Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 90 (2016) 125 – 131

Procedia Computer Science

International Conference On Medical Imaging Understanding and Analysis 2016, MIUA 2016, 6-8 July 2016, Loughborough, UK

Incorporating local and global context for better automated analysis of colorectal cancer on digital pathology slides

Alexander I. Wright[a]*, Derek Magee[b], Philip Quirke[a], Darren Treanor[a,c]

[a]*Section of Pathology and Tumour Biology, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds LS9 7TF, UK*
[b]*School of Compiuting, Faculty of Engineering, University of Leeds, Leeds LS2 9JT, UK*
[c]*Leeds Teaching Hospitals Trust, Leeds LS9 7TF, UK*

Abstract

Phenotypic information derived from visual characteristics of colorectal cancer (CRC) is routinely used for diagnosis and recommendations for treatment. Previously published studies show that the ratio of tissue types within CRC is prognostic. Such studies generate large amounts of data, combining expert classifications with x-y coordinates, which has previously been used to train image analysis algorithms. This paper describes extensions to algorithms employed in previously published work, using pixel clustering as a pre-processing step before normalised cuts in order to reduce the size of the graph for unsupervised segmentation. Image segments are processed for features and given a candidate classification which is weighted by neighbouring segment classes. Global slide features are incorporated to mitigate inconsistencies in overall appearance caused by histological and biological differences. The proposed algorithm increases agreement with the ground truth from 75% to 79% on a dataset of 7,159 images across 157 digital slides.

Peer-review under responsibility of the Organizing Committee of MIUA 2016

Keywords: digital pathology; colorectal cancer; automated analysis; unsupervised segmentation; contextual analysis;

1. Introduction

With over 40,000 new cases and 16,000 deaths per year, colorectal cancer (CRC) is currently the second highest cause of cancer mortality in the UK[1]. Histopathological examination of cancer tissue provides pathologists with phenotypic information from visual characteristics of the disease[2] which is used to predict response to therapies.

* Corresponding author. Tel.: +44-113-343-8509; fax: +44-113-343-8431.
E-mail address: a.wright@leeds.ac.uk

Peer-review under responsibility of the Organizing Committee of MIUA 2016
doi:10.1016/j.procs.2016.07.034

These predictions facilitate the provision of appropriate treatments for individual patients, possibly avoiding exposure to toxic radiotherapies and expensive drugs. This visual information is traditionally acquired via a glass slide tissue sample and microscope, but the rise of high resolution digital pathology scanners allows pathologists to inspect tissue using standard computer screens and software tools for more consistent and quantifiable manual analysis. The RandomSpot stereology tool[3] is used routinely at a number of institutions and has been used by pathologists to help identify the prognostic capabilities of the ratio of tumour to stroma within a patient's cancer (T:S)[4]. This is a time consuming and laborious task which requires pathologists to classify several hundred points on a single slide in order to obtain an appropriate sample size, and therefore automation of the task is highly desirable.
Previous work used existing clinical data to extract regular sized image patches at specific co-ordinates, with the associated pathologist-classified ground truth label[5]. Results showed that the automated analysis of the images was more accurate on smaller patches (64x64px), which was inconsistent with pathologist scoring, yielding significantly higher human agreement on larger images (≥256x256px). The following conclusions were drawn:

1) Algorithm accuracy was lower on larger image patches because they were more likely to contain multiple classes of tissue within them
2) Pathologist accuracy was lower on smaller images because the visual information surrounding the patch (context) is important for making decisions

