Automated Classification of Breast Cancer Stroma Maturity From Histological Images

Sara Reis*, Patrycja Gazinska, John H. Hipwell, Thomy Mertzanidou, Kalnisha Naidoo, Norman Williams, Sarah Pinder, and David J. Hawkes

Abstract—**Objective: The tumor microenvironment plays a crucial role in regulating tumor progression by a number of different mechanisms, in particular, the remodeling of collagen fibers in tumor-associated stroma, which has been reported to be related to patient survival. The underlying motivation of this work is that remodeling of collagen fibers gives rise to observable patterns in hematoxylin and eosin (H&E) stained slides from clinical cases of invasive breast carcinoma that the pathologist can label as mature or immature stroma. The aim of this paper is to categorise and automatically classify stromal regions according to their maturity and show that this classification agrees with that of skilled observers, hence providing a repeatable and quantitative measure for prognostic studies. Methods: We use multiscale basic image features and local binary patterns, in combination with a random decision trees classifier for classification of breast cancer stroma regions-of-interest (ROI). Results: We present results from a cohort of 55 patients with analysis of 169 ROI. Our multiscale approach achieved a classification accuracy of 84%. Conclusion: This work demonstrates the ability of texture-based image analysis to differentiate breast cancer stroma maturity in clinically acquired H&E-stained slides at least as well as skilled observers.**

Index Terms—**Breast cancer, histopathology, image classification, stroma maturity.**

I. INTRODUCTION

THE histopathological evaluation of a biopsy or surgical specimen is considered the gold standard for breast cancer diagnosis, enabling malignancy to be confirmed, the nature of the tumour established and the distribution and extent of the disease determined. Invasive breast carcinoma (IBC) is the most common malignancy to affect women worldwide [1]. Invasive carcinomas are morphologically subdivided according to their growth patterns and degree of differentiation, which in turn indicates how much they resemble normal breast epithelial cells. Routine pathologic reporting of IBC includes the assessment of histological type and grade. The majority (60% to 75%) of invasive breast cancers are classified as ductal/no special type (NST), whilst the remainder of sub-type (e.g, lobular, tubular, mucinous etc) show distinct prognostic significance but are less common [1]. Tumour grade is based on the degree of differentiation of the tumour tissue and can be applied to all types of cancer. It is a composite, semi-quantitative score of tubule formation, nuclear pleomorphism/atypia and mitotic activity in the malignant epithelium [2]. A higher grade implies a more aggressive tumour, with a poorer prognosis.

The stroma compartment, i.e. the mesenchymal tissue surrounding the tumour, has a major role in the control and regulation of physiological processes. Stromal regions can be characterised according to their maturity. For certain types of cancers, as the tumour develops its microenvironment undergoes dynamic changes, such as increased production of fibrous tissue (desmoplasia) with abundant fibroblasts, which results in what is defined as immature stroma. Other cancers do not induce a desmoplastic reaction [3]. In these cases, the stroma is primarily composed of mature collagen fibres.

Stroma assessment can be useful for diagnosis. For instance, tubular cancers can overall show similar characteristics to benign lesions, such as radial scars [1]. However, the presence of immature stroma surrounding the tubular structures is a useful feature to confirm malignancy. Another particular case where pathologists might assess the stromal component is when determining the type of cancer. For example, lobular cancers do not normally present with a stromal reaction [4].

Ueno *et al.* [5] established a histological categorization of fibrotic stroma to reflect its biological behaviour in advanced rectal cancer. They classified stromal tissue into three categories: mature, intermediate and immature. Their analysis of stromal regions from a dataset of 862 patients found that five- and ten-year survival rates were poorest in the group with immature stroma and best in those with mature stroma. Other authors use the term "reactive" stroma to describe stromal areas enriched with fibroblasts and myofibroblasts within the tumour compartment. These are also characteristics of immature stroma. They

suggested that this reactive stroma is a common feature and an intrinsic property of some tumours [6], [7].

Recent evidence has shown that the tumour microenvironment plays a crucial role in regulating tumour progression by different mechanisms, such as suppression of the immune response and remodelling of the extracellular matrix (ECM) [8]. In particular, the remodelling of collagen fibres at the tumour-stromal interface has been shown to adversely affect survival [9]. Moreover, increased stromal density is directly correlated with a greater risk of developing breast cancer [10].

There has been a continuous search to find better image-based biomarkers that are indicative of disease and can be used to predict patient outcome or stratify patients with breast cancer [11]. Various genetic studies have revealed distinct stromal gene expression signatures in the human breast. They also found that stromal expression could be used to stratify and predict disease outcome independently of other clinical prognostic factors [3], [6], [12]. These observations suggest that stroma influences tumour initiation, growth and progression.

With recent advances in digital histopathology, quantitative analysis of haematoxylin and eosin (H&E) images is now feasible. Consequently there has been much interest in the development of methods for tissue classification (malignant versus normal regions), object segmentation (nuclei and mitosis count) and grading [13]–[16]. These include the application of statistical (gray-level co-occurrence matrix, local binary patterns), structural, model-based and transform-based methods (Fourier, Gabor, Difference-of-Gaussians and wavelet transforms) for extracting textural image features.

In recent years, there has been a growing interest in the changes that occur in the tumour microenvironment, in particular in the stroma, and their relationship to cancer progression. Beck *et al.* [13] found that the best histological predictors of patient survival were not from the carcinoma itself, but from the adjacent stromal tissue. Out of 6642 features extracted from H&E stained microarray (TMA) sections of breast carcinoma, eleven were predictive of 5-year survival, with three being stromal morphological features. However, there has not been any specific research into the classification and quantification of different breast stroma types and their prognostic value in breast cancer. An in-depth study of stromal features could reveal useful information on tumour progression and prognosis.

Previous studies have observed a link between collagen alignment and patient outcome, however the majority of these studies have been performed using second harmonic generation (SHG) images [9], [17], [18]. This technique produces high contrast images of the collagen fibres and extracellular matrix of H&E stained tissue slides. To the best of the authors' knowledge, this is the first study that attempts to categorise and automatically classify stromal regions in clinically acquired H&E stained slides. This will enable examination of the hypothesis that stromal maturity is related to growth and metastatic potential of tumours, and thus can be used for prognosis.

These stromal features could also aid in the interpretation of the radiological signal and in particular relating the radiological features to microscopic changes. Variations in the stroma may cause detectable changes in water mobility (diffusion MRI [19]) and mechanical properties (Shear-Wave Elastography [20]). This, in turn, could enhance the ability of non-invasive pre-operative imaging to predict prognosis and also treatment outcomes (e.g. response to primary chemotherapy).

