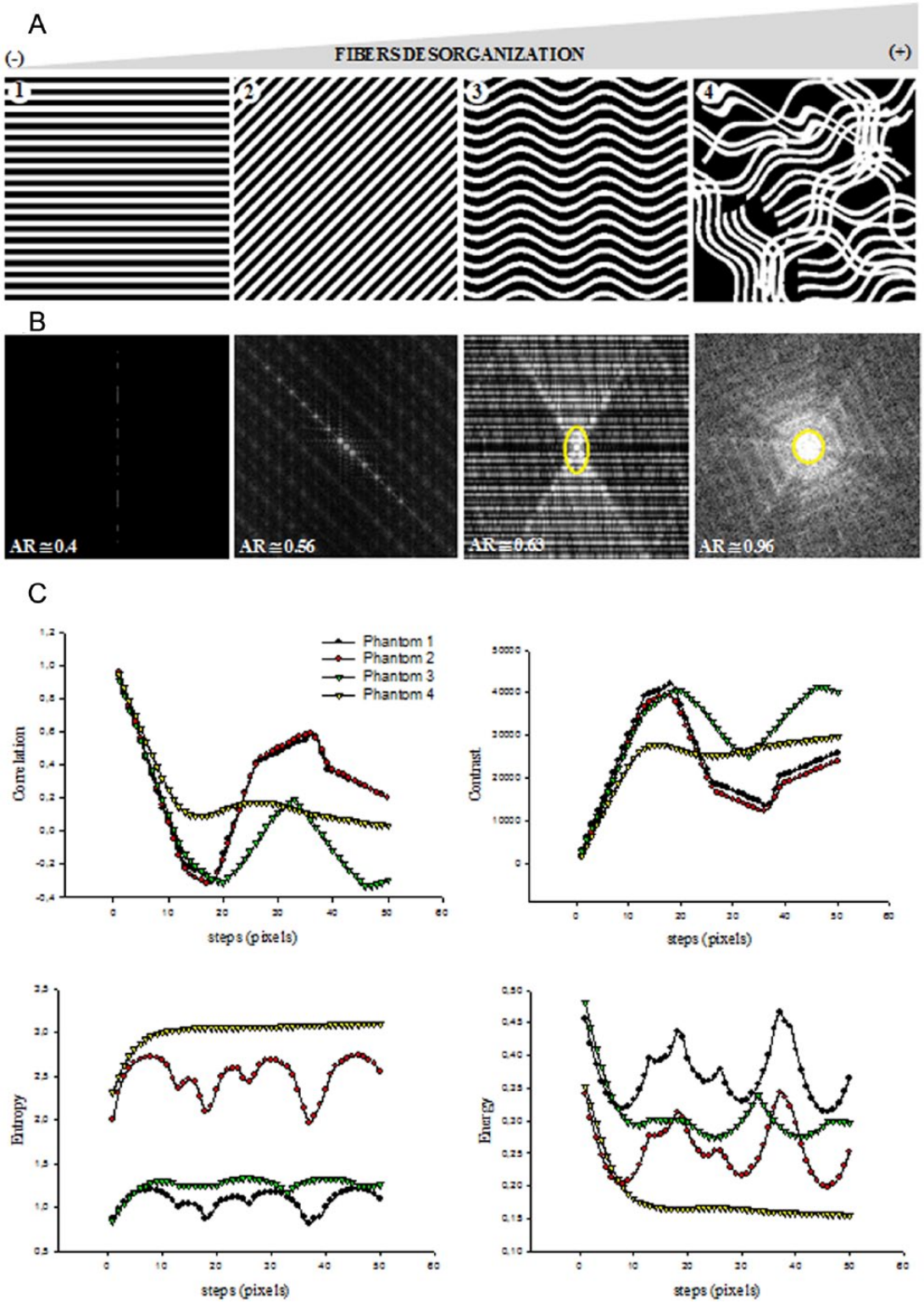

Figure 3. (A) Phantoms used to represent straight (normal) and wavy (pathologic) collagen fibers of stromal ovarian tumors—1: normal; 4: pathologic; 2 and 3: intermediate condition representations. (B) FFT outcomes of phantoms shown in (A) indicating AR. (C) GLCM-texture parameters (correlation, contrast, entropy, and energy) used to characterize phantoms, as a function of *d* (0-50) in the directions of 0°, 45°, 90°, and 135°. Black line: phantom 1 (straight fibers); red line: phantom 2 (tilted straight fiber); green line: phantom 3 (wavy fibers); and yellow line: phantom 4 (disorganized fibers). AR indicates aspect ratio; FFT, fast Fourier transform; GLCM; gray-level co-occurrence matrix.

software (range, 1-3 μm). This operation allows us to determine that fibers have a width of approximately 2.5 μm. Because all ovarian stroma images used are 1024 × 1024 pixels (212.25 μm × 212.25 μm), fiber width would be around 12 pixels. The phantoms were generated with known thickness, spacing, and orientation fibers. Simulated fibers were built with a width and

interspacing of 13 pixels, in agreement with the data of Hu et al.[20] Two phantoms of straight fibers were built with 0° and 45° orientation angles, one with wavy fibers and the other having random fiber distribution (Figure 3A).