This paper reports on work undertaken in order to compensate for these two issues by applying unsupervised segmentation in larger image patches, and including contextual information to assist with machine learning. Normalised cuts[6] is a graph based segmentation algorithm which uses both similarity and dissimilarity metrics to partition a graph into two or more sub-graphs. Images are treated as regular graphs which can be segmented with globally optimised clustering, but requires computationally expensive per-pixel, pairwise comparisons. Tao et al[7] propose a weighted mean-shift algorithm to reduce the colourspace of images before the application of normalised cuts. By clustering the image into similar coloured areas, the size of the graph (and complexity of the affinity matrix) is greatly reduced, facilitating the application of the algorithm to complex histological slide images[8]. This work has been extended further by its application to a hierarchical pyramid for the analysis of ovarian cancer tissue microarrays[9]. The normalised cuts algorithm has also been modified for use in this field by include adding extra features to improve accuracy of segmentation on gastroeneterology images[10]. Other approaches to segmentation in colorectal histology images include weakly supervised algorithms for learning gland or nuclear shape[11] and sparse dictionary based representations of structures[12]. In most cases, approaches involve either pixel level clustering or model fitting[13]. Due to the ground truth data relating to single x-y coordinates rather than objects and structures, there is a need for unsupervised segmentation so that these co-ordinates can be expanded into tissue regions before analysis of features.

Contextual information can be incorporated at different resolutions in order to provide a more complete description of the visual space. In histopathology, local context typically relates to neighbouring tissue classes and the pattern that the tissue forms (or lack thereof). Global context describes the overall condition of the tissue and/or slide, the level of staining and the type of cancer being analysed, which helps the pathologist to understand the visual information at a microscopic level. These relationships between resolutions have previously been mapped in automated solutions using Bayesian networks[14], label regularisation[15], rotationally invariant contextual analysis (spin-context)[16], or simply providing a 'context vector' as a set of features in order to pre-analyse images[17].

2. Methods

2.1 Data

The experiments reported in this paper use a subset of an existing dataset from one colorectal cancer trial[18]. The dataset contains image data from 2,214 cases, which comprises of mostly of stage II cancer patients, with the remaining cases being stage III. Half of the patients received chemotherapy and the other half did not. Most cases had only one glass slide digitally scanned, using an Aperio AT scanner at 20x magnification (0.5 µm per pixel), stained with haematoxylin and eosin stains (H&E). For this study, the subset was comprised of 157 cases, which had been preselected by a pathologist in order to provide a dataset representative of typical workload), and each case had been analysed using the RandomSpot stereology tool, using a target number of 50 spots per digital slide (with a

fitting tolerance of 15%). These cases contained a total of 7,159 pre-scored co-ordinates, which were extracted as images whereby the co-ordinate was at the centre of the image. All images were classified into one of eight types: tumour; lumen; necrosis; mucin; stroma; muscle; inflammation; blood vessels.

2.2 Processing Overview

Based on the conclusions of a previous study[5], image patches were extracted at 256x256 pixels in order to preserve surrounding contextual information. Initial attempts used colour normalisation[19] in order to compensate for variation in staining intensities, but was removed so that staining information could be preserved for analysis using global context (see section 2.6). An implementation of the colour deconvolution algorithm[20] was used to separate stains before unsupervised segmentation. Once segmented, features were extracted from the centre segment (where the pathologist classification had been applied to), and a random forest[21] algorithm was trained on the feature set (see machine learning section). Once trained, the algorithm was then applied to all segments, which were used for surrounding contextual analysis.

2.3 Unsupervised Segmentation

Oversegmentation using simple linear iterative clustering (SLIC)[22] was applied to each image (Figure 1 c), dividing into approximately 168 superpixels, which resulted in each superpixel being approximately the size of one nucleus (10μm^2, given that the slides were scanned at 0.5μm per pixel). For every superpixel pair, a similarity metric was computed. This metric consisted of Euclidean distance, median intensity difference, mean absolute difference of intensity, texture difference (comparing textons[23]) and maximum intervening contours. Similarity metrics were used to create a symmetric affinity matrix, and normalised graph cuts was applied in order to cluster the superpixels into image segments. A fixed number of sixteen segments was chosen in order to create images approximately 64x64px in size, which yielded the highest accuracy result using the original algorithm.