In previous work [21], we proposed an algorithm based on a support vector machine (SVM) classifier applied to a set of quantitative texture features to automatically classify stromal regions from images of H&E sections according to their maturity. Derivative-of-Gaussians (DtG) have been used for many applications, such as edge detection, retinal blood vessels extraction or texture analysis [22]–[24]. In this paper, we demonstrate the use of multi-scale Basic Image Features (BIF) and Local Binary Patterns (LBP) in combination with random decision trees classifier can be used for classification of breast cancer stroma. The computation of BIFs involves classifying the output obtained from convolution of an image with a bank of DtG filters into one of seven categories. These categories correspond to distinct local image structures, as defined by local symmetries [25].

BIFs have been shown to out-perform alternative methods when applied to standard textures [25] and applied to the segmentation of phase contrast microscopy (PCM) images of embryonic stem cells [26] and as a measure to characterise mammographic features according to their orientation with respect to the nipple [27].

In the following sections we describe our methodology and present results from a cohort of 55 patients diagnosed with IBC, where analysis was done in 169 regions-of-interest (ROI).

II. MATERIALS AND METHODS

A. Patients and ROI Selection

The images used in this study were acquired from cases in the King's Health Partner's Cancer Biobank. A total of 55 H&E stained slides from a series of patients diagnosed with IBC were chosen. The sections, stained according to standard protocols were scanned using a Hamamatsu NanoZoomer 2.0-HT (Hamamatsu, Japan) at an objective magnification of x20 with a resolution of 0.46 μm per pixel.

Stromal regions were identified and manually traced on the digitised images. The number of annotated regions per slide varied between two and five, depending on the amount of stroma tissue in the sample that was available for classification (i.e. stroma-rich tumours tend to have more regions selected than stroma-poor tumours, see Appendix).

The chosen stromal regions were then assessed by an experienced breast pathologist (SEP) and a trainee pathologist (KN). For that purpose, a graphical user interface (GUI) was designed as follows: two ROIs were randomly paired and displayed simultaneously and classified as either mature or immature by the observer. This was repeated until every ROI was shown twice, ensuring that no image pair was repeated. We displayed every ROI twice to assess intra-observer agreement. Whilst a forced choice (pair-wise ranking) experiment was considered, the combinatorial scale of the resulting experiment was prohibitive and therefore not adopted. The present study design was therefore chosen to minimise the observer's workload whilst retaining the self-regulating benefit of pair-wise comparison.

The rating was performed on a continuous scale, and ranged from -1 to 1, where -1 corresponds to "highly immature" and 1 to "highly mature". We adopted a continuous scale following discussions with the pathologist - stromal changes are gradual and there can be different levels/stages of maturity present around the tumour. In the Appendix section we provide examples of ROIs and the corresponding scores provided by the pathologist. A total of 179 ROIs were rated. Images that were given a score of zero or that showed disagreement were reviewed by the pathologist and given a final maturity score. Ten ROIs were excluded from the final dataset due to being unsuitable for the purpose of this experiment. These include DCIS regions, regions of old scarring that were initially confused with stroma, ROIs that were too small or the area shown had little stroma to assess; ROIs that were scored zero by the pathologist). After this exclusion, a total of 169 ROIs were annotated according to their maturity (mature = 60; immature = 109).

The intra-observer variability was tested by computing the intraclass correlation coefficient (ICC) and the inter-observer variability was computed by calculating the simple arithmetic mean of the duplicate observations from each observer and then computing the Pearson correlation coefficient. ICC was computed using the R package psych [28], and a two-way model was applied. These results are shown in more detail in the Appendix. The ICC calculated to determine agreement within each observer was 0.91 (SEP) and 0.68 (KN), while the between-observer correlation coefficient was $R^2 = 0.67$.

The dimensions of the annotated regions had an average area of 285 μm^2 (the maximum area selected was 1661 μm^2 and the minimum was 23 μm^2) and were selected in collaboration with a technician on a qualitative basis. Background, non-stromal regions (such as clusters of tumour cells or adipose cells) were removed by creating a binary mask based on the pathologist's annotations.

B. Criteria for Histological Characterisation of Breast Cancer Stroma

Fig. 1 shows two representative examples of mature (a) and immature (b) stroma. Breast cancer stroma was classified as mature when composed of highly organised fine, elongated collagen fibres, normally following the same orientation with elongated, flattened nuclei of fibrocytes wrapped into its multiple layers. Stroma was classified as immature when consisting of randomly oriented fibres surrounded by a high concentration of fibroblasts and exhibiting regions with oedema [5]. The initial goal of this experiment was to classify distinctive stromal regions that present the features explained above. Thus, areas comprising high levels of inflammatory cells were excluded from this experiment.

C. Methods

A schematic summarising the image-processing pipeline is shown in Fig. 2. We evaluated two approaches for extracting stromal texture descriptors: single- and multi-scale LBP and the combination of the Light Lines filter response (computed from the BIFs image classification) image with local binary patterns.
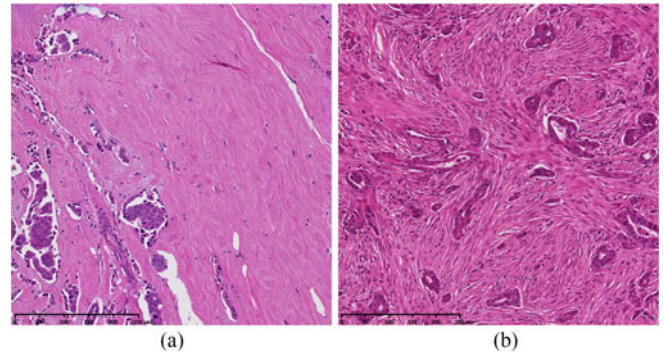

Fig. 1. Example images of (a) mature and (b) immature stroma tissue. Image magnification: x20. Mature stroma presents highly organised fine, elongated collagen fibres, normally following the same orientation with elongated, flattened fibrocyte nuclei. Immature stroma consists of randomly oriented fibres surrounded by a high concentration of fibroblasts and regions of oedema.

Before feature extraction and to reduce stain variation across different ROIs, the H&E images were colour-normalised using nonlinear mapping of a source image to a target image using a representation derived from colour deconvolution and later converted to grey-scale luminance by pixel-wise computation of a linear combination of the red, green and blue channels [29].