Statistics

For multigroup comparisons, 1-way analysis of variance (ANOVA) with a post hoc Tukey-Kramer test was used. We performed Welch test (unpaired and with normal distribution) for 2-group comparisons (normal vs tumor). The levels of significance were as follows: $^{*}P < .05$, significant; $^{**}P < .01$, very significant; $^{***}P < .001$, extremely significant. Data were analyzed with GraphPad InStat (GraphPad Software, San Diego, CA, USA) and graphic with SigmaPlot 2001 V7.0 (Systat Software Inc. [SSI], Richmond, CA, USA). Posterior analysis was performed only with parameters with values of $P < .01$ and $P < .001$ and presented as the lower bound (LB) and upper bound (UB) 95% confidence interval.

Results

Phantom analyses

Anisotropy (FFT). The transform of a rectangular pulse is a *sync* function. Considering the first 2 phantoms (Figure 3A) as a series of square pulses on the *y* axis with 0° or 45°, it follows that the FFT is a *sync* function along axis *v*, perpendicular to the direction of fibers (Figure 3B). By observing the images, it is concluded that image FFT provides information about fiber orientation, showing a pattern with predominantly perpendicular orientation to it. Third phantom is of periodic nature because both the *y* and *x* axes cross between black and white spaces of curves. For this reason, FFT has frequency variations along entire plane with a concentration in the center plane, tending to form an ellipse whose major axis is perpendicular to the fiber direction. Last phantom FFT of disordered fibers presents frequency variations throughout the *uv* plane with concentration peaks of major amplitude in the central region, tending to form a circular pattern, in contrast to elliptical patterns in phantoms with higher order. This allows us to detect structures that lack predominant orientation or regular organization in fibers by the FFT circular pattern. From FFT calculations, phantoms with random lines showed an AR close to 1 compared with more order lines that are close to 0 (Figure 3B).

Geometry analyses—texture (GLCM). Figure 3C depicts the different texture variable calculi over the phantoms. These parameters can obtain the following quantitative values: contrast, between 0 and 65 500; entropy, between 0 and less than 10; energy, between 0 and 1, and correlation, between −1 and 1. Phantoms 1 and 2 exhibit similar correlation and contrast curves, being a remarkable fact, because both images are just rotated versions of each other. However, entropy and energy curves exhibit a similar morphology behavior, but having an offset between same phantoms. Thus, as can be easily observed, when the lines are regularly organized, correlation and contrast curves depict pronounced peaks and valleys, which are related to line width. Wavy fiber phantoms (3) exhibit high contrast (between 25 000 and 40 000, from pixel 20 onward), and because any movement direction records the variation in gray levels caused by the geometry itself, it is verified in any direction. This situation looks similar to the randomly arranged fiber phantom (4), in which all directions exhibit high contrast (around 25 000, from pixel 10 onward). In the first phantoms, entropy is low (high regularity degree). Conversely, randomly arranged fibers (phantom 4) have high entropy values (greater structural complexity). For energy, it is observed that when moving along the fiber direction orientation, energy is greater than that verified in other directions. Conversely, the disordered fiber phantoms (4) exhibit low energy values in all directions (close to 0), indicating less uniformity of the system analyzed.

Ovarian tissues analyses

Anisotropy and geometry analyses were performed, and values of the lower bound (LB) and upper bound (UB) 95% confidence interval are presented. Table 2 shows the AR and different variable textures quantified by GLCM. In the first case, all carcinomas showed extremely significant differences when compared with normal ovary tissue. For mucinous subtype and its variants (borderline subtype was not considered because it was only identified in a single patient), extremely significant differences between normal tissue and mucinous adenocarcinoma, as well as between adenoma and adenocarcinoma subtypes, were found. For serous subtype, no statistically significant differences were obtained.

For the texture analysis, curves were plotted for each of the 4 calculated variables (correlation, contrast, energy, and entropy) from steps 1 to 50. Figure 4 illustrates the behavior of the 4 variables for normal vs pathologic tissues (left column), normal tissue vs serous subtype (middle column), and normal tissue vs mucinous subtype (right column). Qualitatively, more aggressive tumors—adenocarcinomas and borderline—show lower energy (less uniformity) and higher entropy (more complexity), which is highly consistent with the previous analysis performed in the phantoms. From these curves, we analyzed statistically each variable in 4 specific steps, preferably where the curves did not exhibit overlapping. Regarding the correlation analysis, it was observed that approximately from step 20 all curves became asymptotic; for that reason, steps 3, 5, 10, and 15 were selected. In contrast, energy and entropy curves showed no overlapping or intersection from step 8; therefore, steps 10, 20, 30, and 40 were selected. For simplicity, Table 2 presents the quantification only for 2 steps (double black arrow in Figure 4) and was performed between tumor types and for each step independently.

