Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

RESEARCH ARTICLE

Prostate cancer tissue classification by multiphoton imaging, automated image analysis and machine learning

Egleidson F. A. Gomes[1] | Eduardo Paulino Junior[2] | Mário F. R. de Lima[3] | Luana A. Reis[1] | Giovanna Paranhos[1] | Marcelo Mamede[4] | Francis G. J. Longford[5] | Jeremy G. Frey[5] | Ana Maria de Paula*[1]

[1]Departamento de Física, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte-MG, Brazil

[2]Departamento de Anatomia Patológica e Medicina Legal, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte-MG, 30130-100, Brazil

[3]Laboratório Analys Patologia, Av. Bernardo Monteiro 735, Belo Horizonte-MG, 30150-270, Brazil

[4]Departamento Anatomia e Imagem, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte-MG, 30130-100, Brazil

[5]University of Southampton, Southampton SO17 1BJ, UK

Correspondence
*Ana M. de Paula, Departamento de Física, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte-MG, Brazil. Email: ana@fisica.ufmg.br

Abstract

Prostate carcinoma, a slow-growing and often indolent tumour, is the second most commonly diagnosed cancer among men worldwide. The prognosis is mainly based on the Gleason system through prostate biopsy analysis. However, new treatment and monitoring strategies depend on a more precise diagnosis. Here we present results by multiphoton imaging for prostate tumour samples from 120 patients that allow to obtain quantitative parameters leading to specific tumour aggressiveness signatures. An automated image analysis was developed to recognise and quantify stromal fibre and neoplastic cell regions in each image. The set of metrics was able to distinguish between non-neoplastic tissue and carcinoma areas by linear discriminant analysis and random forest with an accuracy of $89 \pm 3\%$, but between Gleason groups of only $46 \pm 6\%$. The reactive stroma analysis improved the accuracy to $65 \pm 5\%$, clearly demonstrating that stromal parameters should be considered as additional criteria for a more accurate diagnosis.

KEYWORDS:
Prostate cancer; multiphoton imaging; diagnosis; reactive stroma; machine learning

1 | INTRODUCTION

Cancer diagnosis is performed by a trained pathologist using conventional microscopy to visualize the histological patterns on biopsy samples. Nonlinear imaging techniques have already been demonstrated as new diagnostic tools that may help improve diagnostic precision and prognosis. A major application is on qualitative and quantitative measurements of the collagen organization in cancer biopsies[1]. The activation of the host stroma microenvironment facing an aggression, as it happens in tissue damage repair, is being observed in many tumours like breast, prostate and colonic cancer. The collagen structure and organization is altered and is thought to be critical in tumour growth and progression[2,3]. Second harmonic generation (SHG) and two-photon excitation fluorescence (2PEF) microscopy allow to study the morphology of the extracellular matrix and to quantify the collagen parameters, including details of these tissue architecture changes, that seems to be associated with stages of tumour development and patient clinical outcome in a large variety of cancers[4–32].

In prostate cancer (PCa) the first nonlinear imaging results, for a limited amount of samples, showed changes in the collagen organization that correlates with the tumour malignancy[33,34]. The PCa grading is based on the Gleason system provided through prostate biopsy analysis on hematoxylin and eosin (HE) stained tissue[35]. The Gleason system attributes values ranging from 1 (most differentiated) to 5 (least differentiated) to the various architectural patterns found in PCa. The primary (i.e. the most prevalent) and secondary grade patterns are added together to give a Gleason score, expressed as a sum.

Since 2005, the Gleason score for prostate core needle biopsies is obtained by adding the primary pattern to the highest remaining one. A new graduation system proposed at the International Society of Urological Pathology Conference (ISUP) was established in 2014, referred to as ISUP grade or simply WHO (World Health Organization) grade. This is based in a 5 prognostic groups that correspond to the Gleason scores: group 1 (3 + 3), group 2 (3 + 4), group 3 (4 + 3), group 4 (sum of scores = 8) and group 5 (sum of scores ≥ 9). The first reports showed that there is an increase on the collagen alignment with the increase of the Gleason grade [33,34]. In addition, recent results [36] for 64 patients showed that the collagen morphology characteristics are associated with biochemical recurrence.

In this work, we present multiphoton imaging by SHG and 2PEF for prostate biopsies from 120 patients. We analysed about 350 biopsy regions on histological slides covering four groups within the Gleason system (table 1). Multiphoton microscopy has the advantage of allowing the imaging of the tissue with no need for the cutting and staining procedures. Nonetheless, in this study, we used the slides that had already been stained on HE when first analyzed as part of a diagnostic routine. In this case, it made it possible to get images in the same areas that the pathologist considered for diagnosis in order to compare the analyses with the Gleason system. In addition, we propose a classification based on the intensity of the reactive stroma (RS) that can be obtained by the multiphoton imaging. The SHG images show signals with high specificity for collagen, that allow quantitative stroma scoring, which is not possible in white light HE analysis.

We developed an automated image analysis software capable of recognizing the collagen fibres and cellular regions in the images to measure the parameters at these specific regions. In a previous work on canine mammary cancer [37], we showed that these image analysis procedures could extract a range of important features that allowed a good discrimination between the specific histological carcinoma. The imaging techniques and image analysis software allowed to grasp the important features to evaluate cell distribution and the collagen fibre network. Here, we demonstrate that by including the RS intensity the PCa groups within the Gleason scores can be better distinguished to provide a more accurate diagnosis. The image classification can be significantly improved by using data analysis methods such as machine learning. The extracted image features were used in linear discriminant analysis (LDA) and random forest analysis that allowed an excellent accuracy in discrimination between non-neoplastic tissue and carcinoma areas and also to track the tumour progression.

2 | EXPERIMENTAL DETAILS

2.1 | Samples

For this work, we selected 120 prostate biopsies from the diagnostic routine at a private pathology laboratory (Laboratório Analys Patologia) in Belo Horizonte, Minas Gerais State, from 2017 to 2019. The samples were fixed in buffered formalin and processed according to the conventional histology technique until inclusion in paraffin blocks. That was followed by a 3 to 5 μm thick cut and staining in HE. We selected the slide that concomitantly contained the largest tumour representation and the highest histological grade among the six core biopsies normally collected from each patient. Four groups were formed with the 120 slides selected according with ISUP histological grade: group 1 (3 + 3), group 2 (3 + 4), group 3 (4 + 3) and group 4 (sum of scores ≥ 8). Inclusion criteria for the study were: absence of any previous treatment for PCa and neoplasia representativeness in the core biopsy, in order to guarantee its presence in consecutive deep cuts. The biopsies were analyzed independently by two pathologists with experience in uropathology. In diverging situations, the consensus diagnosis was used. The slides were then scanned and two or more representative tumour region's were imaged. Comparative normal prostate parenchyma regions were also selected, resulting in a total of 351 analyzed regions. Table 1 shows the number of analyzed areas in each group.