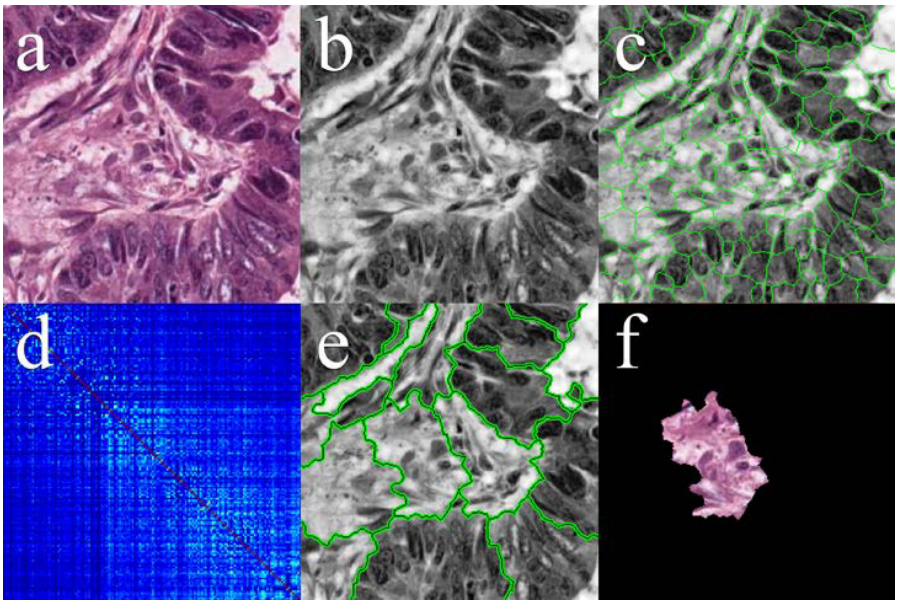

Figure 1 - The SLIC and normalized graph cut process with segment classification (a) original image patch (256x256px), (b) Haematoxylin channel after colour deconvolution, (c) SLIC superpixel segmentation (d) affinity matrix depicting similarity of superpixel pairs, (e) image segments comprised of clustered superpixels, using normalised cuts, (f) segment at the centre of the image that the pathologist score applies to.

2.4 Machine Learning

After segmentation, image segments were processed individually to generate features, to be used in machine learning based classification. Features used were median hue, saturation and intensity values after converting the RGB image to HSV colour space, median intensity of haematoxylin and eosin staining channels after colour deconvolution, median staining values after removing background pixels (value = 240 set by the digital slide

scanning software), percentage foreground intensity for H&E channels, standard deviation of intensity, textural properties (energy and homogeneity) using gray level co-occurrence matrices[24], and nuclear count, detection and mean area[25]. For training, features were extracted on the center segment of each image, so that it could be associated with pathologist classification at that co-ordinate. A random forest was trained using the features and a binary classification of tumour or stroma, whereby each of the possible eight classifications were grouped into the parent class. The random forest algorithm used 500 trees with three predictors sampled for splitting at each node. Cases were randomly grouped for ten-fold cross validation (as opposed to randomly grouping individual images) in order to test the accuracy of the algorithm using unsupervised segmentation alone.

2.5 Local Context

In order to improve predictions made by the random forest algorithm in segments where confidence was low, the trained algorithm was then applied to all segments within a given image patch, providing a probability of tumour (see heatmap in Figure 2 b). Neighbouring segment classification probabilities were then used to weight the classification probability of the centre segment, based on the assumption that a single segment is more likely to be the same class as the majority of its neighbors. By using the product of the surrounding tissue class probabilities, weighted by the length of the adjacent perimeter length as a percentage, artefacts from oversegmented regions are avoided. It is important to note that more complex contextual cues and biological spatial rules could be included at this stage, if using more tissue classes.