In summary, from the statistical analysis (Table 2), it was concluded that AR is useful to confirm the pathologic state of a sample, whereas the textural variables—at specific step selected—allowed to distinguish between subtypes of carcinomas. In Table 2, texture parameters showed very significant

Table 2. Aspect ratio and texture variables.

ANISOTROPY					
ASPECT RATIO COMPARISON BETWEEN TUMORS					
	NORMAL[A]	MIXED[B]	ENDOMETRIOID[C]	S. ADENOCARCINOMA[D]	M. ADENOCARCINOMA[E]
	(0.51–0.57)	(0.61–0.70)***	(0.65–0.73)***	(0.59–0.68)***	(0.62–0.71)***
ASPECT RATIO COMPARISON BETWEEN MUCINOUS VARIANTS					
	NORMAL[A]	**ADENOMA[B]**	**BORDERLINE[C]**	**ADENOCARCINOMA[D]**	
	(0.51–0.57)[d]	(0.46–0.53)[d]	—	(0.62–0.71)[a,b]	
GEOMETRY					
TEXTURE VARIABLES COMPARISON BETWEEN TUMORS					
CORR.	**NORMAL[A]**	**MIXED[B]**	**ENDOMETRIOID[C]**	**S. ADENOCARCINOMA[D]**	**M. ADENOCARCINOMA[E]**
5	$(2.7 \times 10^{-4}–5.6 \times 10^{-4})$	$(2.4 \times 10^{-4}–3.7 \times 10^{-4})^{c,d}$	$(1.3 \times 10^{-4}–12.0 \times 10^{-4})^{**,b,d}$	$(8.3 \times 10^{-4}–1.4 \times 10^{-4})^{b,c,e}$	$(1.4 \times 10^{-4}–2.9 \times 10^{-4})^{d}$
10	$(8.2 \times 10^{-5}–2.0 \times 10^{-4})$	$(8.8 \times 10^{-5}–1.5 \times 10^{-4})^{d,e}$	$(4.8 \times 10^{-5}–9.5 \times 10^{-5})^{**}$	$(2.5 \times 10^{-5}–4.6 \times 10^{-5})^{b}$	$(4.2 \times 10^{-5}–1.1 \times 10^{-4})^{b}$
Cont.					
20	(1807.9–3136.3)	(1825.5–2716.6)[d]	(2833.8–3907.3)	(4590.5–6601.7)[b]	(2837.7–4381.7)
30	(1826.8–3156.0)	(1841.3–2734.4)[d]	(2738.1–4006.3)	(4650.9–6710.2)[b]	(2887.6–4437.5)
Ener.					
20	$(1.9 \times 10^{-3}–3.9 \times 10^{-3})$	$(3.5 \times 10^{-3}–8.0 \times 10^{-3})^{c,d}$	$(1.2 \times 10^{-3}–2.1 \times 10^{-3})^{b}$	$(1.3 \times 10^{-3}–2.3 \times 10^{-3})^{b}$	$(2.0 \times 10^{-3}–3.6 \times 10^{-3})$
30	$(1.9 \times 10^{-3}–3.8 \times 10^{-3})$	$(3.4 \times 10^{-3}–7.7 \times 10^{-3})^{c,d}$	$(1.2 \times 10^{-3}–2.1 \times 10^{-3})^{b}$	$(1.3 \times 10^{-3}–2.3 \times 10^{-3})^{b}$	$(1.9 \times 10^{-3}–3.5 \times 10^{-3})$
Entr.					
20	(6.9–7.4)	(6.6–7.1)[c,d]	(7.4–7.7)[b]	(7.4–7.8)[b]	(7.0–7.5)
30	(6.9–7.4)	(6.6–7.1)[c,d]	(7.3–7.6)[b]	(7.4–7.7)[b]	(6.9–7.4)
TEXTURE VARIABLES COMPARISON BETWEEN SEROUS VARIANTS					
CORR.	**NORMAL[A]**	**ADENOMA[B]**	**BORDERLINE[C]**	**ADENOCARCINOMA[D]**	
5	$(2.7 \times 10^{-4}–5.6 \times 10^{-4})^{d}$	$(2.4 \times 10^{-4}–4.7 \times 10^{-3})^{d}$	$(1.3 \times 10^{-4}–3.3 \times 10^{-4})$	$(8.3 \times 10^{-5}–1.4 \times 10^{-4})^{a,b}$	
10	$(8.2 \times 10^{-5}–2.0 \times 10^{-4})^{d}$	$(9.9 \times 10^{-5}–1.9 \times 10^{-4})^{d}$	$(5.5 \times 10^{-5}–1.6 \times 10^{-4})^{d}$	$(2.5 \times 10^{-5}–4.6 \times 10^{-5})^{a,b,c}$	

(Continued)

Table 2. (Continued)