The presence and intensity of reactive stroma (RS) were also evaluated (see table 1) according to a modified classification system based on a previous published one proposed by Ayala et al. [3], that reflects mainly the proportion of collagen fibre deposition: RS1: Absent or up to 5% collagen fibres in the RS; RS2: Significant to moderate RS (6% to 49% collagen rich stroma compared to the amount of neoplastic cells); RS3: Intense collagen deposition (stromogenic tumours), in which RS accounts for at least the same proportion of neoplastic cells.

The study was approved by the UFMG ethics committee, document CAAE 09288112.0.0000.5149 in 2017. All methods were performed in accordance to the relevant guidelines and regulations and the written consent for the reported experiments was obtained from all participants before their biopsy procedures.

TABLE 1 Number of analysed areas in each Gleason score and reactive stroma groups. The non-neoplastic regions are labeled as Normal, and the sum of scores ≥ 8 as 4 + 4+.

Normal	3 + 3	3 + 4	4 + 3	4 + 4+	Total
40	67	73	79	92	351
11.4%	19.1%	20.8%	22.5%	26.2%	
RS1	RS2	RS3	Total		
93	190	26	309		
30.1%	61.5%	8.4%			

2.2 | Multiphoton imaging

The experimental setup is a scanning microscopy system based on fluorescence by two photon excitation and second harmonic generation using femtosecond laser pulses on a home-modified confocal microscope. The details of the setup has been described in our previous works[34,37]. In brief, we used 140 fs pulses centred at the wavelength of 800 nm, with average power of about 7 mW at the sample position (fluence of 7×10^{-4} J/cm^2). The beam is focused on the sample microscope slab by a 20× objective with a high numerical aperture (NA = 0.9) and the images are obtained in back scattering geometry. The SHG is measured at 400 nm using a filter with bandwidth of 20 nm by directing the signal to a non-descanned detector placed just above the objective. The fluorescence signal from the eosin dye staining the tissue is measured by the standard descanned detector of the confocal unit at the wavelength band of 560 to 600 nm. The transmitted laser image is also collected by the condenser of the microscope. We measured images at the optimized focus position and at ±1 μm from this position to allow for noise filtering by the software. The images were acquired for areas of 0.47×0.47 mm^2.

3 | IMAGE ANALYSIS PROCEDURES

Histological diagnosis is based on cell types and architectural patterns. Recently, attention is also being given to reactive stroma collagen fibres organization[2]. Guided by the pathologists descriptions of these features we developed an image analysis to try to obtain their quantitative measurements. We have developed a software package (PyFibre, available on GitHub[38]) capable of recognizing collagen fibres an cellular regions individually, which is necessary for quantitative measurements. It is important to get these segmentation to obtain the quantitative measurements of described features in each specific areas. The software allows to automate the image analysis and to efficiently obtain results for a large amount of images.

PyFibre automatically reads the images, that are identified by a tag in their file names as SHG and 2PEF. The images are filtered to reduce white and shot noise and then we compose an image that includes the measured SHG, 2PEF and the transmission images. After that, the analysis is performed in order to identify the fibres and the cellular regions. The details of the procedures for separating the fibres are described in the sections 3.1 and 3.2. Then the images are segmented in fibre and cellular areas as described in sections 3.3 and 3.4.

The quantitative properties of cellular and fibre regions are all saved in a metrics database file to be used for further analysis. The measured parameters are calculated after the image segmentation providing a dataset of 29 metrics. The details of these extracted metrics are described in table 2 . They are grouped into those calculated for the entire image, and those calculated for each segment, fibre and network that are identified by the analysis. For the segment metrics, a single metric value for the global image representing the average metric per segment type are also calculated. For example, the “SHG Fibre Intensity Mean" represents the average SHG signal intensity of all pixels within fibre segments.

3.1 | Fibre Network Extraction

Fibril networks may be considered graphs, with the fibres themselves as edges between n nodes at the regions they interconnect. A graph can be described by an adjoint matrix, an $n \times n$ matrix that records whether an edge is present between each node in a system. The degree of each node determines the number of edges it possesses.

In the literature, the FIRE (FIbeR Extraction) algorithm is designed to trace out a network on top of an image containing fibrous textures[39]. Here, we implement a modified version aiming to enhance the resolution in extracting the fibre details. The major change as compared to the FIRE algorithm is the use of a Sato filter[40] for the pre-processing, including a threshold for the edge length, r_0, and allowing to have edges between nodes.

Typically a filter can be applied to the input image prior to applying the FIRE algorithm in order to enhance any tubular-like regions. We apply a simple Sato or “tubeness filter”[40], based on the Hessian eigenvalues of each pixel, which has been shown[41] to have similar performance to the much more complex Curvelet transform (CT) method. A number of nucleation points is generated at each local maximum on our smoothed distance matrix that lies above a threshold value (typically 2 pixels). A set of parent nodes are then assigned to the pixel coordinates of these nucleation points. Propagation of subsequent child nodes is then performed by identifying further local maxima within a given search region of each parent node,

TABLE 2 Extracted metrics and their descriptions.

Name	Description	Group
No. Segments	Number of segments in image.	Global
No. Fibres	Number of fibres extracted in image.	Global
Area	Area of segments in pix^2.	Segment
Coverage	Ratio of pixel area covered by segments to total area of image.	Segment
Intensity Mean	Mean intensity of all image pixels in segments.	Segment
Intensity STD	Pixel intensity standard deviation of all pixels in segments.	Segment
Intensity Entropy	Pixel intensities average Shannon entropy of all pixels in segments.	Segment
Angle SDI	Ratio of the mean to max values in the pixel angle distribution.	Segment
Coherence	Coherence of structure tensor constructed from all pixels in the segment.	Segment
Local Coherence	Average coherence of structure tensors for each pixel in the segment.	Segment
Circularity	The circumference of a circle with area equals to the segment area divided by the perimeter of the segment	Segment
Eccentricity	The focal distance of an ellipsoid with the same second-moments as the segment divided by its major axis length	Segment
Fibre Length	Mean length of the extracted fibres.	Fibre
Fibre Waviness	Mean waviness (length over displacement) of the extracted fibre.	Fibre
Fibre Network Degree	Average number of edges in each node of the reduced FIRE network.	Network
Fibre Network Eigenvalue	Reduced network adjacency matrix maximum eigenvalue.	Network
Fibre Network Connectivity	Connectivity of the reduced network	Network
Fibre Network Cross-Link Density	Average number of cross links (intersection of edges) per extracted fibre in the network	Network

termed Local Maximum Points (LMPs). If an LMP is a successful candidate then a new child node will be created at its location, with an edge assigned to the parent node. The importance of using an edge length threshold r_0 is to allow us to both move the LMPs after their original assignment and also control the resolution at which each fibre can be traced without altering the size of the local LMP search. Finally, allowing edges to be created between any node, rather than just nodes on the same branch, aims to capture cross links within the network structure itself, rather than needing to infer them by identifying intersecting edges. Some cleaning is applied after all child nodes have been propagated in order to remove any artefact from the FIRE algorithm. In general, these aim to remove any redundant or unconnected nodes, as well as small (approx. 4 nodes) networks that are likely to correspond to noise.

