Invasive ductal breast carcinoma detector that is robust to image magnification in whole digital slides

Matthew Balazsi
Paula Blanco
Pablo Zoroquiain
Martin D. Levine
Miguel N. Burnier, Jr.

Invasive ductal breast carcinoma detector that is robust to image magnification in whole digital slides

Matthew Balazsi,[a,b,*] **Paula Blanco,**[b] **Pablo Zoroquiain,**[b] **Martin D. Levine,**[a] **and Miguel N. Burnier Jr.**[b]
[a]McGill University, Centre for Intelligent Machines, Electrical and Computer Engineering, 3480 University Street, City, Montreal, Quebec H3A 2A7, Canada
[b]McGill University, Department of Pathology, Henry C. Witelson Laboratory, 1001 Boulevard Decarie, Block E, Montreal, Quebec H4A 3J1, Canada

Abstract. Invasive ductal breast carcinomas (IDBCs) are the most frequent and aggressive subtypes of breast cancer, affecting a large number of Canadian women every year. Part of the diagnostic process includes grading the cancerous tissue at the microscopic level according to the Nottingham modification of the Scarff-Bloom-Richardson system. Although reliable, there exists a growing interest in automating the grading process, which will provide consistent care for all patients. This paper presents a solution for automatically detecting regions expressing IDBC in images of microscopic tissue, or whole digital slides. This represents the first stage in a larger solution designed to automatically grade IDBC. The detector first tessellated whole digital slides, and image features were extracted, such as color information, local binary patterns, and histograms of oriented gradients. These were presented to a random forest classifier, which was trained and tested using a database of 66 cases diagnosed with IDBC. When properly tuned, the detector balanced accuracy, F1 score, and Dice's similarity coefficient were 88.7%, 79.5%, and 0.69, respectively. Overall, the results seemed strong enough to integrate our detector into a larger solution equipped with components that analyze the cancerous tissue at higher magnification, automatically producing the histopathological grade. © *2016 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JMI.3.2.027501]

Keywords: computer vision; whole digital slides; histopathology; machine learning; invasive ductal breast carcinoma; cancer.

Paper 15220RR received Nov. 16, 2015; accepted for publication Apr. 18, 2016; published online May 18, 2016.

1 Introduction

1.1 Motivation

An estimated 25,000 Canadian women were diagnosed with breast carcinoma in 2015,[1] representing the highest incidence of cancer in the female population. Breast carcinomas are most commonly in the invasive ductal form (IDBC),[2] which yields an 85% survival rate after five years.[3] An important facet of the diagnostic process is evaluating the tumor aggressivity based on the microscopic properties of the tissue, as performed by a pathologist. To this end, cases expressing IDBC are assessed using the Nottingham modification of the Scarff-Bloom-Richardson (SBR) grading system, which has become the standard for pathologists since 2002 after being recommended by the International Union against Cancer[4] and the World Health Organization.[5] It has had success due to its objective nature and strong correlation with patient prognoses.[6] Despite its strict guidelines, inter-rater variability has been thoroughly investigated using the SBR grading system.[7–10] However, a recent study that included 732 pathologists reported that 8.3% of raters had an agreement rate of <95%;[10] therefore, the system can be improved upon.

Automating the process through software would reduce inter-rater variability and ensure more consistent results to the benefit of the patients. Recent technological advances have made this possible, notably with the advent of digital pathology. Slide scanners have been the basis of this because they produce very large images of microscopic tissue, commonly referred to as whole digital slides (Fig. 1). This paper proposes a solution for the first step in the automation process: detecting IDBC in whole digital slides.

1.2 Background

Fully automated histopathological grading systems are not yet clinically available; however, researchers have been working toward this end. Some have departed from the prognostic factors traditionally analyzed by pathologists and are leveraging machine learning techniques to find new factors for tumor classification. For example, Beck et al.[11] have found that certain features relevant for their automated classifier were previously assumed to be irrelevant by pathologists. Therefore, such techniques may provide a deeper understanding of malignancy.

Conversely, methods that more faithfully replicate the workflow of a pathologist will undoubtedly encounter fewer barriers to clinical use. The strategy would be to first automatically segment IDBC from whole digital slides and subsequently analyze the cancerous tissue according to the SBR grading system (Fig. 2). One research group, in particular, has produced a peer-reviewed paper for every step required for accomplishing this goal, from finding areas of interest[12] to the tissue analysis.[13,14]

Finding areas of interest can be challenging due to the overwhelming size of whole digital slides. To overcome this, researchers often tessellate the whole digital slide into a rectangular grid of independent areas. Although simple, the

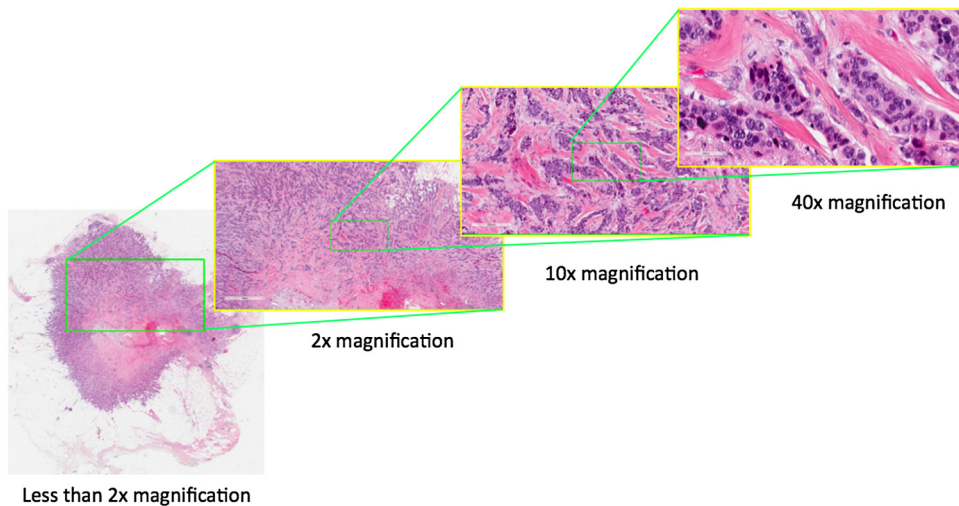*Address all correspondence to: Matthew Balazsi, E-mail: matthew.balazsi@gmail.com