1) Basic Image Features: Computation of BIFs consists of classifying the output obtained from convolution of an image with a bank of DtG filters into one of seven categories. Each category corresponds to distinct local image structures based on local symmetry types: slopes, radially symmetrical dark and bright blobs, dark and light lines, saddle points and flat regions (no structure) [25], [30]. c_{ij} is defined as the response of the convolution of an image I with one of the DtG filters, where i and j represent the order in the x and y directions, respectively. The scale-normalised response is computed using the following equation:

$$s_{ij} = \sigma_B^{i+j} c_{ij} \tag{1}$$

where σ_B is the scale (standard deviation) of the DtG filters. Subsequently, (2) and (3) are computed: λ corresponds to the image Laplacian and γ is a term that measures the variance over directions of the second directional derivative.

$$\lambda = s_{20} + s_{02} \tag{2}$$

$$\gamma = \sqrt{(s_{20} + s_{02})^2 + 4s_{11}^2} \tag{3}$$

These equations are computed for each individual pixel and based on the largest of εc_{00}, $2\sqrt{s_{10}^2 + s_{01}^2}$, λ, $-\lambda$, $2^{-\frac{1}{2}}(y - \lambda)$, $2^{-\frac{1}{2}}(y + \lambda)$ and γ, pixels are then classified into one of the seven categories, resulting in the final BIFs image I_B. ε controls what amplitude of structure is tolerated before a region is considered to have no structure (i.e. this is the "flat" region noise threshold).

2) Combining BIFs With Local Binary Patterns:

a) Light-lines DtG filter: Consultation with a experienced breast pathologist (SEP) concluded that one of the primary features for defining stroma maturity is its alignment and linear coherence, and we inferred that the use of only one of
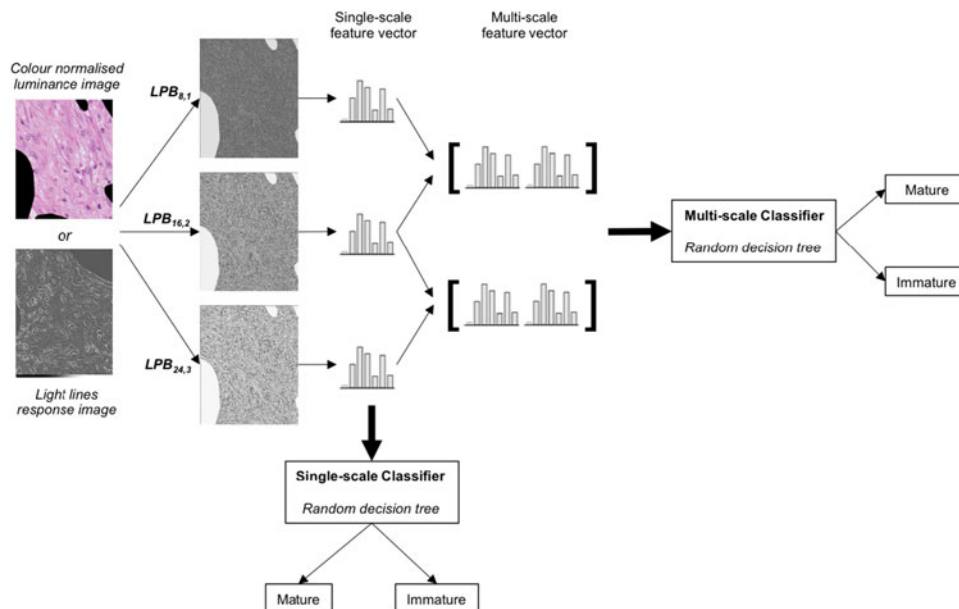
Fig. 2. Overview of the image-processing pipeline. Two approaches are evaluated: LBP with the colour normalised luminance image as an input, and LBP in combination with the light lines filter. Single- and multi-scale feature vectors are constructed at various scales and LBP neighbourhoods.
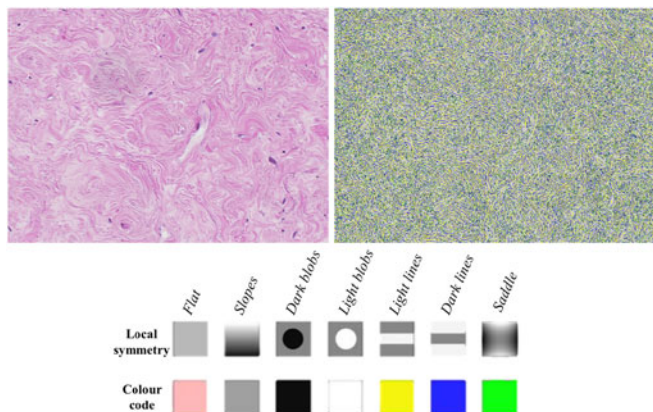

Fig. 3. The result of convolving an H&E image with a bank of DtG at different scales results in a BIFs image where each pixel is classified in one of seven categories.

the categories from the BIFs classification - the light lines filter response - was sufficient for describing stroma maturity, as per our previous findings [21]. A representative image can be seen in Fig. 4. We propose the combination of the light lines filtering computed at multiple scales, with the local binary patterns operator to capture the spatial representation of the image and thus enable the classification.

b) Local Binary Patterns: LBP is a local texture descriptor that captures the proportion of the micro patterns in the texture of an image [31]. The LBP operator compares each pixel, g_c, in an image to P pixels, g_p, in a neighbourhood with radius R and records whether their intensity is greater than that of g_c:

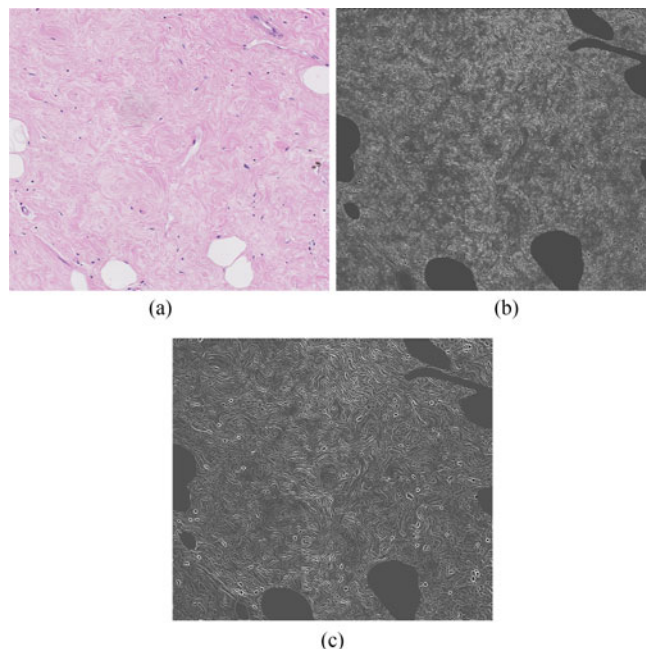$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} 2^p s(g_p - g_c) \tag{4}$$