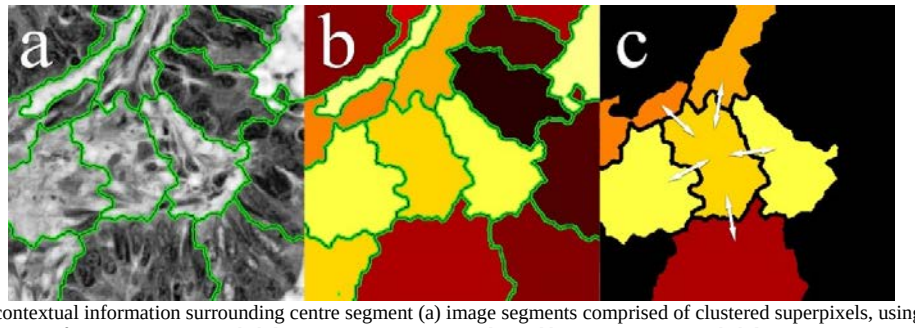

Figure 2 – Local contextual information surrounding centre segment (a) image segments comprised of clustered superpixels, using normalised cuts (b) colour heatmap of tumour segment probability (c) centre segment with neighbouring segment probabilities

2.6 Global Context

There are many factors involved when considering the global appearance of histological tissue samples. Most notable is the inconsistency in colour, which can be attributed to issues such as section thickness, variations in staining chemical compounds, application of the stains, quality of the glass slides and cover slips, length of time between slide generation and digital slide scanning, slide scanning hardware, and colour profiles of the digital scanners to name a few. The clinical dataset used was designed to provide five year survival data, and as such, has large variation in appearance (see Figure 3a).

This variation becomes problematic when using image and stain intensities as training features, and therefore must be compensated for in the algorithm. Colour normalisation provides a robust method for correcting for inconsistent staining in tissue, however overcompensates in areas that do not require normalising, such as larger areas of mucin, lumen, tissue tearing or retraction artefact. Figure 3 (b) plots the average (mean) median intensity of tumour and stroma images patches (y axis) against the median intensity of their parent whole slide (x axis) at low power (scaled to 3x objective zoom). This indicates that the visual information derived from local features can be enhanced when combined with global features. It is therefore proposed that using visual features of global contextual information in conjunction with the local image features will give a reliable comparative measure, with respect to colour variation. The global features used were median intensity, median haematoxylin and eosin staining channel intensity and median haematoxylin staining channel intensity after applying a simple threshold to remove background pixels (mean intensity minus half of one standard deviation intensity). These global features were calculated prior to image patch analysis, then added to the feature vector for each image patch respectively before training and testing.

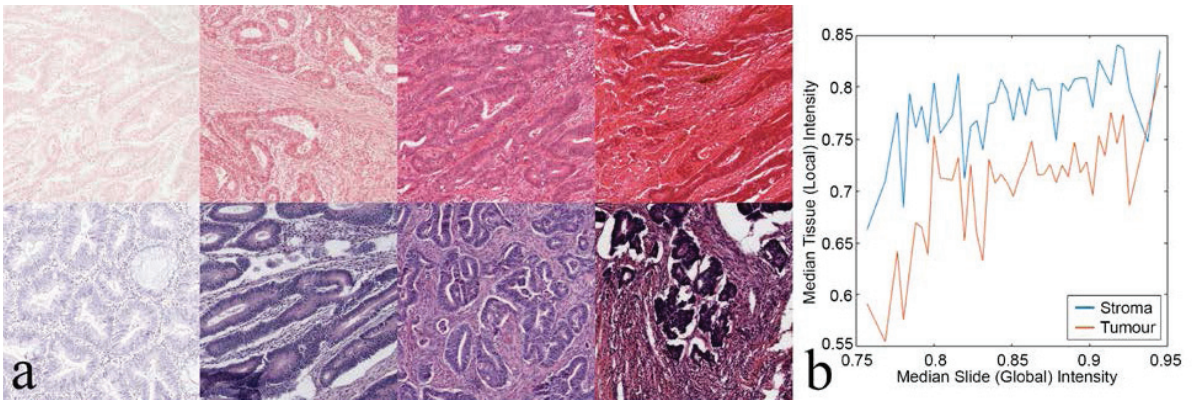

Figure 3 - (a) examples of different levels of staining in the clinical dataset (b) plot of median slide intensity compared to median tumour and stroma intensity within that slide