Fig. 1 An example of a whole digital slide. At 40× magnification, this file contains 53,642 × 55,614 pixels and three color channels, for a total size of 8.95 gigabytes uncompressed.
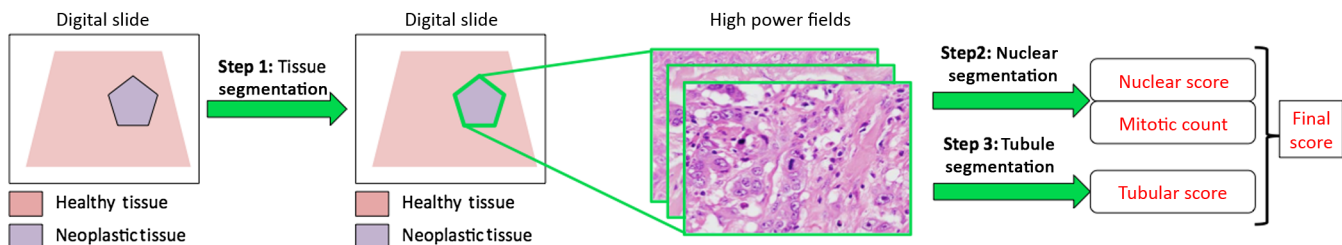

Fig. 2 Automating the SBR grading system would first require a segmentation of the cancerous areas followed by an analysis of the cellular properties of the tissue.

tessellation arbitrarily splits the slide image, which may have the effect of distorting important pathological features. To further lower computational time, methods either sample a subset of the rectangular areas[12] or reduce the magnification level during analysis.[15,16] The former may be undesirable for clinical applications since the omitted areas may contain information that would ultimately affect tumor classification. Similarly, the latter technique may lead to a misinterpretation of the data due to the reduced magnification. Despite the potential flaws, researchers have developed IDBC detectors with very favorable results.

1.3 Our Solution

Our paper presents a method for achieving the first step of the process as seen in Fig. 2, that is, automatically finding areas expressing IDBC within whole digital slides. Our approach most resembles a design recently proposed by Cruz-Roa et al.[16] while additionally addressing two of the common weaknesses described in Sec. 1.2. First, a grid with malleable borders replaces the rectangular tessellation. This generates areas that attempt to encapsulate consistent pathological features. Second, our method is shown to be robust against a wide range of magnification levels.

Our broad strategy was to find areas where cells proliferated chaotically within the whole digital slide. Although disorganized cell growth is a strong indicator of IDBC, we acknowledge that other types of cancer also fit this criterion. For example, ductal carcinomas *in situ* (DCIS), which are often co-expressed with IDBC, share similar cellular traits with IDBC. Although

detecting DCIS would not be strictly wrong, its grading process does not follow the SBR system and should therefore be omitted in the final segmentation.

Despite this limitation, our solution successfully detects IDBC in whole digital slides. Briefly, the images are first tessellated using the simple linear iterative clustering (SLIC) segmentation technique,[17] grouping homogeneous regions into a single tile. As opposed to other segmentation methods in digital pathology,[12,15,16,18,19] the tessellation was not constrained to a rectangular grid, but followed the borders of objects in the image. Next, the resulting tiles were represented by image features describing their color and textural information. Finally, the image features were used to train and test a random forest (RF) classifier, which determined if a given image tile expressed IDBC (Fig. 3). The astounding size of digital slides imposed
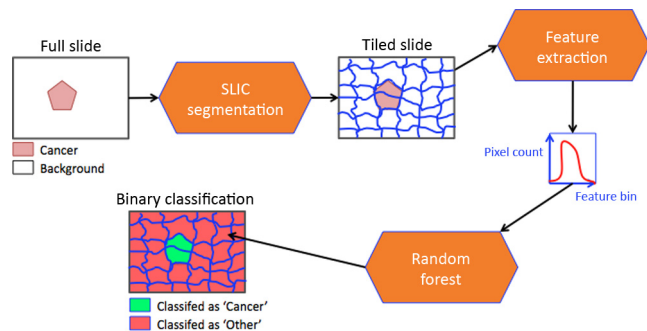

Fig. 3 System overview.

a heavy computational load on our design, which was overcome by exploiting cloud computing, more specifically using Amazon Web Services' m1.xlarge ec2 instances. (The term "instances" is used by Amazon to describe the type of cloud computer.) This heightened capability permitted cases to be resolved in less than an hour.

Generally speaking, the detector performed reasonably well compared to competing methods when trained on 50 cases and tested on 16. Despite some errors, our design faithfully indicated regions expressing IDBC as labeled by a pathologist. Therefore, our approach could realistically be integrated into a larger solution designed to actually automatically grade IDBC according to the SBR system.