TEXTURE VARIABLES COMPARISON BETWEEN SEROUS VARIANTS				
CORR.	**NORMAL(A)**	**ADENOMA(B)**	**BORDERLINE(C)**	**ADENOCARCINOMA(D)**
Cont.				
20	(1807.9–3136.3)^{d}	(2856.9–4181.4)	(2970.4–4494.0)	(4590.5–6601.7)^{a}
30	(1826.8–3156.0)^{d}	(2998.8–4419.8)	(3035.2–4669.2)	(4650.9–6710.2)^{a}
Ener.				
20	$(1.9 \times 10^{-3}–3.9 \times 10^{-3})^{b,c}$	$(8.1 \times 10^{-4}–2.1 \times 10^{-3})^{a}$	$(9.1 \times 10^{-4}–3.9 \times 10^{-3})^{a}$	$(1.3 \times 10^{-3}–2.3 \times 10^{-3})$
30	$(1.9 \times 10^{-3}–3.8 \times 10^{-3})^{b,c}$	$(8.1 \times 10^{-4}–2.1 \times 10^{-3})^{a}$	$(8.8 \times 10^{-4}–3.9 \times 10^{-3})^{a}$	$(1.3 \times 10^{-3}–2.3 \times 10^{-3})$
Entr.				
20	(6.9–7,4)^{b,c}	(7.6–8.0)^{a}	(7.4–8.0)^{a}	(7.4–7.8)
30	(6.8–7,4)^{b,c}	(7.5–7.9)^{a}	(7.3–7.9)^{a}	(7.4–7.7)
TEXTURE VARIABLES COMPARISON BETWEEN MUCINOUS VARIANTS				
CORR.	**NORMAL(A)**	**ADENOMA(B)**	**BORDERLINE(C)**	**ADENOCARCINOMA(D)**
5	$(2.7 \times 10^{-4}–5.6 \times 10^{-4})^{b}$	$(1.4 \times 10^{-4}–2.1 \times 10^{-4})^{a}$	$(1.2 \times 10^{-4}–3.3 \times 10^{-4})$	$(1.4 \times 10^{-4}–2.9 \times 10^{-4})$
10	$(8.2 \times 10^{-5}–2.0 \times 10^{-4})$	$(5.8 \times 10^{-5}–1.0 \times 10^{-4})$	$(3.8 \times 10^{-5}–1.5 \times 10^{-4})$	$(4.2 \times 10^{-5}–1.1 \times 10^{-4})$
Cont.				
20	(1807.9–3136.3)^{b}	(3681.0–5374.8)^{a}		
30	(1826.8–3156.0)^{b}	(3813.6–5614.2)^{a}		

Abbreviations: Cont., contrast; Corr., correlation; Ener., energy; Entr., entropy; M, mucinous; S, serous.
The Welch test was used for 2-group comparisons (normal vs tumor stage).
For multigroup comparisons, tissues are identified with letters. Within a row, values with different superscript letters are statistically different (at least $P < .01$), following analysis of variance. For example, in Geometry, results in column Mixed are indicated with letter (b), the numbers below Mixed have other letters, such as c and d, which means that Mixed has statistically significant differences with endometrioid (c) and serous adenocarcinoma (d).
Both variables are presented as lower bound (LB) and upper bound (UB) 95% confidence interval. Letters in parentheses are used to identify the stage of the tissue. Comparisons are made in 1 row.
$^{**}P < .01$, very significant; $^{***}P < .001$, extremely statistically significant differences.