3. Results

To provide a comparable metric, the algorithm from the previous work was run on the new subset of data (see 2.1 Data), in order to assess improvements in accuracy. The dataset of 7,159 images from 157 slides was processed using ten-fold cross validation for training and testing the random forest algorithm, and agreement between the original pathologist score and algorithm was recorded per image. Once trained, the algorithm took approximately 9.2 minutes to analyse one case containing 50 spots (11 seconds per image). Figure 4 (a) shows the accuracy of the context based algorithm compared to the baseline algorithm, using boxplots of both of the ten folds. The increase in accuracy of .75 to .79 was shown to be statistically significant using a Wilcoxon rank sum test ($p < .01$).

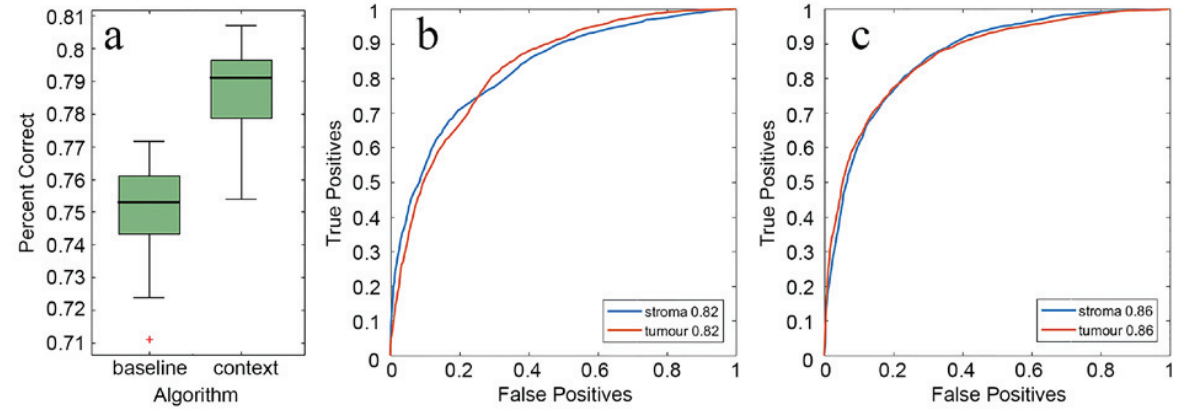

Figure 4 – Results for both baseline and context-based algorithm (a) box plots comparing accuracy of both algorithms ten-fold cross validation (b) ROC curves for baseline algorithm tumour AUC = .82, stroma AUC = .82 (c) ROC curves for context based algorithm tumour AUC = .86, stroma = .86

Table 1 (left) and Table 2 (right) - Confusion matrices showing mean pathologist – baseline algorithm agreement. Baseline algorithm: Accuracy = .75, sensitivity = .75, specificity = .75, Kappa = .5 (moderate agreement). Context algorithm: Accuracy = .79, sensitivity = .79, specificity = .79, Kappa = .57 (moderate agreement).

		Baseline Algorithm	
		Tumour	**Stroma**
Pathologist	**Tumour**	2132	833
	Stroma	573	2111

		Context Algorithm	
		Tumour	**Stroma**
Pathologist	**Tumour**	2410	555
	Stroma	661	2023

Table 1 and 2 show confusion matrices, illustrating the baseline algorithm over-estimates the presence of stroma, and the context-based algorithm overestimates tumour. The context-based algorithm improves agreement (kappa = .5 increased to .57). Figure 4 (b) and (c) show ROC curves for the baseline and context-based algorithms, which show area under the curve of .82 and .86 respectively. It should be noted that these results are not directly comparable to previous work as a smaller dataset was used.