2 Methods

2.1 Datasets

To the best of our knowledge, no freely available database of IDBC slides exists; therefore, it was necessary to create our own. First, glass slides from 66 cases of breast mastectomies were retrieved from the McGill University Hospital Centre pathology registry (study 3229). There was a variable number of slides per case, most often exceeding 20. Retaining all of the slides would have been superfluous for the purpose of our project; therefore, a pathologist selected one slide per case, chosen to maximize the tumor expression in each slide. In the third step, digital reproductions of the slides were created using the Aperio Scanscope AT Turbo slide scanner at 40× magnification with a resolution of 0.2485 microns per pixel. Finally, a pathologist annotated tissues of interest on every slide using the Imagescope software (Aperio Imagescope v11.1.2.760).

Our dataset can be considered small since the number of instances (66 cases of IDBC) is overshadowed by the amount of data per case (an average of 1.5 gigabytes per whole digital slide when compressed). Additionally, the dataset is unbalanced since different pathological entities covered irregularly large areas of the image. Both of these attributes were considered during the design process (Sec. 2.4).

2.2 Tiling Whole Digital Slides Using Simple Linear Iterative Clustering Segmentation

The first step of our solution was to tessellate the whole digital slides into more manageable subsets, allowing us to analyze each one independently. To accomplish this, the image was tiled nonuniformly. This computational process required that the tiles have malleable borders, permitting them to follow the tissue boundaries or other histopathological structures in the images. This step essentially required a region segmentation method applied to the entire digital slide. One such method is the SLIC superpixel (SP) method.[17] The SLIC algorithm groups similar pixels in the image, forming larger region segments with shapes that reflect the objects within the image.[20–22]

For our project, we employed the Images and Visual Representation Group version of the SLIC software. From our experience with this software, we have found that a superpixel size SP_{size} of 10,000 pixels and compactness factor SP_m of 30 was ideal for our dataset. This was determined by optimizing the boundary recall rates and the reconstruction error on our data. (For the sake of brevity, we do not demonstrate this validation step in the current paper. Interested readers are referred to the original thesis.[23])

2.3 Representing Image Tiles Using Chrominance and Textural Features

Superpixels provided a coherent tiling of the whole digital slide. Reaching the ultimate objective also required determining which superpixels were positive for cancer. To accomplish this, we transformed the image representation of each superpixel into a feature space based on chrominance and texture. Since both the shape and size of the superpixels vary, we focused on features that could be represented by histograms. When properly normalized, two areas of different size (i.e., superpixels) that express similar patterns should yield similar histograms.

More specifically, we sought a combination of image features that could describe a variety of histopathological tissue patterns. The data were represented by the two color spaces, the RGB and CIELab,[24] and one gray-scale color channel. Within every color space, chrominance features are obtained using color histograms.[25,26] Next, textural features are computed using rotation-invariant local binary patterns (LBPs)[27–29] over a wide range of radii. Finally, edge information is characterized by histograms of oriented gradients (HOG).[30–33] This combination of descriptors provides 16,128 features per superpixel, as described in Table 1.

2.4 Decision Making with Random Forests

So far we have presented methods for splitting digital slides into meaningful tiles using SLIC segmentation and defining them mathematically using features. The final step requires a classifier to decide whether or not a given image tile indicates cancer. We used an RF classifier[34,35] for this purpose (Table 2).

There are several advantages to RFs, and many of them are of great interest to the application discussed in this paper. First, the overall performance of an RF is competitive with other state-of-the-art methods. They can also process a large number of features elegantly without explicitly reducing the dimension of the space, which is vital when dealing with the uncommonly large number of features required in our design. Another significant advantage is that RFs report the most important features for classifying the data, thereby permitting a clear understanding of how the classifier has reached its decision. Finally, the RF classifier has proven to be effective when dealing with unbalanced datasets,[35] which is also very important since our dataset does not contain equal amounts of healthy and cancerous tissue, as discussed in Sec. 2.1.

2.5 Assessing the Classifier Performance Using Cross-Validation

In Sec. 2.4, we have introduced a technique for automatically classifying data based on a training set. It is also necessary to measure the classifier performance for the purposes of validation and optimization. Our approach produces a binary classification, thereby distinguishing IDBC from all other histopathological tissue types. Comparing the classification results to the ground truth as provided by expert pathologists introduces four possible outcomes (Table 3). In certain cases, the superpixel boundaries did not always coincide with the ground truth annotations and could therefore contain a mix of negative and positive labels. To resolve this conflict, the final label for a given superpixel was determined by majority vote.

Table 1 A given tile, or superpixel, was represented by (b) seven color channels. Every channel contained (a) 2304 features, giving (c) 16,128 features for all seven channels.

(a)

Feature	Radius	# of bins
LBP	1	256
LBP	2	256
LBP	3	256
LBP	4	256
LBP	5	256
LBP	10	256
LBP	20	256
HOG magnitude	—	256
HOG orientation	—	256
Color	—	256
Total	—	**2304**

(b)

Color space	# of channels
RGB	3
Lab	3
Gray scale	1
Total	**7**

(c)

# of channels	7
# of bins per channel	2304
# of features per superpixel	**16,128**

Table 2 Description of the input parameters for RF classifiers.

Parameter	Description
$RF_{NumTrees}$	The number of decision trees in the RF.
RF_{Depth}	The maximum depth of every decision tree.
$RF_{NumTests}$	The number of features tested at every node during the training process. The feature that best splits the data will be retained for classification purposes.
$RF_{Confidence}$	Output probabilities above this threshold will be considered positive for IDBC.