Fig. 4. The digitised H&E section is convolved with a bank of DtG kernels and each pixel is classified in one of seven categories. Because we are interested in the linear structure of the stroma, we only take into account the light lines category and use that as the input for texture feature extraction. (a) Original H&E region of interest; (b) resulting lightlines image at $\sigma_2 = 0.5$ μm and (c) $\sigma_5 = 1$ μm.

where $s(x)$ is:

$$s(x) = \begin{cases} 1, & \text{if} x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

The results from the neighbours are used to construct a binary code. Pixels which have a value less than the value of the central
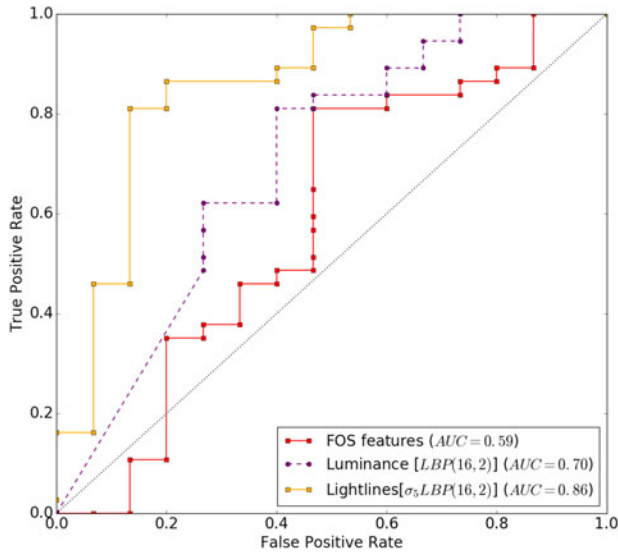
Fig. 5. ROC curves for FOS, Luminance and Lightlines/LBP single-scale classifiers that performed the best. The Lightlines/LBP classifier area under the ROC curve was 0.86 ($\sigma_5 = 2$ and $\mathrm{LBP}_{(16,2)}$), while the Luminance classifier was 0.70 ($\mathrm{LBP}_{(16,2)}$) and the FOS classifier 0.59.

pixel, g_c, are set to 0, and the pixels that have a larger or equal value are set to 1. The binary codes are translated into an integer and the histogram of these values is used to represent the texture of the image.

The original LBP was defined in a rectangular 3×3 pixel neighbourhood. Ojala *et al.* [31] introduced the circular neighbourhood and applied it to different neighbourhood sizes. By changing the radius of a circular neighbourhood (R) and the number of pixels along the circle (P), the LBP feature $\mathrm{LBP}_{P,R}$ can then be retrieved at different scales. Another addition to the original LBP operator is the use of minimised uniform patterns to achieve rotation invariance and reduce the length of the feature vector, denoted as $\mathrm{LBP}_{P,R}^{riu2}$. Some binary patterns occur more commonly in texture images than others. Patterns are called uniform if they contain at most two transitions on a circular ring from 1 to 0 or vice versa. Examples such as 00000000 (0 transitions) and 01110000 (2 transitions) are considered uniform, while 11001001 (4 transitions) and 01010010 (6 transitions) are not. During computation of the LBP labels, each uniform pattern has its own separate label whereas all the non-uniform patterns are labelled within a single label.

c) Multi-Scale approach: Lightlines computation of the input image were computed at a range of scales. The local spatial structure of the resulting images was then captured using $\mathrm{LBP}_{P,R}^{riu2}$. The multi-scale feature vector was constructed by concatenation of the LBP histograms obtained across the scales considered. For this experiment, combinations of two light line filter scales were included. Experiments showed that increasing the combinations of scales to three or more did not yield better performance.

D. Classification

Using the proposed framework, two sets of feature vectors were created from the original set of H&E stromal regions using

both single- and multi-scale approaches. Every combination of scales was tested but only the best results obtained for each category after parameter value optimisation are reported in the next section. The data was divided into training (70%) and test (30%) sets. The division into training and test set was done on a patient level in order to prevent ROIs from the same patient being present in both training and test sets and thus introducing bias to the classifier. From the 55 patients, 39 were included in the training set and 16 in the test set. A random forest (RF) classifier was trained and its optimal parameters were chosen via 10-fold cross validation of the training set. The pre-trained RF classifier was then used to classify the images from the test set. The output of the classifier was a binary label, with 1 for immature and 0 for mature. Scikit-learn, an open source library for Python was used for this purpose [32]. Classification performance was evaluated by comparison of the classifiers output with ground truth labels created by the pathologist. Furthermore, qualitative comparisons were made between single- and multi-scale approaches by the means of percentage of incorrectly classified images (using the metrics precision, recall and F1-score) and area under the ROC curve.

III. RESULTS

The performance of single- and multi-scale RF classifiers were evaluated on the test set of 52 (30% of the total available data) stromal ROI. Of these images, 15 (21%) represented mature stroma and 37 (9%) immature stroma.
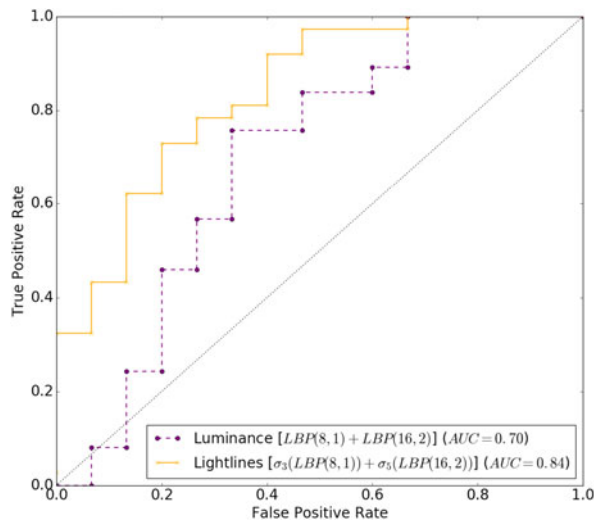
A. Single-Scale Approach

The first experiment evaluated the performance of our proposed method using the single-scale approach. Combination of scales (for the lightlines filter) with various LBP neighbourhoods were tested in order to find which combination gave the best performance. Five scales were individually tested ($\sigma_1 = 0.25, \sigma_2 = 0.5, \sigma_3 = 0.75, \sigma_4 = 1$ and $\sigma_5 = 2$) in combination with three LBP neighbourhoods ($\mathrm{LBP}_{(8,1)}$, $\mathrm{LBP}_{(16,2)}$ and $\mathrm{LBP}_{(24,3)}$). We have also previously tested our method against features based on first-order statistics (FOS). These features were the mean, standard deviation, integrated density, skewness and kurtosis [33]. Fig. 5 presents the ROC curves for the best single-scale classifiers. Table I summarises the main classification metrics computed to evaluate the classifiers performance. Combination of the lightlines filter with LBP achieved higher performance (area under the ROC curve (AUC = 0.86) than using LBP on their own (AUC = 0.70) or simple FOS features (AUC = 0.59), resulting in an improvement of 16% and 28% in accuracy, respectively.