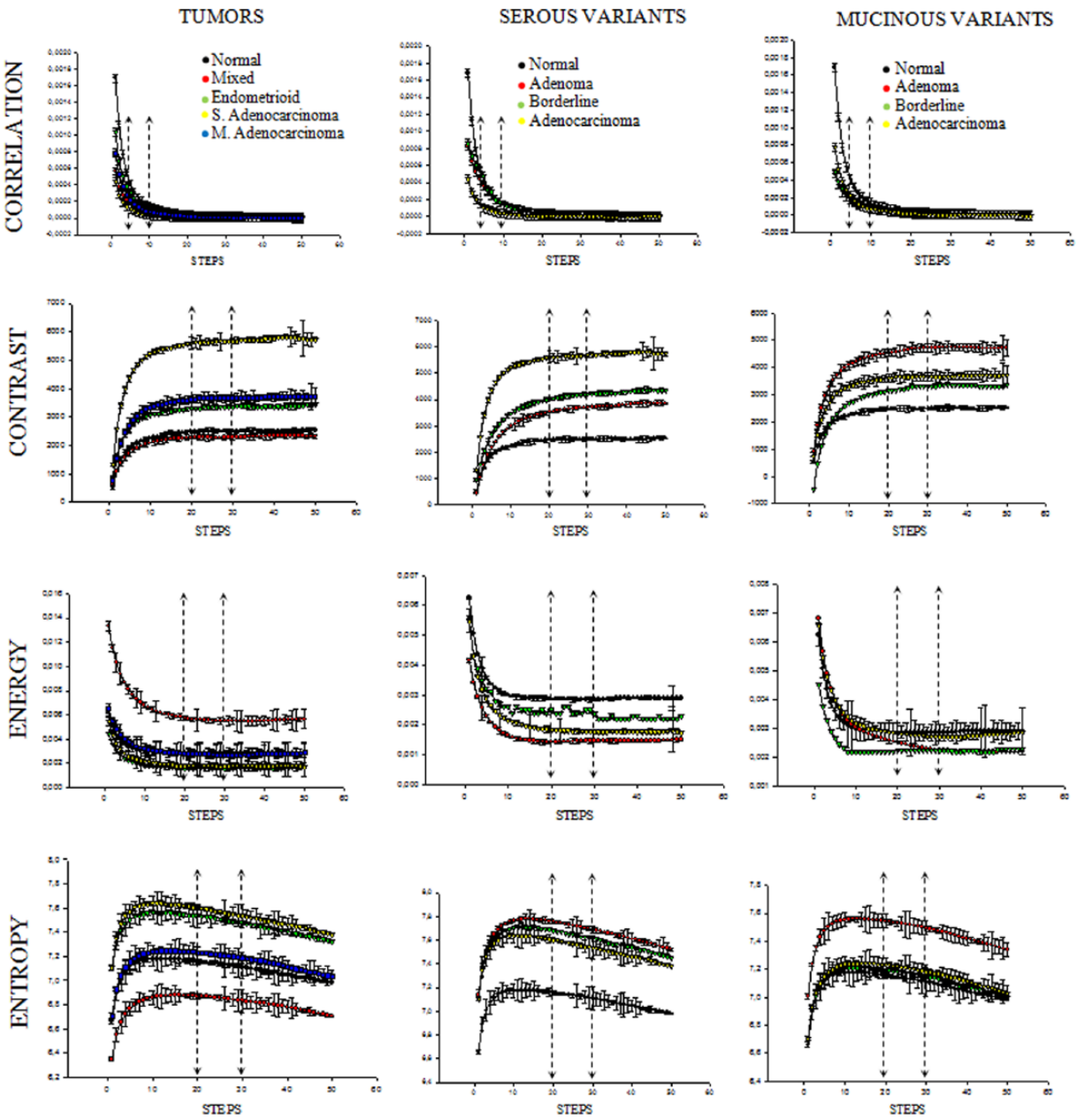

Figure 4. Texture analysis using the gray-level co-occurrence matrix. Each point represents the regions of interest average for that step. Correlation (first row), contrast (second row), energy (third row), and entropy (fourth row) values in stromal ovarian tissues vs distance pixels, ranging from 1 to 50 pixels (0.35-17.30 µm) in 0°, 45°, 90°, and 135° image directions. Normal vs pathologic tissues (left column), normal tissue vs serous subtype (middle column), and normal tissue vs mucinous subtype (right column). Double black arrow marks the steps where quantification was done. M indicates mucinous; S, serous.

differences. Evaluating the confidence intervals and the usefulness of each texture variables, a protocol was built to systematize the analysis of second-harmonic ovarian biopsy microscopic images.

Protocol

To differentiate between healthy and pathologic tissues and to enable the distinction between subtypes and degrees of malignancy, the following steps are performed:

1. Take SHG images of size 1024 × 1024 pixels;
2. Proceed to the identification of ovarian epithelium;
3. Choose (immediately below the epithelium) a number of ROIs (at least 4) of 120 × 120 pixels;
4. For each ROI, obtain the ellipses/circles calculating the FFT and also AR;
5. Average between them, getting a single value for the AR;
6. Calculate the following textural variables for each ROI:

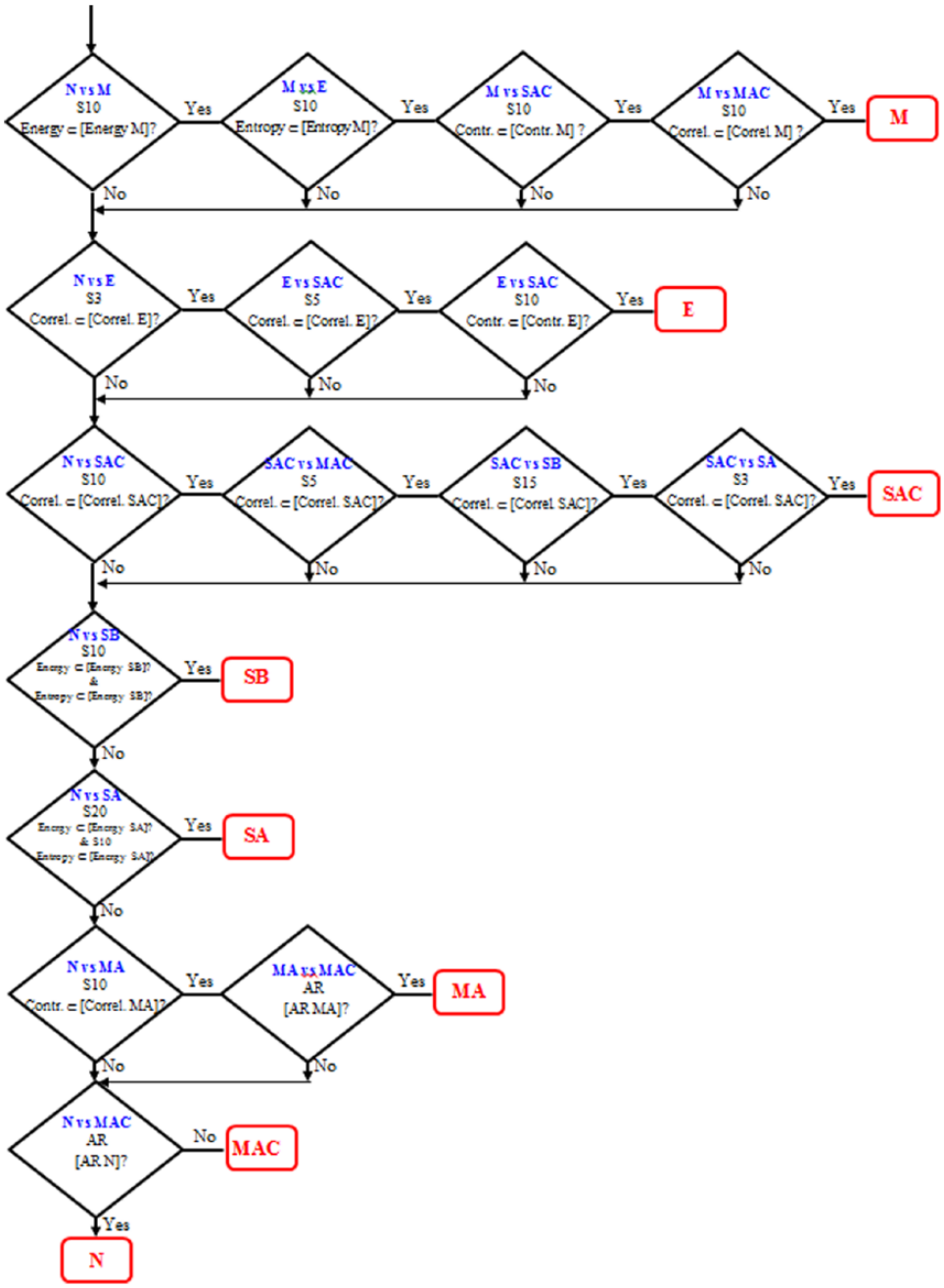

Figure 5. Decision tree for ovarian cancer detection using gray-level co-occurrence matrix (GLCM) features extracted from steps 3, 5, 10, and 15 for correlation and from steps 10, 20, 30, and 40 for contrast, energy, and entropy. AR indicates aspect ratio; Contr., contrast; Correl., correlation; E, endometrioid; M, mixed; MA, mucinous adenoma; MAC, mucinous adenocarcinoma; N, normal; SA, serous adenoma; SAC, serous adenocarcinoma; SB, serous borderline; S_n, step n; vs, versus. [ ] indicates confidence interval; ⊂, included in.

(a) Energy for steps 10 and 20;
(b) Entropy to step 10;
(c) Contrast to step 10;
(d) Correlation for steps 3, 5, 10, and 15;

7. Average each variable between ROIs;
8. Sample classification using the decision tree.

Decision tree for classification detection using AR and GLCM features is shown in Figure 5.

Discussion and Conclusions

Careful examination of biopsies is quite powerful, but many visual analysis tasks performed in the laboratory are prone to error and are monotonous and subjective.[24-26] Nevertheless, even today, most of the imaging studies in biological sciences do not use fully automated analysis,[24] so a need remains for image analysis software that allows interactivity but records any choices made by the user and automates repetitive steps whenever possible.

Usually, it is quite easy to analyze one region of a single image, but current imaging studies generally require analysis of several regions in numerous images. This reiteration can render even the simplest examination tasks quite complicated in practice. The principal difficulties include recording ROIs and linking them to the resulting quantifications. One approach to improve the reproducibility of image analysis is to use programming or scripting to automate the analysis process, thus removing variability from interactive user input.

Following this premise and also attending the recent concern about reproducibility of scientific research,[27] in this work we present a successful image analysis strategy to address these issues. Using Octave and ImageJ, that support scripts written in their own macro language, we developed an improvement and automated plugins, both for AR determination from axis of ellipse figures obtained with FFT and for texture variables obtained with GLCM.[28] In previous works, our group and other researches obtained the axis of the ellipse by manually drawing the shape of the figure on the image of the FFT.[29,30] In that case, it was fundamental to trace the edge of the figure that forms the set of frequencies obtained to thereby highlight the ellipse (major and minor axes) formed. Given that this varies according to the image, and should be chosen trying to achieve a perceptibly elliptical shape, the methodology is subject to errors of subjective assessment of the operator. Now, we calculated automatically the AR after processing the image of FFT. The results obtained are coincident with previous one,[5,11,18,29] which means values close to 0 indicate samples with predominant alignment (normal samples) and therefore more anisotropy. The values near to 1 indicate more randomly arranged fibers (tumor samples), which means it presents isotropy. For GLCM determination, previous plugin for every variable required a new run each time so that step and angle are modified. Now, with a customized version of TextureToo plugin, the user can enter the initial and final steps for a given angle. The entire range is calculated with 1-stop increments in a single program execution. The results are exported in a file that contains the image name, followed by the corresponding step number, orientation, and texture variable, all separated by commas.