Table 3 The four possible outcomes of binary classification.

	Ground truth label: IDBC	Ground truth label: Other
Detector result: IDBC	True positive (TP)	False positive (FP)
Detector result: Other	False negative (FN)	True negative (TN)

To characterize the system behavior, we used several measures of accuracy. First, the sensitivity [Eq. (1)], specificity [Eq. (2)], balanced accuracy (BAC)[36] [Eq. (3)], and F1 score[37] [Eq. (4)] were useful for evaluating how the system performed as a binary detector. Dice's similarity coefficient (DSC)[38,39] [Eq. (5)] was employed to measure how the system segmented regions of interest. As described in Eq. (5), a perfect overlap of two regions A and B will yield a coefficient of 1.

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{1}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \tag{2}$$

$$\text{BAC} = \frac{\text{Sensitivity} + \text{Specificity}}{2}, \tag{3}$$

$$\text{F1 score} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP}, \tag{4}$$

$$\text{DSC} = \frac{2 \cdot |A \cap B|}{|A| + |B|}. \tag{5}$$

We note that measuring a detector's performance may be biased by the selection of test cases as well as the actual training and testing cases available. To compensate for this, we employed k-fold cross-validation[40] and ensured that no case was tested more than once.

2.6 Selection of Parameters

The proposed detector was substantiated through rigorous testing on our dataset, which consisted of 66 cases diagnosed with IDBC. However, before proceeding to this step, it was necessary to select the seven input parameters associated with our method. To recapitulate, the whole digital slide was first sampled at a given magnification (M) and subsequently tessellated using the SLIC algorithm, which required setting the superpixel size (SP_{size}) and compactness factor (SP_m). Although feature extraction did not require parameter selection, the RF classifier depended on the parameters $RF_{NumTrees}$, RF_{Depth}, $RF_{NumTests}$, and $RF_{Confidence}$. In this paper, we will demonstrate the process only for validating the image magnification (M), although this same procedure was repeated for the other six parameters.[23]

Selecting the image magnification was important since it directly affected the computation time. For this reason, it was necessary to examine the effect of image magnification on the classification result. To this end, classifiers were trained

and tested by sampling the whole digital slides at 2×, 4×, and 10× magnifications. Comparing the classifier performance at different magnifications may be flawed due to the inherent randomness associated with the SLIC tessellation process. To overcome this issue, we designed an experiment where the superpixels at all three magnifications covered identical tissue areas and the data that were passed to the classifier differed only in their image resolution. To achieve this, the tessellation at 4× was projected onto the slide at 2× and 10× magnifications. Four RFs per magnification, described in Table 4, were trained on six randomly selected cases and subsequently tested on another six.

2.7 Validation

Once the input parameters were selected (Table 5), the algorithm was then tested using our entire dataset. We performed k-fold cross-validation[40] using 7, 28, and 50 training cases. Cross-validation with 7 and 28 training cases required two folds each and could be sensitive to the separation of the cases. Therefore, we repeated the validation process five times to measure whether

Table 4 The parameters used for testing the effect of image magnification on the classifier performance.

Test name	SP_{Size}	SP_m	$RF_{NumTrees}$	RF_{Depth}	$RF_{NumTests}$
Test 1	10,000	30	500	50	200
Test 2	10,000	30	500	50	500
Test 3	10,000	30	500	75	200
Test 4	10,000	30	500	75	500

Table 5 The parameters used for the cross-validation of 66 cases of IDBC.

M	SP_{Size}	SP_m	$RF_{NumTrees}$	RF_{Depth}	$RF_{NumTests}$
2×	10,000	30	200	50	400

Table 6 The effect of magnification on the detector's performance.

M	2×	4×	10×
Sensitivity	91.80% ± 0.8	92.10% ± 0.2	88.40% ± 1.4
Specificity	79.60% ± 0.4	80.50% ± 0.8	83.80% ± 0.3
BAC	85.70% ± 0.4	86.20% ± 0.4	86.10% ± 0.7
F1 score	63.30% ± 0.8	64.20% ± 0.8	66.10% ± 0.8
DSC	0.54 ± 0.005	0.55 ± 0.005	0.57 ± 0.004

the results varied with respect to the random assignment of cases into both folds.

Additionally, we produced a receiver operating characteristics (ROC) curve[41] for the parameter $RF_{Confidence}$. More specifically, the RFs computed the probability that a region was positive for cancer, and any result above the confidence threshold $RF_{Confidence}$ was considered positive (Fig. 4). Changing $RF_{Confidence}$ directly affects the decisions made by the detector. Therefore, the ROC curve permitted us to determine the optimal value of $RF_{Confidence}$, while the area under the ROC curve (AUC) quantified the overall classifier performance.

Finally, we compared the effects of tessellating whole digital slides into a square grid versus superpixels. Therefore, we repeated the k-fold cross-validation with 50 training cases where all whole digital slides were broken into 100×100 pixel subimages.