4. Discussion and Conclusion

Automated analysis of colorectal cancer is non-trivial due to the many biological, histological and technical considerations that affect the variation in appearance of tissue on digital slides. This variation affects the appearance of slides in terms of colour, texture and shape, and so model based segmentation may only be appropriate for certain types or grades of cancer, where appearances are similar. Unsupervised segmentation allows for images to be split into regions that are similar in appearance and can be classified based on their visual characteristics, which provides more flexibility for the variable appearance of cancer tissue. The current algorithm shows an improvement over previous work, which used smaller patch sizes instead of unsupervised segmentation, and did not allow for surrounding contextual information to be included. The inclusion of unsupervised segmentation meant that training images were much less likely to include more than one tissue class per segment, thus reducing ambiguity in the feature set. Unsupervised segmentation was applied by initially discretising the pixel data of each image patch into a much smaller set of superpixels, which roughly represented the size of one nucleus. This resulted in a graph cut problem which is less computationally expensive than calculating similarities per-pixel, whilst maintaining segment boundaries at a nuclear level. By applying segmentation to the image patches, the surrounding segment classes provide contextual information which was used to weight the probability of the random forest prediction. In previous work, this local contextual information has been shown to be important to pathologists for scoring images, and therefore should be a consideration when developing image analysis solutions to automate the pathologist task. Currently the algorithm uses the assumption that higher amounts of surrounding tissue increase the likelihood of that same classification. This could be extended to include important biological structural rules, for example, lumen must be surrounded by tumour, or that a candidate lumen segment lying between stroma and tumour is more likely to be a retraction artefact. Contextual information was also applied at the macro level, on order to provide information about the slide as a whole. This global information is important when considering sets of data that have been stained inconsistently, or in this case, as part of a five year longitudinal study, where the fading of stains before digital slide scanning was inevitable. By combining the features of the slide with the feature set for each image patch, the algorithm is able to learn that the relationships between tumour and stroma features will change, based upon such global features. This work could be extended to incorporate a prediction of the type of cancer being analysed, which could help improve classification accuracy. For example, mucinous adenocarcinomas are considered mostly tumour, but the light appearance of the mucin is likely to be classified as weakly stained stroma at the microscopic level. Finally, it is worth noting that the algorithm was trained and tested on a subset of one clinical trial dataset, and therefore accuracy of the algorithm is likely to increase with the expanded dataset.

Acknowledgements

The author’s PhD research is funded by Yorkshire Cancer Research.

References

1 Cancer Research UK, “Bowel cancer incidence statistics - Cancer Research UK.” 2013.
2 Todman A., Naguib R. N. G, Bennett M. K., “Visual characterisation of colon images,” in *Proceedings of Medical Image Understanding and Analysis*, 2001, pp. 161–164.
3 Wright A. I., Grabsch H. I., Treanor D. E., “RandomSpot: A web-based tool for systematic random sampling of virtual slides,” *J. Pathol. Inform.*, pp. 56–60, 2015.
4 West N. P., Dattani M., McShane P., Hutchins G., Grabsch J., Mueller W., Treanor, D. Quirke P., Grabsch H., “The proportion of tumour cells is an independent predictor for survival in colorectal cancer patients.,” *Br. J. Cancer*, vol. 102, no. 10, pp. 1519–23, May 2010.
5 Wright A., Magee D., Quirke P., Treanor D. E., “Towards automatic patient selection for chemotherapy in colorectal cancer trials,” in *SPIE*