B. Multi-Scale Approach

In the second experiment we evaluated a multi-scale approach by combining different lightlines filter scales over various LBP neighbourhoods. Fig. 6 presents the ROC curves for the two best multi-scale classifiers. As in the first experiment, the combination of lightlines with LBP yields the best performance. The difference in performance between single-scale and multi-scale

TABLE I
COMPARISON OF THE PERFORMANCE OF THE CLASSIFICATION METHODS
PRESENTED HERE

Feature type		Classes	Precision	Recall	F1-score	Misclassified ROIs	AUC
Single-scale	FOS	Mature	0.33	0.13	0.19	13	0.59
		Immature	0.71	0.89	0.79	4	
		Average	0.60	0.67	0.61		
	Luminance	Mature	0.60	0.40	0.48	9	0.70
		Immature	0.78	0.89	0.83	4	
		Average	0.73	0.75	0.70		
	Lightlines	Mature	0.64	0.50	0.56	7	0.86
		Immature	0.82	0.89	0.86	4	
		Average	0.77	0.78	0.78		
Multi-scale	Luminance	Mature	0.56	0.33	0.42	10	0.70
		Immature	0.76	0.89	0.82	4	
		Average	0.70	0.73	0.70		
	Lightlines	Mature	0.69	0.60	0.64	7	0.84
		Immature	0.85	0.89	0.87	2	
		Average	0.80	0.81	0.80		

Fig. 7. Comparison between the classifier output and observer continuous score. The points in red correspond to the misclassified ROIs.
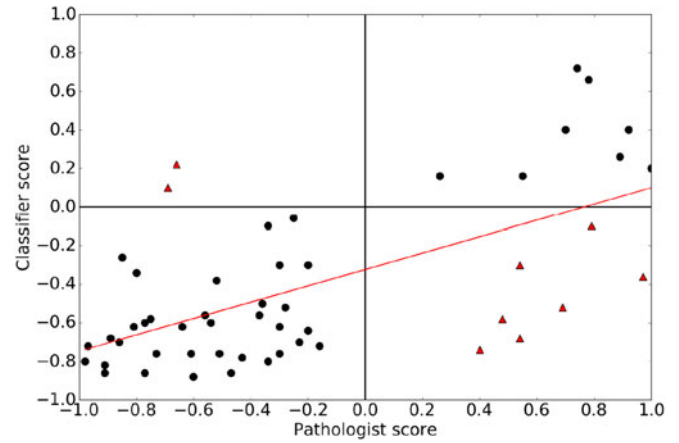

Fig. 6. ROC curves for Luminance and Lightlines/LBP multiscale classifiers that performed best. The Lightlines/LBP classifier area under the ROC curve was 0.84 ($\sigma_3 = 0.75$ and LBP$_{(8,1)}$; $\sigma_5 = 2$ and LBP$_{(8,1)}$), while the Luminance classifier was 0.70 (LBP$_{(8,1)}$ and LBP$_{(16,2)}$).

Fig. 8. Examples of true positives for both classes and their respective classifier scores: (a) 0.91 and (b) 0.58 immature; (c) 0.86 and (d) 0.53 mature.
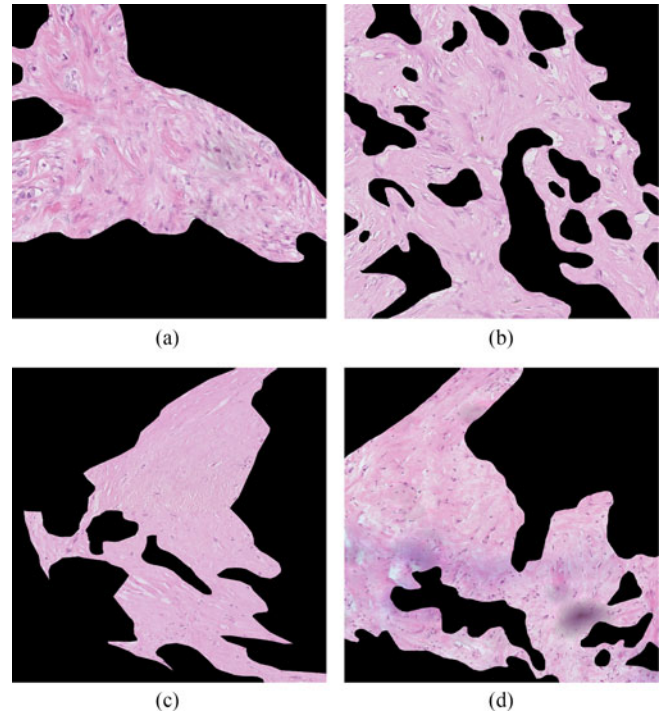
did not show an improvement in the AUC. However, the number of misclassified ROIs dropped in the case of the Lightlines/LBP multi-scale classifier. Fig. 7 shows the pathologist's continuous score on the x-axis and the Lightlines/LBP multi-scale classifier probability output on the y-axis for the test set ($n = 52$). The classifier outputs a probability for each class that ranges from 0 to 1. In order to compare the two measures, we rescaled the classifiers' output range to -1 to 1:

$$\text{Classifier score} = 2 * \text{prob}(0) - 1 \qquad (6)$$

where $\text{prob}(0)$ corresponds to the predicted probabilities for the class "mature" ranged from 0 to 1. A linear regression was fitted and is shown in red on Fig. 7. The fit yielded a significant correlation between the two variables, with a $ps < .001$ and a correlation coefficient $R^2 = 0.62$. The estimated slope was 0.42, indicating a positive correlation.

Fig. 8 shows examples from the test set of true positives from both classes with their respective predicted class probability from the Lightlines/LBP multi-scale classifier. From the 52 ROIs that comprised the test set, 9 were misclassified by the Lightlines/LBP multi-scale classifier, belonging to 6 of the 16 patients. Some of these ROIs are shown in Fig. 9.