3 Results

3.1 Effect of Image Magnification on the Classifier

As reported in Table 6, increasing the magnification from 2× to 10× produced no significant improvement in terms of BAC, a 3% improvement for the F1 score, and a 0.06 improvement in terms of DSC. Considering the amount of supplementary data analyzed, increasing the magnification produced minor improvements. We assume that this was a testament to the large variety of features used to describe the tissue. For example, the radii of the LBP codes ranged from 1 to 20 pixels; therefore, it was capable of capturing the same phenomena at different scales.
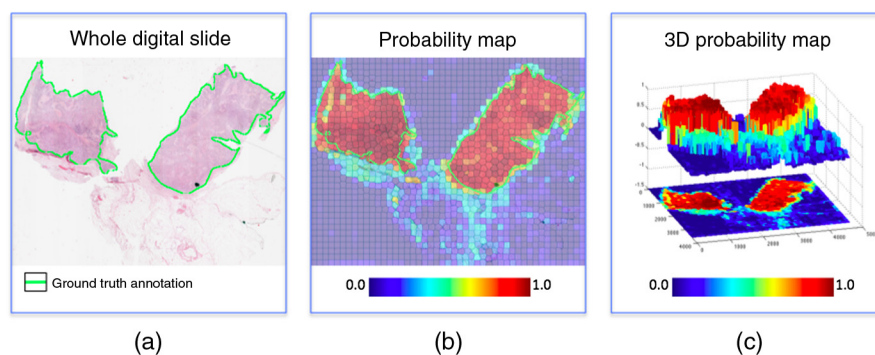

Fig. 4 RFs output a probability map of IDBC. The cancer confidence threshold dictates at what probability level a given region is considered to be positive. (a) An example of a whole digital slide at low magnification. (b) The output of the classifier is a probability map of IDBC expression. (c) The probability map can be viewed as a series of peaks and valleys. The confidence threshold dictates the minimal peak elevation considered for classifying an area as IDBC.

Table 7 Cross-validation results using 66 cases of IDBC. "Square" indicates experiments in which the superpixels were replaced by a square grid.

	—	—	—	Square	Ref. 16
# of training cases	7	28	50	50	84
# of folds	2	2	5	5	—
$RF_{Confidence}$	0.6	0.6	0.6	0.7	—
Sensitivity	80.72 ± 7.4%	79.52 ± 3.9%	84.03 ± 1.6%	77.94 ± 4.3%	—
Specificity	90.92 ± 3.2%	93.30 ± 1.2%	93.31 ± 2.1%	95.20 ± 2.3%	—
AUC	0.93	0.94	0.95	0.95	—
BAC	85.82 ± 2.6%	86.41 ± 1.5%	88.67 ± 1.3%	86.79 ± 1.5%	84.23%
F1 score	74.40 ± 3.8%	76.60 ± 1.4%	79.51 ± 6.3%	77.57 ± 2.8%	71.80%
DSC	0.64 ± 0.07	0.66 ± 0.01	0.69 ± 0.12	0.66 ± 0.02	—

3.2 Testing the Design on a Large Database of Cases

The cross-validation results are summarized in Table 7. The classifier performed reasonably well and demonstrated similar metrics as in Ref. 16, although both methods could not be directly compared since they were tested on different data and no publicly available dataset currently exists. Moreover, tessellating the digital slides using superpixels as opposed to a square grid improved all metrics. Interestingly, the gains achieved through superpixel tessellation were slightly greater than when increasing the magnification to 10× (Table 6).

Finally, as would be expected, increasing the number of training cases seemed to improve all performance metrics (Table 7), including the AUC values (Fig. 5). However, this does not address the issue of whether failure modes exist that could not be overcome by using additional training cases. This required a careful visual inspection of the actual results. To recapitulate, analysis of the image colors and textures was essentially designed to find epithelial cells that proliferated in a disorganized pattern. For the most part, this strategy effectively recreated the ground truth annotations, as dictated by expert pathologists (Fig. 6).

However, the cellular architecture in IDBC also bears a resemblance to other histopathological entities, such as DCIS. Differentiating these two would require more information than what is currently analyzed by the detector. This produced areas in the image that were actually seen to express DCIS but were misclassified as IDBC (Fig. 7). This illustrates that an analysis system of the type discussed in this paper would be
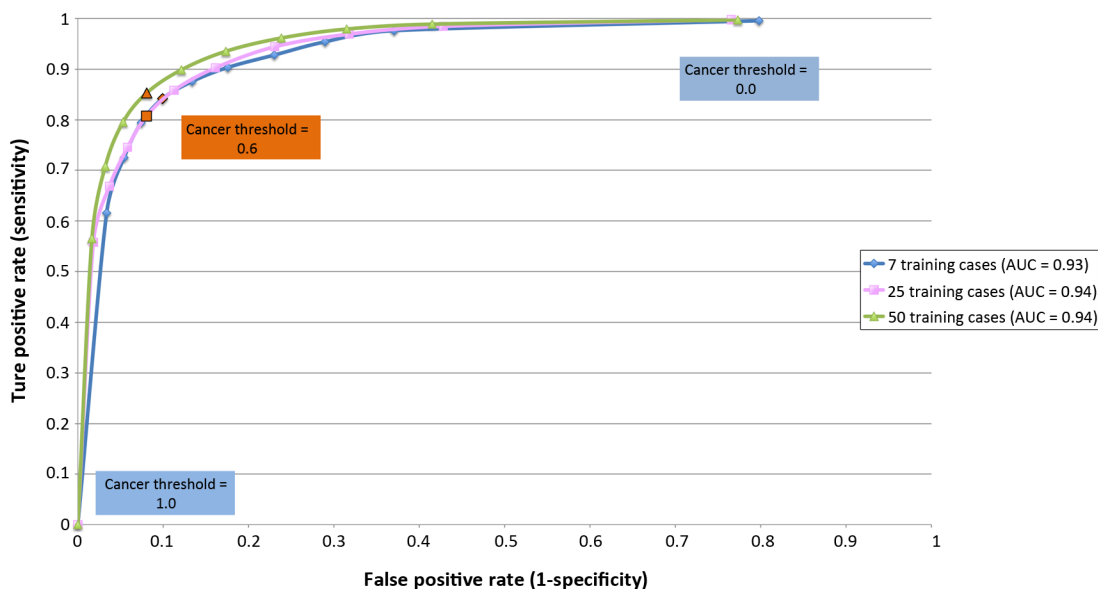