3.2 | Fibre Assignment

In order to calculate the fibre properties we need to be able to obtain the individual fibres in the network. This is a non-trivial task, since it essentially involves deciding on which of the many possible paths in the network should be assigned as fibres. Therefore, for continuity, we attempt to apply a similar procedure as in FIRE algorithm[39] when deciding upon the most appropriate fibres.

We start by identifying an "external" parent node that only contains 1 edge as these must be the beginning or the end of any fibre. The single node in common to the parent is automatically added to the fibre and is assigned as their primary child node. The algorithm then traces back along any nodes connected to the active child node and rebuilds each fibre based on the LMP

criteria of the original FIRE algorithm that extracted the network (i.e. minimize edge lengths and inter-node angles). If it is not possible to locate a connecting node that lies within a given angle or distance from the propagating node, then this will be considered the end of the current fibre and the algorithm will restart on any remaining parent nodes that have not already been assigned to a fibre.

The process is illustrated in figure 1 . Note that as a consequence child nodes may be assigned to multiple fibres, but parent nodes may only be assigned to a single fibre.

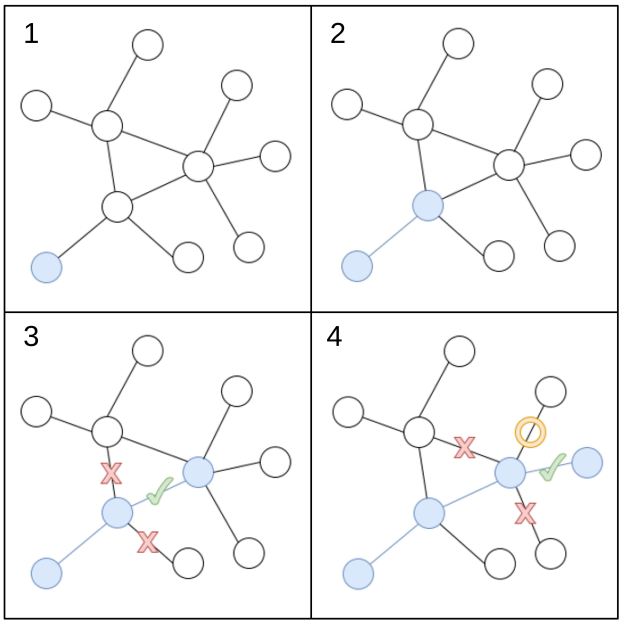

FIGURE 1 Diagram of the fibre assignment algorithm. 1: Identify of fibre by selecting an "external" parent node with only 1 edge. 2: Add primary child node with edge in common to parent. 3: Add further child nodes that meet original FIRE algorithm LMP criteria. If multiple nodes are possible, choose most optimal (i.e. shortest edge length, smallest inter-nodal angle). 4: Terminate fibre if child node is either "external" or no connected nodes meet FIRE algorithm LMP criteria.

3.3 | Colour Segmentation

The colour segmentation is based on a common method of segmentation for microscope-stained HE images included in software packages such as CurveAlign[42] that uses clustering of RGB values in a each color pixel vector. This can be considered a reliable technique since gram-straining yields cellular regions of a standard pigment.

Here we obtained a composite RGB image from the multi-photon signals in order to make use of such colour segmentation techniques. The greyscale intensity values from SHG, 2PEF and transmission data taken at a fixed biopsy region are transformed into an RGB image with relatively standardized pigments. In this method, the SHG image is important to obtain a better image segmentation, as it shows only signal for the collagen fibres.

Once the collagen fibre networks are extracted from the SHG image, a binary mask is created to represent the initial identification of fibre segments, based on the location of each collagen network. These segments are then used to weight regions within the SHG channel of the composite RGB image, so as to enhance their contribution in each pixel vector.

In order to enhance the clustering of colours and yield smooth segments we apply some smoothing techniques on the raw composite image that mimic those carried out by the CurveAlign software BDcreationHE algorithm. Segmentation is then performed by clustering of each pixel's RGB components, independent of its location. For computational efficiency we use the MiniBatchKMeans function as implemented in scikit-learn, which has been shown to achieve very similar performance to its full KMeans implementation[43]. It must be noted that KMeans is a stochastic algorithm, and therefore is not guaranteed to return the optimum solution at every run.

3.4 | Fibrous and Cellular Region Assignment

In order to deal with the stochastic nature of the KMeans algorithm we parameterized a function in color space $v(r, g, b, i)$ (representing the unit vector red, green and blue components and overall magnitude intensity) that represents the boundary between "fibrous" and "cellular" regions. Each cell cluster is then ranked according to the L1 distance of all centroids from $v(r, g, b, i)$, resulting a cost function that is inversely proportional to the average proximity of each centroid to the boundary. The KMeans run that creates the lowest average cost over a number of runs is then chosen as the optimal solution for our segmentation filter.

After each cluster has been assigned either a "fibrous" or "cellular" label, a number of extra checks are performed to review whether the segmentation process has yielded regions that would be expected from a biological perspective. In particular, segments below a threshold pixel area or mean intensity value are removed. This aims to minimize the impact of numerical artefacts that may have been introduced into our analysis pipeline.

3.5 | Linear discriminant analysis

The metrics obtained by the image analysis are used to try to distinguish between the diagnoses based on the LDA methodology. The LDA is a technique used to reduce the dimensionality of a large set of data by a projection of data into a new space in order to get a good data separability [44,45]. It is a supervised clustering method that considers the class labels (that is given by the diagnostic labels) to search for the axes that present the maximum data separation.