The modification of these processing tools allows the accurate automatic calculation of anisotropy and texture variables, shortening processing times and avoiding errors of subjective judgment of the operator. We compared time calculations with previous works and the time consumed was reduced considerably. When serous tumor types were analyzed (a total of 33 samples) with this new strategy, the final results (after selecting the ROIs) were obtained in 8 minutes compared with more than 50 minutes needed in original tray.[18,30] These improvements are important because these metrics (AR and GLCM) are systematically used for diagnosis of ovarian pathologies.[16,31]

Our strategy is fundamentally applicable for some small- and medium-sized data sets (<200 images) where interactive manual ROI segmentation is the available option (step 3 in our protocol). This allows us to analyze many SHG images and build a diagnosis protocol, which use a classification tree to differentiate between biopsies. Interactive segmentation may simply be much more efficient than automatic segmentation where tuning the segmentation algorithm parameters can be extremely time-consuming. A positive aspect to mention is that an interactive approach allows the comparison of different analysis parameters. For example, using this protocol, we tested several serous adenocarcinoma biopsies, and all were successfully classified. In the near future, we are planning to obtain and test more biopsies for all human ovary tumor types with the idea of continuing to add more samples to our protocol. We do expect that confidence intervals (presented in Table 2) become more accurate as more samples will be analyzed with this methodology. Our tactic allows interactivity but records any choices made by the user and automates repetitive steps whenever possible. In this line, recently an open-source Java-based plugin for the Fiji distribution of ImageJ, called Slide Set, was created.[24] This provides a framework for interactive and reproducible image analysis, controlled through a graphical user interface with intuitive menu-based commands. For very large data sets, where the number of images alone makes interactivity impractical, several software packages that support fully automated image analysis workflows with little scripting exist, such as CellProfiler, PhenoRipper, and the Protocols feature of Icy.[32-34]

In conclusion, it was shown that the processing and analysis of SHG images allow to characterize different ovarian tumors with specific stroma disorder through textural differences. Also, anisotropy measurements indicate that collagen network in serous, mucinous, endometrial, and mixed tumors is markedly different from the corresponding normal tissue, demonstrating there are structural changes in fibrillary collagen matrix in malignant tissues and these changes can be measured with SHG techniques. As stated above, SHG images from collagen fibers could become a promising technique to understand carcinogenesis. In this sense, early identification of structural changes in the matrix allows the early detection and treatment of disease.

Careful investigations on the selection of the fields of view and the interpretation of the quantitative parameters are required, and other quantitative methods can also be complemented to provide comprehensive information associated with the progression of ovarian cancer. In the future work, we are planning for more biopsies, performing an image classification using a support vector machine algorithm and analyzing the performance of the classifier by finding the average and standard deviation of the area under the receiver operator characteristic curve for the 100 training sets. This curve will provide us the sensitivity and specificity of the diagnosis method.

In summary, we have demonstrated the feasibility of performing classification of collagen fibril morphology based on first- and second-order texture statistical parameters derived

from SHG images. With better understanding of early disease, we hope to develop an optical diagnostic test for ovarian cancer.

Author Contributions

JA conceived the manuscript. AZ, JL, and KSS designed the experiments. JA, AZ, JL, KSS, and LE analyzed the data. AZ and JA wrote the first draft of the manuscript. HFC and CLC contributed with images acquisition. HFC, LA, and FB collaborated with handling and diagnostic of biopsies. LE and VHC contributed to the writing of the manuscript. JA, HFC, CLC, LA, LE, FB, and VHC made critical revisions and approved final version. All authors have reviewed and approved the final manuscript.