Fig. 5 The ROC curve with respect to $RF_{Confidence}$. The classifier performed slightly better as the number of training cases increased, as demonstrated by the AUC values.
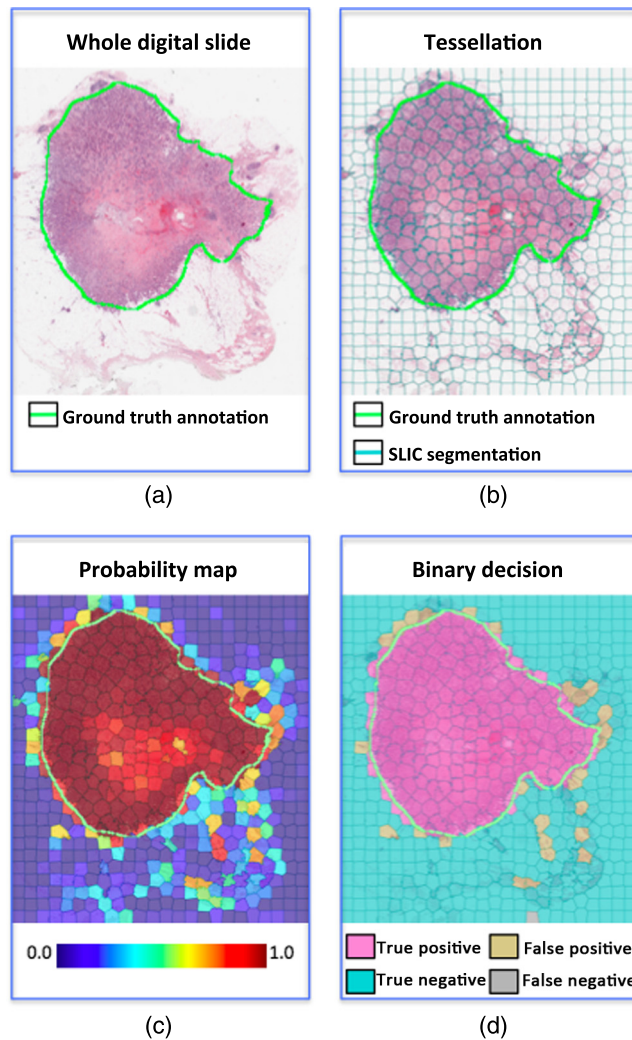
Fig. 6 An example of the IDBC detector output, using 50 training cases: (a) the original whole digital slide, (b) tessellation using SLIC segmentation, (c) the probability map of cancer, and (d) the final decision with an $RF_{Confidence}$ of 0.6%.

suitable for selecting specific areas in the digital slide for a more detailed examination. This second step would then permit analysis for DCIS on a cellular basis. Indeed, this was the actual goal of our research.

4 Conclusion

4.1 Summary

The solution proposed in this paper performed consistently over different magnifications and tessellation resolutions. Therefore, any future application using our framework could potentially adjust these parameters to fit their own specific needs without depreciating the performance. Furthermore, the overall solution can be easily applied to other cancer types simply by changing the training data. This is a testament to the image features used for the analysis, which is able to successfully characterize and detect the differences among different tissue types.

On the other hand, several design choices introduced limitations. First, and perhaps foremost, the tessellation process, which imparted many advantages, unfortunately required heavy computational power, which was accessible only using cloud computing. Therefore, our approach could not easily be

reproduced with conventional computers. Researchers who do not have the necessary hardware can either modify the SLIC SP algorithm or employ a rectangular tessellation. Unfortunately, the latter would incur a performance hit to the system. Second, analyzing the RGB color space is recommended only in conjunction with highly controlled histopathological and imaging procedures, due to a lack of color normalization in the space. Finally, the solution was verified only on slides where IDBC was very prominent. The detector was not tested to determine whether it was capable of finding small foci of IDBC.