The 29 metrics extracted from the software form a 29-dimensional input space and we used the linear functions of the input vector as the decision surfaces. We first rescaled the measured features to show a Gaussian distribution between -1 and 1. After that, the LDA is performed on the dataset separating a training set and a test set. For that the whole dataset is split into 6 sets, 5 of them as training and 1 as test, considering a proportional representation of each cancer type by a stratified K-fold cross validation (CV) method. And this procedure is repeated 8 times to get an average accuracy value. Also to obtain the most relevant metrics in our LDA get models we applied a recursive feature elimination and cross-validation (RFECV). However as the optimal score is achieved only when considering more than 18 features, we present the LDA results always for the full dataset. The analyses were performed using the standard LDA library in Python.

3.6 | Random Forest analysis

As an alternative to LDA, we trained a random forest algorithm to identify samples with the presence of tumours, as well as to identify the tumour type. Random forest is a combined classifier that consists in an ensemble of decision trees where the output is the mode of the predicted classes [46]. It was chosen because it is particularly resistant to over fitting when dealing with a small dataset, such as is the case.

Before running random forest, we performed an exploratory data analysis to filter any correlated features from the 29 metrics extracted by the PyFibre. A correlation heat map was plotted, and we removed any features where $r \geq 0.98$. Each feature was scaled using a quantile-based transformation. For each feature, the original values are mapped to a uniform distribution via an estimate of its cumulative distribution function, then the obtained values are mapped to a Gaussian distribution using the quantile function. This process reduces the impact of any outliers. As in LDA, we applied a stratified K-fold cross-validation, splitting the cohort six times.

4 | RESULTS AND DISCUSSION

Figure 2 shows examples of the multiphoton images obtained from a biopsy sample. The measured area from a Gleason 3+3 tumour is indicated in the HE bright field image in (a) showing the acinar component randomly arranged between the stroma. The intensity map images of the 2PEF, in red, and SHG, in green, are presented in (b) and (c), respectively. The multiphoton images highlight the details of the acinar and stromal tumour components, the SHG image shows the collagen fibres and the 2PEF shows the emission of the eosin dye used to stain the tissue.

The automated image analysis allows to segment the images into the cellular (d) and collagen fibre content (e). Each connected segment is shown in a random colour. The extracted collagen fibre network, in red, and the extracted single fibres, in random colours, are shown in (f) and (g), respectively. A colour map of the collagen fibre alignment angles is shown in (h). The measurements of the degree of alignment of the collagen fibres and their arrangement patterns are based on the structure tensor analysis [29,47–52]. According to the proposed reactive stroma classification system, the biopsy region shown in (a) was classified as RS2, showing over 5% collagen fibres (light pink strands) in the background.

After segmenting the images to obtain the specific fibre and cellular regions, the measured parameters are obtained specifically for the pixels within each segment. The obtained values for the 29 metrics from all images are used for a comprehensive LDA and random forest analyses as presented in the following sections.

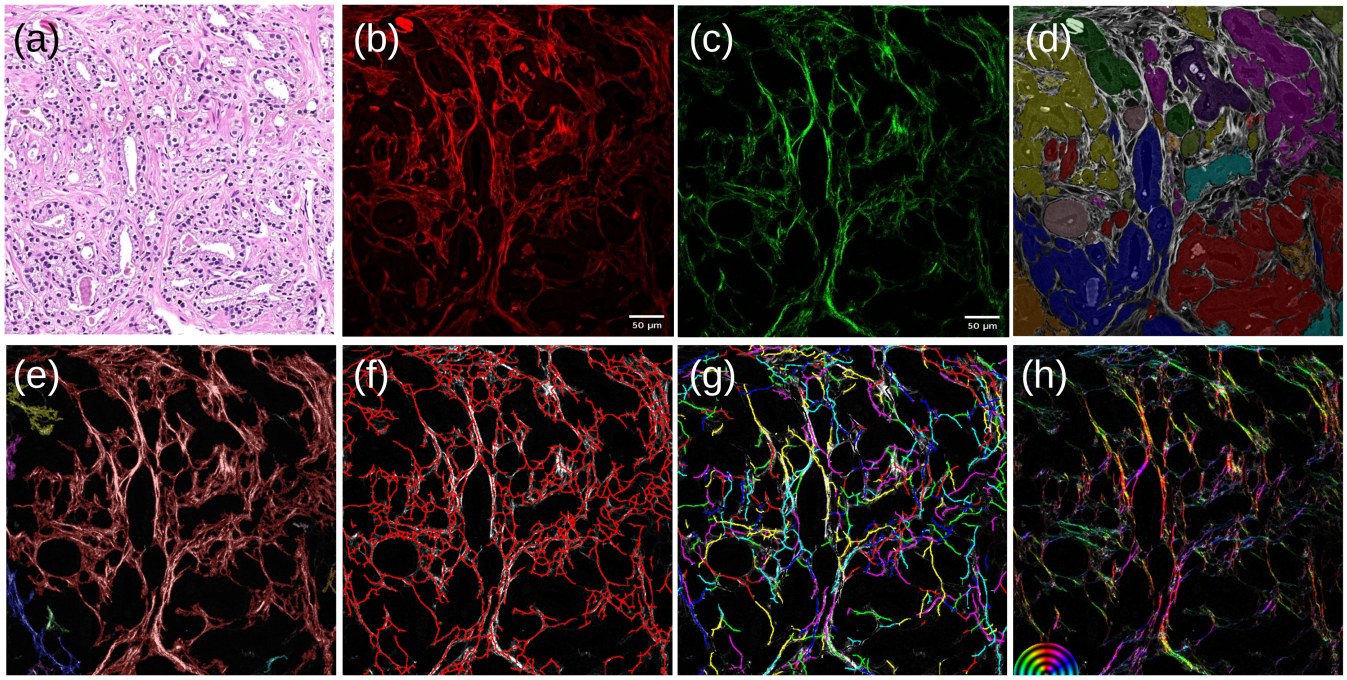

FIGURE 2 Acquired images: (a) HE bright field microscopy image, (b) 2PEF and (c) SHG. Images extracted by the software: (d) cellular segments and (e) fibre segments, each connected segment is shown in a random colour, and the cell and fibre segments are on top of the grey scale 2PEF and SHG intensity maps, respectively. (f) the collagen fibre network in red, and (g) the individual extracted fibres (each fibre is in a random colour), on top of the grey scale SHG intensity map. (h) the collagen fibre orientation angles (colour map scale on the left). The scale bar of 50 μm applies to all the extracted images and the HE image.