IV. DISCUSSION

In this paper we have presented an image analysis method for the automatic classification of IBC stroma maturity in manually chosen regions of interest from digitised H&E stained
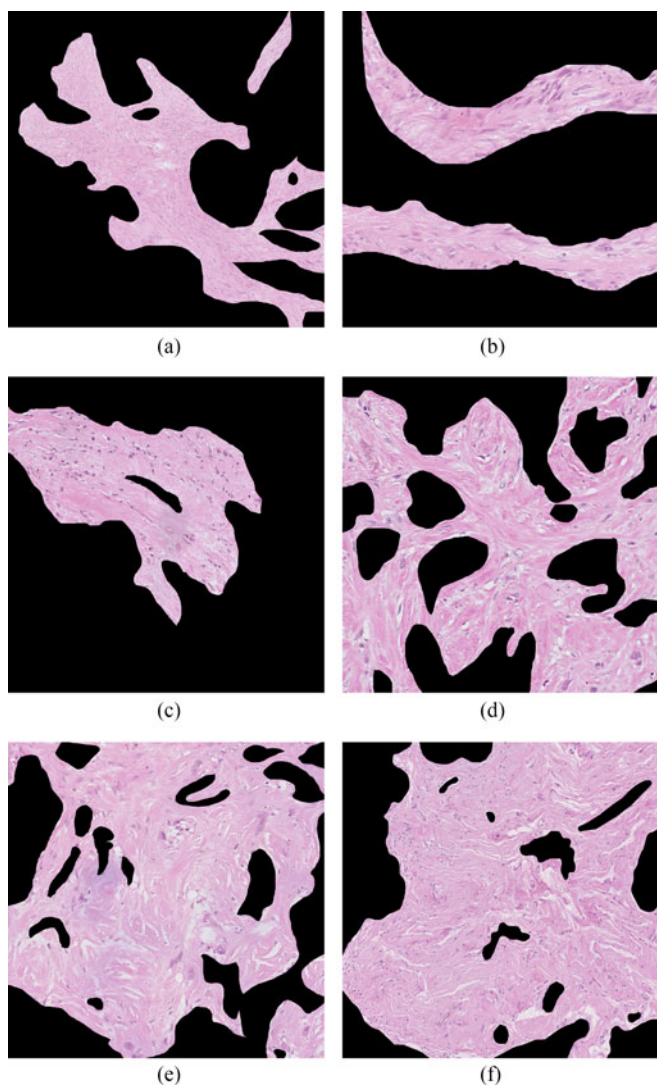
Fig. 9. ROIs that were incorrectly classified by the classifier. (a) and (b) were labelled immature and (c), (d), (e) and (f) were labelled mature by the pathologist.

breast tissue slides. The results for this cohort of patients, demonstrate the ability of our method to distinguish between mature and immature stroma in IBC. To the best of the authors' knowledge, this is the first study that attempts to categorise and automatically classify stromal regions in H&E stained slides that are routinely acquired clinically. Previous related work has been focused on differentiating distinct tumour-associated collagen signatures (TACS) based primarily on collagen fibre alignment with respect to tumour-stromal boundaries in second harmonic generation (SHG) images, therefore requiring the acquisition of additional images, which is time-consuming and requires the use of specialised equipment [17], [34].

The use of LBP, in combination with the filtered images, was found to improve classification accuracy (28% higher than using FOS features and 16% compared to using LBP features on their own). In contrast, the combination of multiple lightlines filter scales did not result in a significant improvement ($p > .05$)

[35]. Applying the LBP algorithm without pre-filtering the image, performed less well. Overall, the results shown here lead us to conclude that the Lightlines/LBP classifier as a means to capture texture features from the H&E digitised images achieves good results in differentiating immature from mature stroma. Out of the 52 regions that comprised the test set, 9 were misclassified, where 7 were labelled mature and 2 immature. The misclassified regions were reviewed by the pathologist in order to determine whether specific pathologic features may have contributed to the misclassification. The ROIs that were misclassified as immature (Fig. 9(d)–(f)) were of particular interest as they exhibit clear features of mature stroma (highly organised collagen strands and low celullarity) but also areas with oedema, which is characteristic of immature stroma. Based on the data presented in Table I, it is evident that there is a slight difference in the classifiers performance when looking at the two classes individually. This can be justified by the fact that our classes were somewhat unbalanced (mature $= 60$; immature $= 109$), and more immature regions were selected and labelled by the pathologist. The graph in Fig. 7 showed that the classifiers' prediction of maturity was correlated with the continuous scores given by the expert pathologist ($R^2 = 0.64$), which is almost as good as the trainee pathologist intra-observer agreement ($R^2 = 0.67$). Moreover, our image dataset covers a wide range of maturity stages as defined by the pathologist, making it a richer and balanced dataset.

Several limitations of the analysis proposed here have to be acknowledged, offering opportunities for improvement in future work. In our experiments we analysed small hand-picked ROI within and around the tumour. Such regions naturally cannot fully represent the overall stromal reaction of the whole tumour, or even the whole histological slide, as tumours can potentially show stromal reaction heterogeneity. Thus, evaluation of the entire area of the peri-tumoural and within the tumour stroma will be performed in future work in order to demonstrate clinical relevance. In this regard, future work will address implementation of whole-slide stroma maturity classification in order to explore this.

Our analysis was based on the assumption that stroma maturity is a binary classification problem. However, this is a simplification, as stromal changes are clearly a complex biological process. As our evaluation was performed mostly in regions where this difference in maturity was clear, difficult cases such as the ones mentioned previously, might be tackled by expanding our binary classification into a multiple-class problem.

Further improvements of the proposed framework could include the investigation of the optimal image size for this particular problem, which would also give us an insight into the minimum amount of information we need in order to distinguish between these two types of textural images (how small can the images be and still be correctly classified). In other words, this would also indicate the scale at which changes in maturity could be detected.

Finally, these stromal features could aid in the interpretation of the radiological signal and relating it to microscopical changes, ultimately enhancing pre-operative imaging for predicting prognosis and selection of patient treatment.
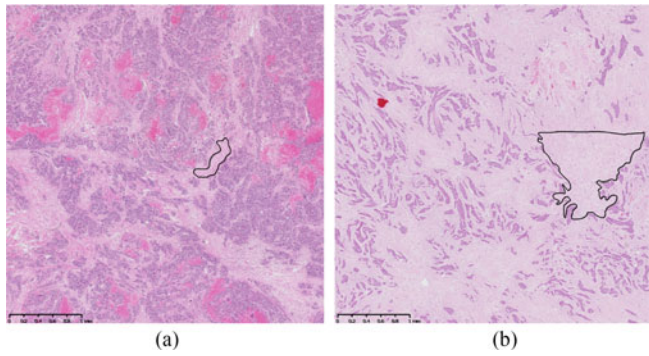
Fig. 10. Examples of (a) stroma-poor and (b) stroma-rich grade 3 NST tumours from our cohort of patients.