9041, Medical Imaging 2014: Digital Pathology, 2014, p. 90410A.
6 Shi J., Malik J., "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.
7 Tao W., Jin H., Zhang Y., "Color image segmentation based on mean shift and normalized cuts.," *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 37, no. 5, pp. 1382–9, 2007.
8 Xu J., Madabhushi A., Janowczyk A., Chandran S., "A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: applications to histopathology image segmentation," *Proc. SPIE*, vol. 7623, no. April 2016, p. 76230Y–76230Y–12, 2010.
9 Janowczyk A., Chandran S., Singh R., Sasaroli D., Coukos G., Feldman M. D., Madabhushi A., "High-throughput biomarker segmentation on ovarian cancer tissue microarrays via hierarchical normalized cuts," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1240–1252, 2012.
10 Devaraj S. A., Kumarasamy T.,. Ganesh V. B. G, Ranjan C. C. C., Narayanan A. G., Prof A., Scholar B. E., Xavier F., "Normalized cuts based segmentation of gastroentrology images using visual features," vol. 1, no. 11, pp. 1376–1382, 2014.
11 Xu, Y. Zhang, J., Chang E. I.-C., Lai M., Tu Z., "Context-constrained multiple instance learning for histopathology image segmentation.," *Int. Conf. Med. Image Comput. Comput. Interv. MICCAI*, vol. 15, no. Pt 3, pp. 623–30, Jan. 2012.
12 Gao Y., Liu W., Arjun S., Zhu L., Ratner V., Kurc T., Saltz J., Tannenbaum A., "Multi-scale learning based segmentation of glands in digital colonrectal pathology images," *Proc. SPIE*, vol. 9791. p. 97910M–97910M–6, 2016.
13 Sirinukunwattana K., Pluim, J. P. W., Chen H., Qi X., Heng P.-A., Guo Y. Wang B., L. Y., Matuszewski B. J., Bruni E., Sanchez U., Böhm A., Ronneberger O.,. Ben Cheikh B, Racoceanu D., Kainz P., Pfeiffer M., Urschler M., Snead D. R. J., Rajpoot N. M., "Gland Segmentation in Colon Histology Images: The GlaS Challenge Contest," pp. 1–24, 2016.
14 Chomphuwiset P., Magee D. R., Boyle R. D., andTreanor D., "Context-Based Classification of Cell Nuclei and Tissue Regions in Liver Histopathology," in *Medical Image Understanding and Analysis*, 2011, pp. 1–5.
15 Roullier V., Lézoray O., Ta V. T., Elmoataz A., "Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization," *Comput. Med. Imaging Graph.*, vol. 35, no. 7–8, pp. 603–615, 2011.
16 McKenna S. J., Amaral T., Akbar S., Jordan L., Thompson A., "Immunohistochemical analysis of breast tissue microarray images using contextual classifiers.," *J. Pathol. Inform.*, vol. 4, no. Suppl, p. S13, Jan. 2013.
17 Magee D., Treanor D., Chomphuwiset P., Quirke P., "Context Aware Colour Classification in Digital Microscopy," in *Medical Image Understanding and Analysis*, 2011, pp. 1–5.
18 QUASAR Collaborative Group, "Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study.," *Lancet*, vol. 370, no. 9604, pp. 2020–9, Dec. 2007.
19 Magee D., Treanor D., Crellin D., Shires M., Mohee K., Quirke P., "Colour Normalisation in Digital Histopathology Images," in *Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*, 2009, pp. 100–111.
20 Ruifrok A. C. & Johnston D. A., "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, 2001.
21 Breiman L., "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
22 Achanta R., Shaji A., Smith K., "SLIC superpixels compared to state-of-the-art superpixel methods," *Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, 2012.
23 Malik J., Belongie S., Leung T., Shi J., "Contour and texture analysis for image segmentation," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 7–27, 2001.
24 Haralick R. M., Shanmugam K., Dinstein I., "Textural features for image classification," *IEEE Trans. Syst. Man. Cybern.*, vol. 3, no. 6, pp. 610–621, 1973.
25 Zhou Y., Magee D., Treanor D., Bulpitt A., "Stain guided mean-shift filtering in automatic detection of human tissue nuclei," *J. Pathol. Inform.*, vol. 4, no. 2, p. 6, 2013.