4.1 | LDA results

The LDA results considering the 29 extracted metrics are presented in Figure 3 (a) for all the Gleason diagnosis groups studied. The full symbols show the training dataset and the crosses show the test dataset for each classification class, with the colours as indicated by the colour bar. There are 57 images in the test dataset. It can be seen that there is a clear separation between the non-neoplastic tissue and the carcinoma samples. The model prediction for the separation of the non-neoplastic tissue and all the carcinoma samples presents an accuracy of $89 \pm 3\%$. However the average accuracy discrimination between all the Gleason score groups in Figure 3 (a) is only of $46 \pm 6\%$. Figure 3 (b) shows the cross validation score accuracy as the number of features included in the LDA. The maximum is achieved for a number of features of around 18, but as the curve is almost constant after that we kept all the features in the LDA graphics in (a).

The tables 3 and 4 present the multi-class classification results for each Gleason score diagnosis. The confusion matrix (tables 3) shows the predicted classes for the images in the test dataset (57 images). The obtained precision, sensitivity and specificity are quite good for the normal and indolent tumours (3+3) and also for the most aggressive ones (4+4+). Nonetheless the sensitivity for the 3 + 4 and 4 + 3 Gleason score classifications is quite poor. Those are indeed the most difficult ones to be diagnosed by the pathologists. A comparison for these classification showed a level of agreement of only 21.4% among 28 expert pathologists in a controlled comparative study[53,54]. Thus the achieved LDA separation is also comparable to the pathologists agreement.

The LDA results considering the groups provided by the stroma analysis are presented in Figure 4 . The intensity of RS, as described in the samples section 2.1, provides a much better discrimination. The collagen morphological changes are important for the cells migration and proliferation and may be a good cancer marker. The LDA results considering all the 29 obtained metrics are presented in Figure 4 for the three groups labeled as RS1, RS2 and RS3. The test dataset includes 50 images. The average accuracy discrimination between the three patterns is $65 \pm 5\%$, that is a considerably better result as compared with the Gleason score classification.

The tables 5 and 6 present the multi-class classification results for the RS groups. The precision, sensitivity and specificity are quite good for all the RS patterns.

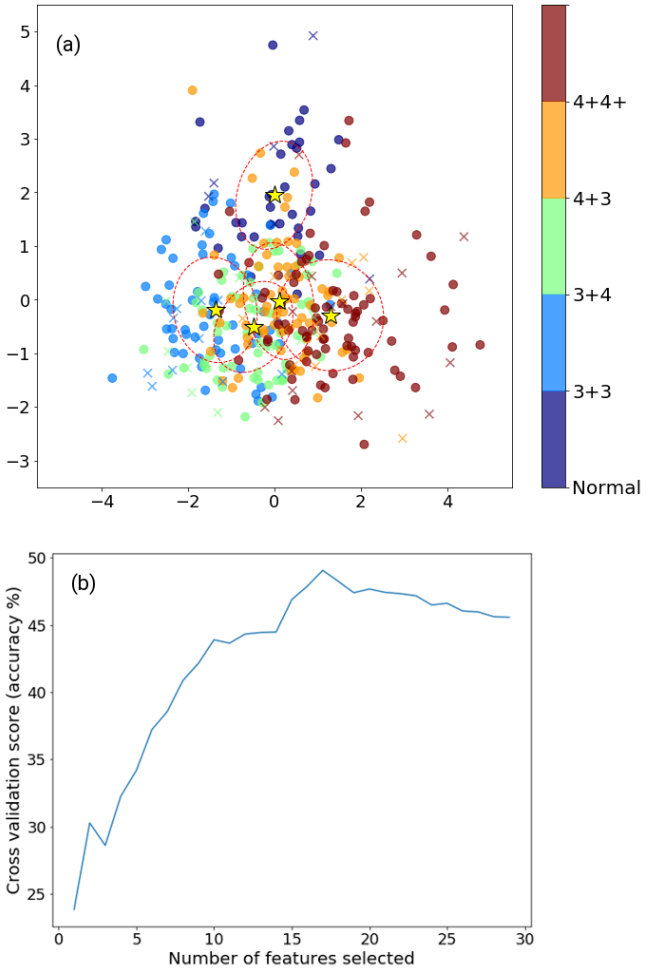

FIGURE 3 (a) LDA plots for the the cancer biopsies separated according to the Gleason score. The full symbols show the training dataset and the crosses show the test dataset for each type with the colours as indicated by the colour bar. The mean of each case group is marked by a yellow star and one standard deviation of the training dataset is shown by the ellipses. (b) Cross validation score as the number of features extracted from the image analysis.

We have also considered the observed stroma patterns to try to further separate the Gleason score groups. For the Gleason scores 3 + 4 and 4 + 3 the measured images in each Gleason score classification were divided into the reactive stroma RS1, RS2. For the 3 + 4 the measured regions are mainly RS2 (58 images), whereas RS1 was seen in only 14 images. The 4 + 3 group had 34 and 32 RS1 and RS2, respectively. The RS3 are mainly present in the high grade tumour group group 4 + 4+. Figure 5 shows the results for these divided groups. Note that they can can be seen as four distinct groups. Nonetheless the number of samples, especially for the 3 + 4 RS1 (14 images) is too small to perform the statistical calculations. However, these preliminary tests points to an average

TABLE 3 Confusion matrix: Gleason score classification.

		Predicted Class				
		Normal	3 + 3	3 + 4	4 + 3	4 + 4+
True Class	Normal	5	0	0	0	1
	3 + 3	0	10	1	0	0
	3 + 4	0	4	2	3	3
	4 + 3	0	1	0	4	8
	4 + 4+	1	1	0	1	12

TABLE 4 Statistical Analysis: Gleason score classification.

	Normal	3 + 3	3 + 4	4 + 3	4 + 4+
Precision	0.83	0.63	0.67	0.50	0.50
Sensitivity	0.83	0.91	0.17	0.31	0.80
Specificity	0.96	0.87	0.98	0.91	0.71

cross validation accuracy of 50 ± 10%. The average precision, sensitivity and specificity for all the groups are about 70%. These results indicate that a composed classification system that includes information about the intensity of the RS to the acini morphology based Gleason system could lead to a more precise diagnosis and prognosis approaches. It is known that the stromogenic tumours have a more aggressive behaviour.

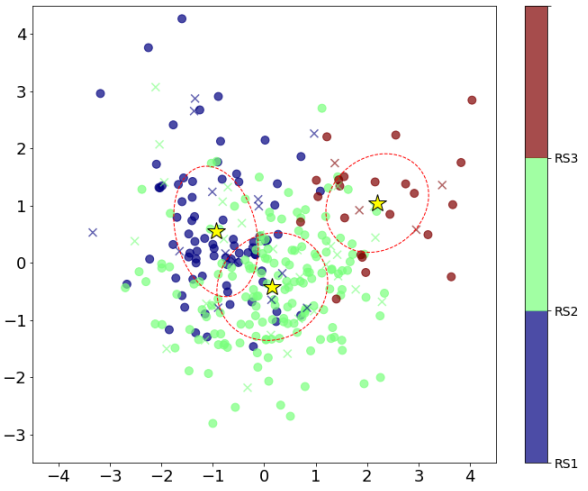

FIGURE 4 LDA plots for the the cancer biopsies separated by the stromal reaction patterns. The full symbols show the training dataset and the crosses show the test dataset for each type with the colours as indicated by the colour bar. The mean of each case group is marked by a yellow star and one standard deviation of the training dataset is shown by the ellipses.