4.1.1 Summary of the results

The method proposed in this paper depends on an initial parameter selection, which we obtained by experimentation. Given this initial procedure, we were able to successfully discover the areas expressing IDBC with a BAC of 88.7%, an F1 score of 79.5%, and a DSC of 0.69. Performance was improved by employing superpixel tessellation compared to a rectangular grid. The superpixels had a tendency to group similar histopathological patterns and followed the edges of the tumors.
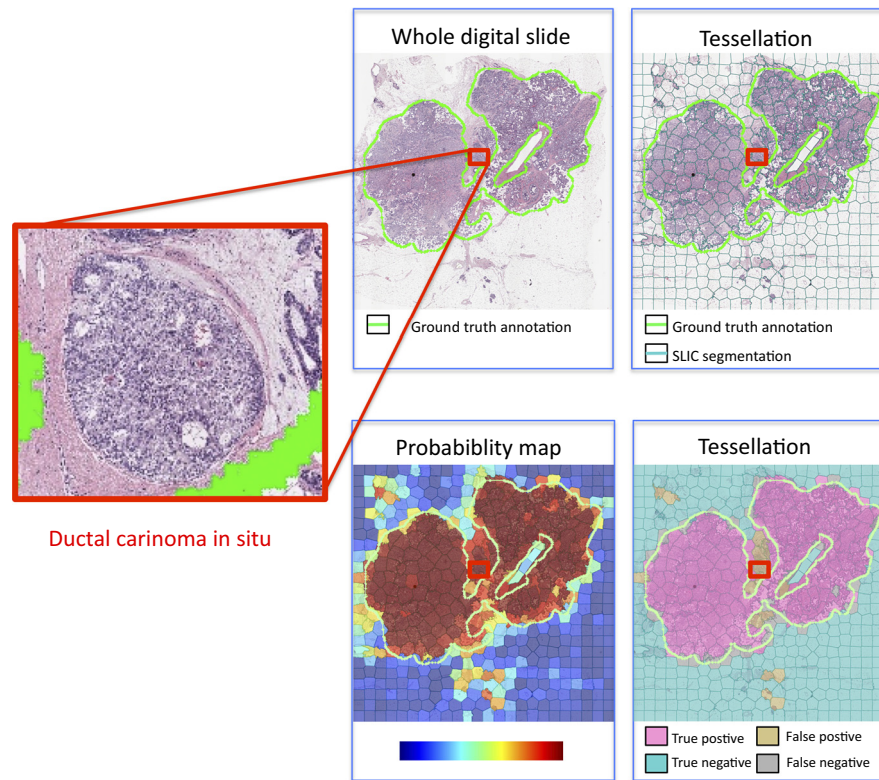
Fig. 7 An example of a false positive result occurring in a region expressing DCIS: (a) the original whole digital slide, (b) tessellation using SLIC segmentation, (c) the probability map of cancer, and (d) the final decision with an RF$_{Confidence}$ of 0.6%.

Moreover, increasing the magnification produced a meager performance gain considering the additional computation burden it imposed. Therefore, researchers can leverage this robustness to significantly reduce computation time.

A close inspection of the results indicates certain common modes of failure, notably in terms of the false positive rate. Generally, the classifier responded strongly to areas where epithelial cells proliferated in a disorganized fashion. Although this is a significant characteristic of IDBC, other histopathological entities sometimes expressed similar traits. For example, DCIS differs from IDBC only in the specific locations in which the cells proliferate with respect to the breast ducts. Otherwise, both manifest similar nuclear morphology and cellular properties. Therefore, distinguishing the two would be difficult given only color and texture image features. Future improvements should include a subsequent step that distinguishes IDBC from other unhealthy tissues in order to further reduce the false positive rates.

4.2 Toward a Clinical Application

The results produced for this paper were more than satisfactory for our ultimate objective. Automatically grading IDBC would require additional components that measure nuclear pleomorphism, tubule formation, and mitosis (Fig. 2).

We believe that automated histopathological grading will eventually benefit patients in several ways. It can alleviate the workload in large pathological centers, or alternatively provide a solution for areas lacking trained pathologists, such as in the developing world. All the while, it will produce results that are consistent and unbiased, providing uniform care for all patients.

References

1. Canadian Cancer Society's Avisory Committee on Cancer Statistics, *Canadian Cancer Statistics 2015*, Statistics Canada, Toronto, Ontario (2015).
2. C. R. Eheman et al., "The changing incidence of in situ and invasive ductal and lobular breast carcinomas: United States, 1999-2004," *Cancer Epidemiol. Biomarkers Prev.* **18**, 1763–1769 (2009).
3. G. Arpino et al., "Infiltrating lobular carcinoma of the breast: tumor characteristics and clinical outcome," *Breast Cancer Res.* **6**(3), R149–156 (2004).
4. Union for International Cancer Control, *Classification of Malignant Tumors*, Wiley, New York (2002).
5. World Health Organization, *International Histological Classification of Tumours. Histologic Types of Breast Tumours*, World Health Organization, Geneva (2002).
6. E. A. Rakha et al., "Prognostic significance of Nottingham histologic grade in invasive breast carcinoma," *J. Clin. Oncol.* **26**, 3153–3158 (2008).
7. L. W. Dalton, D. L. Page, and W. D. Dupont, "Histologic grading of breast carcinoma. a reproducibility study," *Cancer* **73**, 2765–2770 (1994).
8. H. F. Frierson et al., "Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma," *Am. J. Clin. Pathol.* **103**, 195–198 (1995).

9. P. Robbins et al., "Histological grading of breast carcinomas: a study of interobserver agreement," *Hum. Pathol.* **26**, 873–879 (1995).

10. T. R. Fanshawe et al., "Assessing agreement between multiple raters with missing rating information, applied to breast cancer tumour grading," *PLoS One* **3**(8), e2925 (2008).

11. A. H. Beck et al., "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Sci. Transl. Med.* **3**(108), 108ra113 (2011).

12. A. Veillard, N. Lomnie, and D. Racoceanu, "An exploration scheme for large images: application to breast cancer grading," in *20th Int. Conf. on Pattern Recognition*, pp. 3472–3475 (2010).

13. A. Veillard, S. Bressan, and D. Racoceanu, "SVM-based framework for the robust extraction of objects from histopathological images using color, texture, scale and geometry," in *11th Int. Conf. on Machine Learning and Applications*, Vol. 1, pp. 70–75 (2012).

14. L. Roux et al., "Mitosis detection in breast cancer histological images. An ICPR 2012 contest," *J. Pathol. Inform.* **4**, 8 (2013).

15. S. Doyle et al., "A boosting cascade for automated detection of prostate cancer from digitized histology," *Lect. Notes Comput. Sci.* **4191**, 504–511 (2006).