V. CONCLUSION

This study has demonstrated the ability of a combination of DtG filtering with LBP to determine the maturity of breast stroma as represented in H&E stained slides. The AUC for a test set of 52 regions of interest was 0.86, decreasing marginally to 0.84 for multiple scales. Inclusion of the lightline DtG pre-filtering significantly improved the discrimination of the method over direct application of the LBP method the grey-scale luminance images. This work demonstrates that texture based classification of H&E images from IBC could have practical clinical applications in distinguishing stromal morphology. Although stroma assessment is not currently a standard procedure for pathologists when performing a diagnosis or in prognostication, the idea of using this component as a prospective feature for evaluating tumour growth, aggressiveness and disease progression is gaining ground. Furthermore, the scale of the regions that we have analysed is ~0.1 to 1.0 mm and therefore at a scale where emerging modalities such as MR diffusion, digital tomosynthesis and high resolution ultrasound can obtain signal. Stromal micro-structure modulates proton diffusion and tissue micromechanics. The classification described in this paper will map to signals from emerging MR diffusion and ultrasound elastography methods, ultimately enabling some form of *in-vivo* histology and cancer risk assessment.

APPENDIX

Figs. 10 show examples of both stroma-rich and stroma-poor tumours. Stroma-poor tumours resulted in smaller ROIs, while in stroma-rich tumours the ROIs were slightly bigger.

Figs. 12 and 11 represent the wide range of scores given by the pathologist. The stromal regions used in this study were assessed by an experienced breast pathologist and scored according to their maturity. Because different levels of maturity are present around the tumour, the rating was done on a continuous scale.

Figs. 13 and 14 show the paired data from each observer (SEP and KN). The Pearson correlation coefficient was computed for each observer. Simple arithmetic mean was computed for the duplicate observations for each observer and plotted in Fig. 15. For evaluating consistency between observers, the Pearson correlation coefficient was computed.
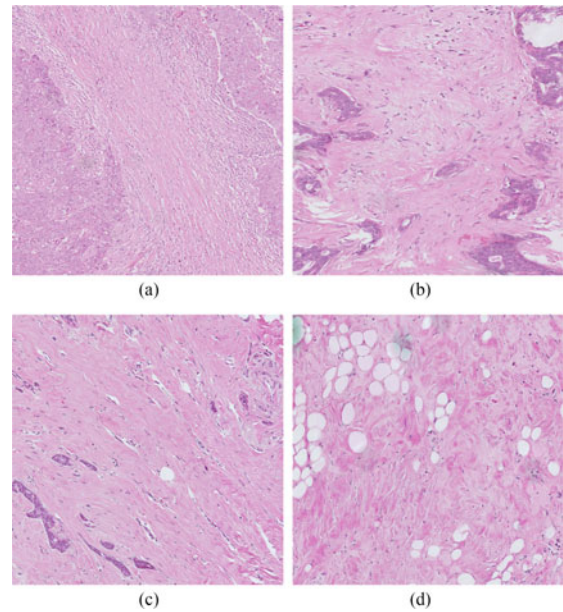

Fig. 11. Representative ROIs and their respective score given by the expert pathologist: (a) 0.25; (b) 0.50; (c) 0.75; (d) 1. x20 magnification. ROIs were resized for illustration purposes.
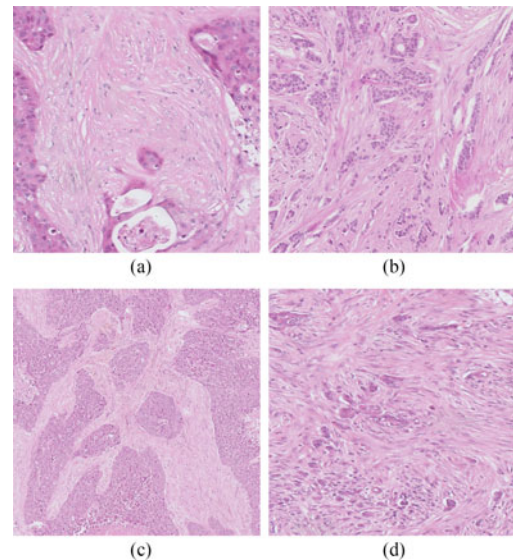

Fig. 12. Representative ROIs and their respective score given by the expert pathologist: (a) −0.25; (b) −0.50; (c) −0.75; (d) −1. x20 magnification. ROIs were resized for illustration purposes.
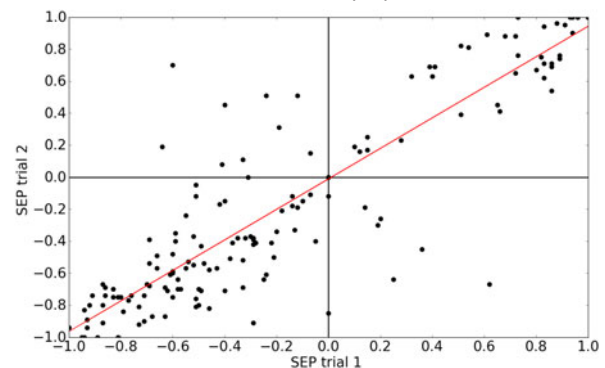

Fig. 13. Paired maturity scores from experienced pathologist (SEP). The ICC computed was 0.91.
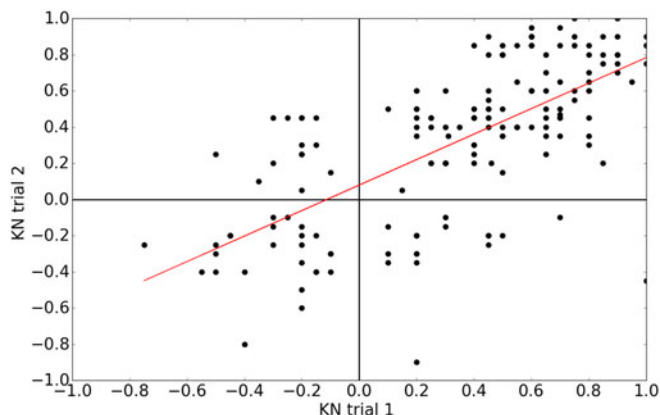
Fig. 14. Paired maturity scores from junior pathologist (KN). The ICC computed was 0.68.
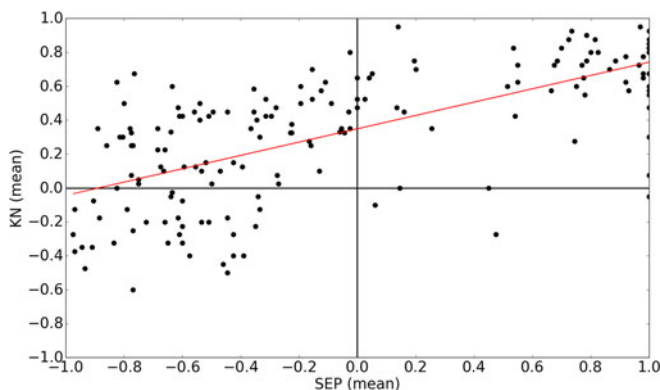

Fig. 15. Simple arithmetic mean was computed for the duplicate observations for each observer. The inter-observer correlation coefficient was $R^2 = 0.67$.