REFERENCES

1. Cicchi R, Vogler N, Kapsokalyvas D, Dietzek B, Popp J, Pavone FS. From molecular structure to tissue architecture: collagen organization probed by SHG microscopy. *J Biophotonics*. 2013;6:129–142.
2. Zoumi A, Yeh A, Tromberg BJ. Imaging cells and extracellular matrix in vivo by using second-harmonic generation and two-photon excited fluorescence. *Proc Natl Acad Sci U S A*. 2002;99:11014–11019.
3. Campagnola PJ, Loew LM. Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. *Nat Biotechnol*. 2003;21:1356–1360.
4. Roth S, Freund I. Optical second-harmonic scattering in rat-tail tendon. *Biopolymers*. 1981;20:1271–1290.
5. Adur J, Pelegati VB, de Thomaz AA, et al. Second harmonic generation microscopy as a powerful diagnostic imaging modality for human ovarian cancer. *J Biophotonics*. 2014;7:37–48.
6. Ghazaryan A, Tsai HF, Hayrapetyan G, et al. Analysis of collagen fiber domain organization by Fourier second harmonic generation microscopy. *J Biomed Opt*. 2013;18:031105.
7. Cicchi R, Kapsokalyvas D, De Giorgi V, et al. Scoring of collagen organization in healthy and diseased human dermis by multiphoton microscopy. *J Biophotonics*. 2010;3:34–43.
8. Chen J, Zhuo S, Jiang X, et al. Multiphoton microscopy study of the morphological and quantity changes of collagen and elastic fiber components in keloid disease. *J Biomed Opt*. 2011;16:051305.
9. Strupler M, Pena A, Hernest M, et al. Second harmonic imaging and scoring of collagen in fibrotic tissues. *Opt Express*. 2007;15:4054–4065.
10. Lin S-J, Hsiao C-Y, Sun Y, et al. Monitoring the thermally induced structural transitions of collagen by use of second-harmonic generation microscopy. *Opt Lett*. 2005;30:622–624.
11. Brown E, McKee T, di Tomaso E, et al. Dynamic imaging of collagen and its modulation in tumors in vivo using second-harmonic generation. *Nat Med*. 2003;9:796–801.
12. Campagnola PJ, Dong CY. Second harmonic generation microscopy: principles and applications to disease diagnosis. *Laser Photonics Rev*. 2011;5:13–26.
13. Zhuo S, Yan J, Chen G, et al. Label-free monitoring of colonic cancer progression using multiphoton microscopy. *Biomed Opt Express*. 2011;2:615.
14. Conklin MW, Eickhoff JC, Riching KM, et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. *Am J Pathol*. 2011;178:1221–1232.
15. Watson JM, Rice PF, Marion SL, et al. Analysis of second-harmonic-generation microscopy in a mouse model of ovarian carcinoma. *J Biomed Opt*. 2012;17:0760021.
16. Tilbury K, Campagnola P. Applications of second-harmonic generation imaging microscopy in ovarian and breast cancer. *Perspect Medicin Chem*. 2015;16:21–32.
17. Ambekar Ramachandra Rao R, Mehta MR, Toussaint KC Jr. Quantitative analysis of biological tissues using Fourier transform-second-harmonic generation imaging. In: Multiphoton Microscopy in the Biomedical Sciences X; February 26, 2010; San Francisco, CA.
18. Adur J, Pelegati VB, de Thomaz AA, et al. Optical biomarkers of serous and mucinous human ovarian tumor assessed with nonlinear optics microscopies. *PLoS ONE*. 2012;7:e47007.
19. Matteini P, Ratto F, Rossi F, et al. Photothermally-induced disordered patterns of corneal collagen revealed by SHG imaging. *Opt Express*. 2009;17:4868.
20. Hu W, Li H, Wang C, Gou S, Fu L. Characterization of collagen fibers by means of texture analysis of second harmonic generation images using orientation-dependent gray level co-occurrence matrix method. *J Biomed Opt*. 2012;17:026007.
21. Mostaço-Guidolin LB, Ko AC, Wang F, et al. Collagen morphology and texture analysis: from statistics to classification. *Sci Report*. 2013;3:Article 2190.
22. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE T Syst Man Cyb*. 1973;3:610–621.
23. Otsu N. A threshold selection method from gray-level histograms. *IEEE T Syst Man Cyb*. 1979;9:62–66.
24. Nanes BA. Slide set: reproducible image analysis and batch processing with ImageJ. *Biotechniques*. 2015;59:269–278.
25. Lamprecht MR, Sabatini DM, Carpenter AE. CellProfiler: free, versatile software for automated biological image analysis. *Biotechniques*. 2007;42:71–75.
26. Vokes MS, Carpenter AE. Using CellProfiler for automatic identification and measurement of biological objects in images. *Curr Protoc Mol Biol*. 2008;109:1–13.
27. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10:712.
28. Luna JSJ, Sanchez Salas K. *Standardization of a method for detecting ovarian cancer, as potential diagnostic tool, from the second harmonic generation in collagen fibers* [graduation thesis] (unpublished). 2015.
29. Williams RM, Flesken-Nikitin A, Ellenson LH, et al. Strategies for high-resolution imaging of epithelial ovarian cancer by laparoscopic nonlinear microscopy. *Transl Oncol*. 2010;3:181–194.
30. Adur J, Pelegati VB, Costa LF, et al. Recognition of serous ovarian tumors in human samples by multimodal nonlinear optical microscopy. *J Biomed Opt*. 2011;16:096017.
31. Adur J, Carvalho HF, Cesar CL, Casco VH. Nonlinear optical microscopy signal processing strategies in cancer. *Cancer Inform*. 2014;2:67–76.
32. Kamentsky L, Jones TR, Fraser A, et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics*. 2011;27:1179–1180.
33. Rajaram S, Pavie B, Wu LF, et al. PhenoRipper: software for rapidly profiling microscopy images. *Nat Methods*. 2012;9:635–637.
34. de Chaumont F, Dallongeville S, Chenouard N, et al. Icy: an open bioimage informatics platform for extended reproducible research. *Nat Methods*. 2012;9:690–696.