TABLE 5 Confusion matrix: RS classification.

		Predicted Class		
		RS1	RS2	RS3
True Class	RS1	5	9	1
	RS2	7	21	3
	RS3	0	0	4

TABLE 6 Statistical Analysis: RS classification.

	RS1	RS2	RS3
Precision	0.42	0.70	0.50
Sensitivity	0.33	0.68	1.00
Specificity	0.80	0.53	0.91

Nonetheless, further clinical data studies would be necessary to better understand and strengthen these findings.

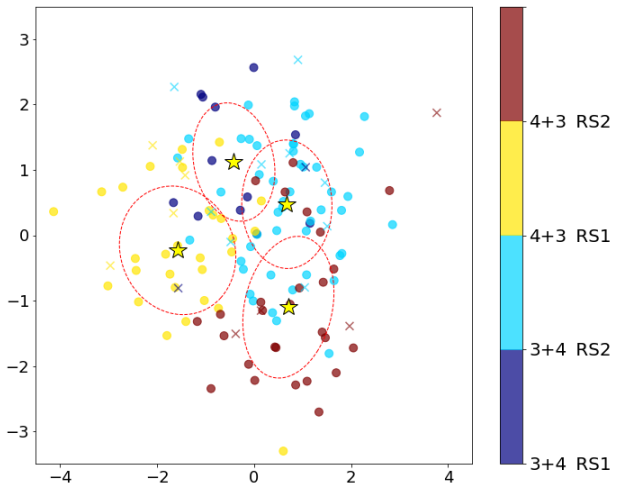

FIGURE 5 LDA plot for the stromal reaction within the 3+4 and 4 + 3 Gleason scores. The full symbols show the training dataset and the crosses show the test dataset for each type with the colours as indicated by the colour bar. The mean of each case group is marked by a yellow star and one standard deviation of the training dataset is shown by the ellipses.

4.2 | Random forest results

In order to further test the extracted parameters, we have also performed the random forest analysis considering the same datasets as in the LDA analysis. First, we analyzed the classified histological cancer types according to the Gleason system. We obtained discrimination between non-neoplastic and carcinoma samples with an accuracy of 91±3%, and discrimination between the four carcinoma groups of 36 ± 6%. In addition, for the intensity of the reactive stroma classification the overall discrimination accuracy is 62±6%. Upon general analysis, the random forest results are very similar to the LDA performance but under performs slightly, probably since it is a more complex model for a small dataset. The obtained confusion matrix and the statistical analyses provide very similar results as in the LDA ones, thus are not repeated here. The results reinforce that the set of metrics are robust parameters for machine learning sample classification methods and that the obtained classification are reliable.

The RF provides, in addition, the metrics importance rank for the groups classification. They provide interesting insights on the actual decisive factor of each metric for the model's decision process. The table 7 presents the obtained results, showing that the importance is quite well distributed over all the metrics. However, note that the extracted fibre metrics are the most important ones, both for the Gleason and the RS classification. In the LDA analysis, the cross validation score as the number of features in figure 3 (b), shows that the optimal number of features for the classification is around 18. From the random forest, within the 18 most important metrics, the features obtained by SHG analysis play a decisive role, 13 out of 18 features are from the fibre segment and fibre network metrics. These results reinforce that the reactive stroma parameters should indeed the considered for a more precise diagnosis.

5 | CONCLUSIONS

In summary we demonstrated the use of nonlinear imaging and automated image analysis procedures that could help improving cancer diagnosis. The SHG and 2PEF images allow measurements of the collagen and cellular regions of the biopsies that provide quantitative fibre and cellular parameters. These extracted parameters, after an automated image segmentation, formed a dataset that by LDA and random forest techniques achieved an excellent separation between non-neoplastic and tumour samples with accuracy larger than 90%. The discrimination for the Gleason score groups are similar to the agreement between pathologists. Furthermore, the results highlighted the importance of including the intensity of the reactive stroma as a further tool for improving diagnosis and prognostication. The average accuracy discrimination between

TABLE 7 Random forest metric importance percentage for the Gleason and RS classification.

Metric	Gleason (%)	RS (%)
No. Fibres	8.58	8.22
Circularity (cell)	5.78	6.42
Local Coherence (cell)	4.93	5.34
Intensity STD (fibre)	4.71	9.41
Angle SDI (cell)	4.68	1.91
Area (cell)	4.51	1.21
Eccentricity (cell)	4.46	2.99
Fibre Network Cross-Link Density	4.11	2.59
Intensity Entropy (fibre)	3.77	9.01
Intensity Mean (fibre)	3.61	3.46
Circularity (fibre)	3.58	2.97
Coherence (cell)	3.49	4.60
Local Coherence (fibre)	3.35	3.18
Fibre Length	3.15	4.43
SHG Angle SDI (fibre)	2.97	3.02
Fibre Waviness	2.93	1.70
Fibre Network Degree	2.77	2.48
Coverage (fibre)	2.74	0.75
Fibre Angle SDI	2.70	0.91
No. Cells	2.62	0.30
Fibre Network Eigenvalue	2.61	1.95
Fibre Segment Area	2.58	2.58
Fibre Segment Eccentricity	2.49	2.05
Fibre Network Connectivity	2.44	2.40
Intensity STD (cell)	2.40	1.65
Coherence (fibre)	2.32	4.52
Coverage (cell)	2.29	2.42
Intensity Mean (cell)	2.03	5.13
Intensity Entropy (cell)	1.38	2.40

the RS patterns is $65 \pm 5\%$. These findings reinforce the role of the reactive stroma on the progression of the prostate tumor, adding useful information for a better cancer classification. The automated procedure has the potential of being used as a high throughput image analysis tool, and we believe that it can be successfully applied to other primary cancer sites.

ACKNOWLEDGEMENTS

We acknowledge the technical support from Laboratório Analys Patologia in sample processing and the help with the diagnosis.

This work was financially supported by the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) and Institute of Science and Technology (INCT) in Carbon Nanomaterials, and by the Engineering and Physical Sciences Research Council, grant No. EP/G03690X/1 at the University of Southampton.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Jonathan N. Ouellette, Cole R. Drifka, Kelli B. Pointer, Yuming Liu, Tyler J Lieberthal, W John Kao, John S. Kuo, Agnes G. Loeffler, Kevin W. Eliceiri, *Bioengineering* **2021**, *8* (2).