REFERENCES

[1] F. P. O'Malley *et al.*, *Breast Pathology: A Volume in the Series: Foundations in Diagnostic Pathology*, 2nd ed. Amsterdam, The Netherlands: Elsevier, Aug. 2011.

[2] C. W. Elston and I. O. Ellis, "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up," *Histopathology*, vol. 19, no. 5, pp. 403–410, Nov. 1991.

[3] A. Planche *et al.*, "Identification of prognostic molecular features in the reactive stroma of human breast and prostate cancer," *PLoS ONE*, vol. 6, no. 5, pp. 1–12, May 2011.

[4] A. E. M. Reed *et al.*, "Invasive lobular carcinoma of the breast: Morphology, biomarkers and omics," *Breast Cancer Res.*, vol. 17, no. 1, pp. 1–11, Jan. 2015.

[5] H. Ueno *et al.*, "Histological categorisation of fibrotic cancer stroma in advanced rectal cancer," *Gut*, vol. 53, no. 4, pp. 581–586, Apr. 2004.

[6] P. Farmer *et al.*, "A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer," *Nature Med.*, vol. 15, no. 1, pp. 68–74, Jan. 2009.

[7] S. Gangadhara *et al.*, "Pro-metastatic tumor-stroma interactions in breast cancer," *Future Oncology*, vol. 8, no. 11, pp. 1427–1442, Nov. 2012.

[8] B. K. Brisson *et al.*, "Type III collagen directs stromal organization and limits metastasis in a murine model of breast cancer," *Amer. J. Pathology*, vol. 185, no. 5, pp. 1471–1486.

[9] M. W. Conklin *et al.*, "Aligned collagen is a prognostic signature for survival in human breast carcinoma," *Amer. J. Patholog*, vol. 178, no. 3, pp. 1221–1232.

[10] M. J. Bissell and W. C. Hines, "Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression," *Nature Med.*, vol. 17, no. 3, pp. 320–329, Mar. 2011.

[11] M. Veta *et al.*, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, May 2014.

[12] G. Finak *et al.*, "Stromal gene expression predicts clinical outcome in breast cancer," *Nature Med.*, vol. 14, no. 5, pp. 518–527, May 2008.

[13] A. H. Beck *et al.*, "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Sci. Translational Med.*, vol. 3, no. 108, Nov. 2011, Paper 108ra113.

[14] J.-M. Chen *et al.*, "New breast cancer prognostic factors identified by computer-aided image analysis of HE stained histopathology images," *Sci. Rep.*, vol. 5, pp. 1–13, May 2015.

[15] N. Linder *et al.*, "Identification of tumor epithelium and stroma in tissue microarrays using texture analysis," *Diagnostic Pathology*, vol. 7, no. 1, pp. 1–12, Mar. 2012.

[16] O. Sertel *et al.*, "Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development," *Pattern Recognit.*, vol. 42, no. 6, pp. 1093–1103, Jun. 2009.

[17] P. P. Provenzano *et al.*, "Collagen reorganization at the tumor-stromal interface facilitates local invasion," *BMC Med.*, vol. 4, no. 1, p. 38, Dec. 2006.

[18] V. Ajeti *et al.*, "Structural changes in mixed Col I/Col V collagen gels probed by SHG microscopy: Implications for probing stromal alterations in human breast cancer," *Biomed. Opt. Express*, vol. 2, no. 8, pp. 2307–2316, Aug. 2011.

[19] C. Bailey *et al.*, "Modelling vasculature and cellular restriction in breast tumours using diffusion MRI," in *Proc. Int. Soc. Mag. Reson. Med.*, vol. 23, 2015, Art. no. 1082.

[20] P. A. Wijeratne *et al.*, "Multiscale modelling of solid tumour growth: The effect of collagen micromechanics," *Biomech. Modeling Mechanobiol.*, vol. 15, pp. 1079–1090, 2016.

[21] S. Reis *et al.*, "Classification of breast cancer stroma as a tool for prognosis," in *Proc. SPIE Med. Imaging*, 2016, vol. 979105.

[22] M. M. Fraz *et al.*, *Retinal Vessel Extraction Using First-Order Derivative of Gaussian and Morphological Processing*. Berlin, Germany: Springer, 2011, pp. 410–420.

[23] X. Wu *et al.*, "Palmprint texture analysis using derivative of Gaussian filters," in *Proc. Int. Conf. Comput. Intell. Security*, Nov. 2006, vol. 1, pp. 751–754.

[24] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[25] M. Crosier and L. D. Griffin, "Using basic image features for texture classification," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 447–460, Jan. 2010.

[26] N. Jaccard *et al.*, "Segmentation of phase contrast microscopy images based on multi-scale local basic image features histograms," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Apr. 2015, pp. 1–9.

[27] J. H. Hipwell *et al.*, "Characterizing breast phenotype with a novel measure of fibroglandular structure," in *Breast Imaging, Lecture Notes in Computer Science*. Berlin, Germany: Springer, Jul. 2012, pp. 181–188.

[28] W. Revelle, *Psych: Procedures for Psychological, Psychometric, and Personality Research*, Northwestern University, Evanston, IL, USA, 2016.

[29] A. Khan *et al.*, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, Jun. 2014.

[30] L. Griffin, "The second order local-image-structure solid," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1355–1366, Aug. 2007.

[31] T. Ojala *et al.*, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[32] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[33] L. B. Mostaço-Guidolin, "Collagen morphology and texture analysis: From statistics to classification," *Sci. Rep.*, vol. 3, Jul. 2013, Art. no. 2190.

[34] J. S. Bredfeldt *et al.*, "Automated quantification of aligned collagen for human breast carcinoma prognosis," *J. Pathology Informat.*, vol. 5, p. 28, Aug. 2014.

[35] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.