16. A. Cruz-Roa et al., "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," *Proc. SPIE* **9041**, 904103 (2014).

17. R. Achanta et al., "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012).

18. S. Kothari et al., "Biological interpretation of morphological patterns in histopathological whole-slide images," in *Proc. of the ACM Conf. on Bioinformatics, Computational Biology and Biomedicine*, pp. 218–225, ACM, New York (2012).

19. H. Kalkan et al., "Automated colorectal cancer diagnosis for whole-slice histopathology," *Lect. Notes Comput. Sci.* **7512**, 550–557 (2012).

20. F. Rodriguez-Telles, L. Abril Torres-Mendez, and E. Martinez-Garcia, "A fast floor segmentation algorithm for visual-based robot navigation," in *Int. Conf. on Computer and Robot Vision*, pp. 167–173 (2013).

21. S. Sahli, D. A. Lavigne, and Y. Sheng, "Saliency region selection in large aerial imagery using multiscale SLIC segmentation," *Proc. SPIE* **8360**, 83600D (2012).

22. A. Namburete and J. Noble, "Fetal cranial segmentation in 2D ultrasound images using shape properties of pixel clusters," in *IEEE 10th Int. Symp. on Biomedical Imaging*, pp. 720–723 (2013).

23. M. Balazsi, "An automated approach for segmenting regions containing invasive ductal breast carcinomas in whole digital slides," Master's Thesis, McGill University, Montreal, Quebec, Canada (2014).

24. C. Connolly and T. Fleiss, "A study of efficiency and accuracy in the transformation from RGB to CIELAB color space," *IEEE Trans. Image Process.* **6**, 1046–1048 (1997).

25. J. Han and K.-K. Ma, "Fuzzy color histogram and its use in color image retrieval," *IEEE Trans. Image Process.* **11**, 944–952 (2002).

26. A. Ferman, A. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," *IEEE Trans. Image Process.* **11**, 497–508 (2002).

27. M. Pietikinen, "Image analysis with local binary patterns," *Lect. Notes Comput. Sci.* **3540**, 115–118 (2005).

28. T. Lindahl, "Study of local binary patterns," Master's Thesis, Linkping University, Department of Science and Technology (2007).

29. Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognit.* **43**(3), 706–719 (2010).

30. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 886–893 (2005).

31. Q. Zhu et al., "Fast human detection using a cascade of histograms of oriented gradients," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 1491–1498 (2006).

32. O. Dniz et al., "Face recognition using histograms of oriented gradients," *Pattern Recognit. Lett.* **32**(12), 1598–1603 (2011).

33. A. Newell and L. Griffin, "Multiscale histogram of oriented gradient descriptors for robust character recognition," in *Int. Conf. on Document Analysis and Recognition*, pp. 1085–1089 (2011).

34. L. Breiman, "Random forests," *Mach. Learn.* **45**, 5–32 (2001).

35. A. Verikas, A. Gelzinis, and M. Bacauskiene, "Mining data with random forests: a survey and results of new tests," *Pattern Recognit.* **44**(2), 330–349 (2011).

36. K. Brodersen et al., "The balanced accuracy and its posterior distribution," in *20th Int. Conf. on Pattern Recognition*, pp. 3121–3124 (2010).

37. D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation," *J. Mach. Learn. Technol.* **2**(1), 37–63 (2011).

38. X. Han and B. Fischl, "Atlas renormalization for improved brain MR image segmentation across scanner platforms," *IEEE Trans. Med. Imaging* **26**, 479–486 (2007).

39. K. H. Zou et al., "Statistical validation of image segmentation quality based on a spatial overlap index: scientific reports," *Acad. Radiol.* **11**(2), 178–189 (2004).

40. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. of the 14th Int. Joint Conf. on Artificial Intelligence*, Vol. 2, pp. 1137–1143, Morgan Kaufmann Publishers Inc., San Francisco, California (1995).

41. T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.* **27**(8), 861–874 (2006).

Matthew Balazsi received his BEng and MEng degrees in electrical and computer engineering from McGill University, Montreal, in 2008 and 2014 respectively. He is currently the founder of Medical Parachute, a company that offers cloud-based data management and analytical tools for medical researchers.

Paula Blanco is a staff pathologist at The Ottawa Hospital, Ottawa, Ontario, Canada and an assistant professor of Pathology and Laboratory Medicine at the University of Ottawa. She is also the current Undergraduate Program Director for pathology at the University of Ottawa in Ottawa, Ontario, Canada.

Pablo Zoroquiain is an assistant professor in the Department of Pathology at Pontificia Universidad Catolica de Chile, Santiago, Chile, where he completed his residency in anatomical pathology. He is currently pursuing a clinical and research fellowship in ocular pathology at McGill University, Montreal. During his fellowship, he has received 3 awards from McGill University and Cedars Cancer Institute. He is the author of 30 publications and has been the guest speaker in several international meetings.

Martin D. Levine received his BEng and MEng degrees in electrical and computer engineering from McGill University, Montreal, in 1960 and 1963, respectively, and his PhD in electrical engineering from the Imperial College of Science and Technology, University of London, London, England, in 1965. He is currently a professor in the Department of Electrical and Computer Engineering, McGill University, and served as the founding director of the Center for Intelligent Machines from 1986 to 1998.

Miguel N. Burnier Jr. was the chairman of the Department of Ophthalmology, McGill University from 1993 to 2008. He is currently a full professor of ophthalmology, pathology, medicine, and oncology at McGill, and he was the Thomas O. Hecht Family Chair in Ophthalmology (1996 to 2012). He is currently the general director of Clinical Research of the McGill University Health Centre Research Institute. He is author and coauthor of over 500 publications, including 300 peer-reviewed papers.