[2] Xavier Catteau, Philippe Simon, Michel Jondet, Michel Vanhaeverbeek, Jean-Christophe Noël, *PLOS One* **2019**, *14* (3), 1–8.

[3] Gustavo Ayala, Jennifer A. Tuxhorn, Thomas M. Wheeler, Anna Frolov, Peter T. Scardino, Makoto Ohori, Marcus Wheeler, Jeffrey Spitler, David R. Rowley, *Clinical Cancer Research* **2003**, *9* (13), 4792–4801.

[4] G Falzon, S Pearson, R Murison, *Physics in Medicine & Biology* **2008**, *53* (23), 6641.

[5] Paolo P Provenzano, Kevin W Eliceiri, Jay M Campbell, David R Inman, John G White, Patricia J Keely, *BMC Medicine* **2006**, *4* (1), 38.

[6] Xiaoxing Han, Ryan M Burke, Martha L Zettel, Ping Tang, Edward B Brown, *Optics Express* **2008**, *16* (3), 1846–1859.

[7] Paolo P Provenzano, David R Inman, Kevin W Eliceiri, Justin G Knittel, Long Yan, Curtis T Rueden, John G White, Patricia J Keely, *BMC Medicine* **2008**, *6* (1), 11.

[8] Adam E Tuer, Serguei Krouglov, Nicole Prent, Richard Cisek, Daaf Sandkuijl, Kazuhiro Yasufuku, Brian C Wilson, Virginijus Barzda, *The Journal of Physical Chemistry B* **2011**, *115* (44), 12759–12769.

[9] Paul Campagnola, *Analytical Chemistry* **2011**, *83* (9), 3224–3231.

[10] Matthew W Conklin, Jens C Eickhoff, Kristin M Riching, Carolyn A Pehlke, Kevin W Eliceiri, Paolo P Provenzano, Andreas Friedl, Patricia J Keely, *The American Journal of Pathology* **2011**, *178* (3), 1221–1232.

[11] Visar Ajeti, Oleg Nadiarnykh, Suzanne M Ponik, Patricia J Keely, Kevin W Eliceiri, Paul J Campagnola, *Biomedical Optics Express* **2011**, *2* (8), 2307–2316.

[12] Kathleen Burke, Ping Tang, Edward B Brown, *Journal of Biomedical Optics* **2012**, *18* (3), 031106.

[13] Ara Ghazaryan, Halley F Tsai, Gor Hayrapetyan, Wei-Liang Chen, Yang-Fang Chen, Myung-Yung Jeong, Chang-Seok Kim, Shean-Jen Chen, Chen-Yuan Dong, *Journal of Biomedical Optics* **2012**, *18* (3), 031105.

[14] Kun Zhang, Callie A Corsa, Suzanne M Ponik, Julie L Prior, David Piwnica-Worms, Kevin W Eliceiri, Patricia J Keely, Gregory D Longmore, *Nature Cell Biology* **2013**, *15* (6), 677–687.

[15] Jeremy S Bredfeldt, Yuming Liu, Carolyn A Pehlke, Matthew W Conklin, Joseph M Szulczewski, David R Inman, Patricia J Keely, Robert D Nowak, Thomas R Mackie, Kevin W Eliceiri, *Journal of Biomedical Optics* **2014**, *19* (1), 016007.

[16] Javier Adur, Vitor B. Pelegati, Andre A. de Thomaz, Mariana O. Baratti, Diogo B. Almeida, L. A. L. A. Andrade, Fátima Bottcher-Luiz, Hernandes F. Carvalho, Carlos L. Cesar, *Plos One* **2012**, *7* (10), 1–13.

[17] Javier Adur, Vitor B Pelegati, Andre A de Thomaz, Mariana O Baratti, Liliana A L A Andrade, Hernandes F Carvalho, Fátima Bottcher-Luiz, Carlos Lenz Cesar, *Journal of Biophotonics* **2014**, *7* (1-2), 37–48.

[18] Xiufeng Wu, Gang Chen, Jianping Lu, Weifeng Zhu, Jingting Qiu, Jianxin Chen, Shusen Xie, Shuangmu Zhuo, Jun Yan, *PLOS One* **2013**, *8* (6), e65933.

[19] Mariano Bizzarri, Alessandra Cucina, *BioMed Research International* **2014**, *2014*.

[20] Yuankai K Tao, Dejun Shen, Yuri Sheikine, Osman O Ahsen, Helen H Wang, Daniel B Schmolze, Nicole B Johnson, Jeffrey S Brooker, Alex E Cable, James L Connolly, et al., *Proceedings of the National Academy of Sciences* **2014**, *111* (43), 15304–15309.

[21] Anders Brabrand, Ian I Kariuki, Monica J Engstrøm, Olav A Haugen, Linda A Dyrnes, Bjørn O Åsvold, Magnus B Lilledahl, Anna M Bofin, *APIMIS* **2015**, *123* (1), 1–8.

[22] Wai Jin Tan, Jie Yan, Shuoyu Xu, Aye Aye Thike, Boon Huat Bay, Hanry Yu, Min-Han Tan, Puay Hoon Tan, *Journal of Clinical Pathology* **2015**, *68* (12), 1033–1035.

[23] Karissa Tilbury, Paul J Campagnola, *Perspectives in Medicinal Chemistry* **2015**, *7*, 21–32.

[24] Ahmad Golaraei, Lukas Kontenis, Richard Cisek, Danielle Tokarz, Susan J Done, Brian C Wilson, Virginijus Barzda, *Biomed. Opt. Express* **2016**, *7* (10), 4054–4068.

[25] Matthew S Hall, Farid Alisafaei, Ehsan Ban, Xinzeng Feng, Chung-Yuen Hui, Vivek B Shenoy, Mingming Wu, *Proceedings of the National Academy of Sciences* **2016**, *113* (49), 14043–14048.

[26] Karin Wang, Fei Wu, Bo Ri Seo, Claudia Fischbach, Weisi Chen, Lauren Hsu, Delphine Gourdon, *Matrix Biology* **2017**, *60–61*, 86–95.

[27] Ashley Case, Becky K. Brisson, Amy C. Durham, Suzanne Rosen, James Monslow, Elizabeth Buza, Pascale Salah, Julie Gillem, Gordon Ruthel, Sridhar Veluvolu, Veronica Kristiansen, Ellen PurÃĪ, Dorothy C. Brown, Karin U. Sørenmo, Susan W. Volk, *PLOS ONE* **2017**, *12* (7), 1–19.

[28] Matthew W Conklin, Ronald E Gangnon, Brian L Sprague, Lisa Van Gemert, John M Hampton, Kevin W Eliceiri, Jeremy S Bredfeldt, Yuming Liu, Nuntida Surachaicharn, Polly A Newcomb, Andreas Friedl, Patricia J Keely, Amy Trentham-Dietz, *Cancer Epidemiol Biomarkers Prev.* **2018**, *27* (2), 138–145.

[29] Rodrigo A Natal, José Vassallo, Geisilene R Paiva, Vitor B Pelegati, Guilherme O Barbosa, Guilherme R Mendonça, Caroline Bondarik, Sophie F Derchain, Hernandes F Carvalho, Carmen S Lima, et al., *Tumor Biology* **2018**, *40* (4), 1010428318770953.

[30] Rodrigo Andrade Natal, Geisilene R Paiva, Vitor B Pelegati, Ludwing Marenco, César A Alvarenga, Renato F Vargas, Sophie F Derchain, Luis O Sarian, Camille Franchet, Carlos L Cesar, et al., *Scientific Reports* **2019**, *9* (1), 7715.

[31] Sara L. Best, Yuming Liu, Adib Keikhosravi, Cole R. Drifka, Kaitlin M. Woo, Guneet S. Mehta, Marie

Altwegg, Terra N. Thimm, Matthew Houlihan, Jeremy S. Bredfeldt, E. Jason Abel, Wei Huang, Kevin W. Eliceiri, *BMC Cancer* **2019**, *19* (1), 490.

[32] L Gole, J Yeong, J C T Lim, K H Ong, H Han, A A Thike, Y C Poh, S Yee, J Iqbal, W Hong, B Lee, W Yu, P H Tan, *Breast Cancer Research* **2020**, *22* (1), 42.

[33] Yuting Ling, Chunhui Li, Kairui Feng, Scott Palmer, Paul L. Appleton, Stephen Lang, David McGloin, Zhihong Huang, Ghulam Nabi, *Journal of Biophotonics* **2016**, 1–8.

[34] Aloísio Miguel Garcia, Felipe Lima Magalhães, Jaqueline S Soares, Eduardo Paulino Junior, Mário F R de Lima, Marcelo Mamede, Ana M de Paula, *Biomedical Physics & Engineering Express* **2017**, *4* (2), 025026.

[35] Jonathan I Epstein, Michael J Zelefsky, Daniel D Sjoberg, Joel B Nelson, Lars Egevad, Cristina Magi-Galluzzi, Andrew J Vickers, Anil V Parwani, Victor E Reuter, Samson W Fine, James A Eastham, Peter Wiklund, Misop Han, Chandana A Reddy, Jay P Ciezki, Tommy Nyberg, Eric A Klein, *European Urology* **2016**, *69* (3), 428 – 435.

[36] Ina P Pavlova, Sujit S Nair, Dara Lundon, Stanislaw Sobotka, Reza Roshandel, Patrick-Julien Treacy, Parita Ratnani, Rachel Brody, Jonathan I Epstein, Gustavo E Ayala, et al., *Journal of personalized medicine* **2021**, *11* (11), 1061.

[37] Luana A. Reis, Ana P. V. Garcia, Egleidson F. A. Gomes, Francis G. J. Longford, Jeremy G. Frey, Geovanni D. Cassali, Ana M. de Paula, *Biomed. Opt. Express* **2020**, *11* (11), 6413–6427.

[38] Francis G.J. Longford, *PyFibre: Python Fibrous Image Analysis Toolkit*, https://github.com/franklongford/PyFibre, Version 2.1.1, (2022).

[39] Andrew M Stein, David A Vader, Louise M Jawerth, David A Weitz, Leonard M Sander, *Journal of Microscopy* **2008**, *232* (3), 463–475.

[40] Yoshinobu Sato, Shin Nakajima, Nobuyuki Shiraga, Hideki Atsumi, Shigeyuki Yoshida, Thomas Koller, Guido Gerig, Ron Kikinis, *Medical Image Analysis* **1998**, *2* (2), 143 – 168.

[41] Jeremy S Bredfeldt, Yuming Liu, Matthew W Conklin, Patricia J Keely, Thomas R Mackie, Kevin W Eliceiri, *Journal of Pathology Informatics* **2014**, *5* (1), 28.

[42] Y. Liu, A. Keikhosravi, G.S. Mehta, C.R. Drifka, K.W. Eliceiri in *Fibrosis: Methods in Molecular Biology*, *Vol. 1627*, L. Rittié (Ed.), New York, NY: Springer, **2017**, pp. 429–451.

[43] D Sculley, in *Proceedings of the 19th International Conference on World Wide Web*, **2010**, pp. 1177–1178.

[44] Aleix M Martínez, Avinash C Kak, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2001**, *23* (2), 228–233.

[45] Keinosuke Fukunaga, *Statistical pattern recognition*, World Scientific, **1993**.

[46] B Dai, R Chen, S Zhu, W Zhang, in *2018 International Symposium on Computer, Consumer and Control (IS3C)*, **2018**, pp. 449–452.

[47] Bernd Jahne, *Digital image processing*, *Vol. 4*, Springer Berlin, **2005**.

[48] Francisco J Ávila, Oscar del Barco, Juan M Bueno, *Journal of Biomedical Optics* **2015**, *20* (8), 086001.

[49] Francisco J Ávila, Oscar del Barco, Juan M Bueno, *Journal of Biomedical Optics* **2016**, *21* (6), 066015.

[50] Francisco J Ávila, Oscar Del Barco, Juan M Bueno, *Journal of Optics* **2017**, *19* (10), 105301.

[51] Francisco J Ávila, Juan M Bueno, *Applied Optics* **2015**, *54* (33), 9848–9854.

[52] Rana Rezakhaniha, Aristotelis Agianniotis, Jelle Tymen Christiaan Schrauwen, Alessandra Griffa, Daniel Sage, CVC vd Bouten, FN Van de Vosse, Michaël Unser, Nikolaos Stergiopulos, *Biomechanics and Modeling in Mechanobiology* **2012**, *11* (3-4), 461–473.

[53] Lars Egevad, Brett Delahunt, Daniel M Berney, David G Bostwick, John Cheville, Eva Comperat, Andrew J Evans, Samson W Fine, David J Grignon, Peter A Humphrey, et al., *Histopathology* **2018**, *73* (1), 8–18.

[54] Derek J Lomas, Hashim U Ahmed, *Nature reviews Clinical oncology* **2020**, *17* (6), 372–381.