Scoring of breast tissue microarrays using ordinal regression: local patches vs. nuclei segmentation

Telmo Amaral[a], Michele Sciarabba[b], Stephen McKenna[a,*], Katherine Robertson[c], and Alastair Thompson[d]

[a]School of Computing, University of Dundee, Dundee, UK,
[b]Department of Human Morphology and Biomedical Sciences "Città Studi", University of Milan, Milan, Italy,
[c]Pathology and Neuroscience, University of Dundee, [d]Surgery and Molecular Oncology, University of Dundee

Abstract. Breast tissue microarrays (TMAs) facilitate the study of very large numbers of breast tumours per histological section, but their scoring by pathologists is time consuming, prone to observer variability, and not without error. This paper reports the use of ordinal regression to predict the scores of TMA spots subjected to progesterone receptor immunohistochemistry. We compare the use of global features obtained via two different methods, one involving and the other dispensing with accurate segmentation of epithelial cell nuclei. In addition, we investigate the effect of analysing only regions of interest (ROIs) within each spot, as opposed to analysing the whole spots. The use of the entropy of the posterior probability distribution over category labels for avoiding uncertain decisions is demonstrated.

1 Introduction

Tissue microarrays (TMAs) are a high-throughput technique proposed by Kononen et al. [1], to facilitate the study of large numbers of tissue samples. TMAs are constructed by taking typically a few hundred cylindrical biopsies (named "cores") from donor blocks of formalin-fixed, wax-embedded tissue and inserting them into a recipient wax block in a grid arrangement. Sections of the TMA block are cut, so that each section contains an array of tissue "spots" (each spot being a section of one of the cores previously inserted into the microarray block). TMA sections provide targets for parallel in situ detection of DNA, RNA, and protein targets in each specimen on the array. TMAs are now extensively utilised in the study of cancers.

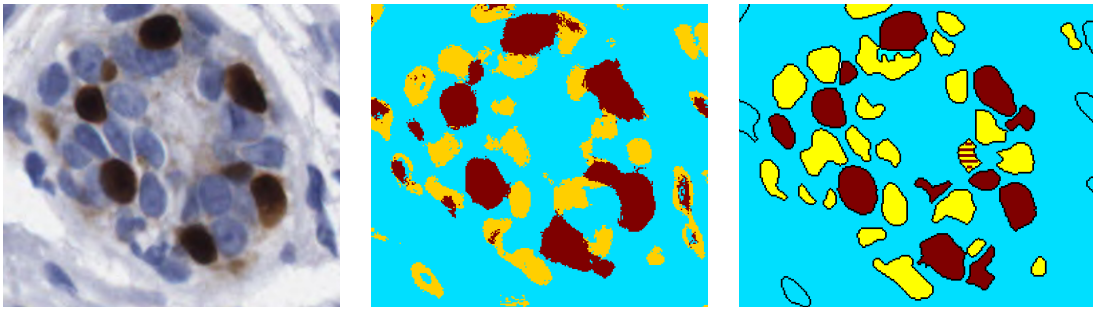
Immunohistochemistry is carried out to detect protein expression in the tissue spots. For example, antibodies directed against progesterone receptor (PR) can be used to detect nuclear expression of PR in breast tumours. The spots are then observed one by one under the microscope and the degree of immunostaining of those spots that contain either normal or tumour tissue is assessed and assigned a score. The scores of breast tissue spots may take the form of Quickscores [2], each composed of two integer values, namely a value between 0 and 6 that estimates the proportion of epithelial nuclei that are immunopositive, and a value between 0 and 3 that estimates the strength of staining of those nuclei (these values will henceforth be referred to as QSP and QSS, respectively). In practice and in general, pathologists do not score each spot as a whole, but rather focus their attention on a certain region of interest (ROI) within the spot.

Applying the above procedure to breast TMA sections incorporating large numbers of tissue samples is time consuming and suffers from inter- and intra-observer variability and perceptual errors. Thus, there is strong motivation for the development of automated methods for quantitative analysis of breast TMA image data. However, accurate segmentation of cells and intra-cellular compartments in images of tissue sections can be problematic for reasons that include cell overlapping, complex tissue structure, debris, and variable appearance.

Most of the published work on automated "ranking" of breast tissue sections aims not at predicting immunohistochemical scores, but at distinguishing between different Bloom-Richardson grades of sections stained with haematoxylin & eosin [3–8]. In contrast, Kostopoulos et al. [9] applied k-nearest neighbour weighted votes classification to colour-textural features, in order to predict the oestrogen receptor's positive status of biopsy images – a procedure that takes into account the percentage of epithelial nuclei that are immunopositive.

In this paper we report the use of ordinal regression to predict the Quickscores of breast TMA spots subjected to progesterone receptor immunohistochemistry, which results in nuclear staining in positive cases. Ordinal regression differs from classification, as the existence of an order between the target categories (such as Quickscore values) is taken into account. We compare the use of global features obtained via two different methods, one based on local patch statistics and the other involving explicit nuclear segmentation. Moreover, we investigate the effect of analysing only regions of interest within each spot, as opposed to analysing the whole of each spot.

*School of Computing, University of Dundee, DD1 4HN Dundee, UK, stephen@computing.dundee.ac.uk



(a) (b) (c)

Figure 1. (a) A small tissue region and intermediate results from the two methods used for extraction of global features: (b) local patch-based pixel classification (dark grey: E+ pixels; light grey: E- pixels; mid-grey: background pixels); and (c) nuclear segmentation (dark grey: E+ nuclei; light grey: E- nuclei; striped: nuclei near to the decision boundary; mid-grey: background).

2 Data

Colour images of 175 breast TMA spots from the National Cancer Research Institute's Adjuvant Breast Cancer (ABC) Chemotherapy Trial [10] with resolution of 0.23 μm / pixel and typical spot diameter of $700\mu m$ (i.e. about 3000 pixels) were used. These contained tissue subjected to progesterone receptor immunohistochemistry. For each spot, Quickscores assigned by a pathologist on two separate occasions were available. A manual annotation of the region of interest associated with each spot was also available. A further 20 spots were used for training only. A circular sub-region (500 pixels in diameter) was randomly selected on each such spot and all epithelial nuclei (approximately 700) within these sub-regions were manually segmented and labelled as immunonegative or immunopositive.

3 Global feature extraction

3.1 Local patch-based method

This feature extraction method involved no explicit segmentation of epithelial nuclei, and has been described in more detail by Amaral et al. [11]. The 20 sub-regions containing manually annotated epithelial nuclei were used to estimate class-conditional distributions for three classes, namely immunopositive epithelial nuclei (E+), immunonegative epithelial nuclei (E-), and background. These were conditioned on the colour and differential invariants (up to 2nd order) computed at a local patch centred on each pixel. Bayes' rule was then used to estimate posterior class probabilities given the local patch centred on a pixel. Figure 1(b) illustrates the results for the small tissue region shown in Figure 1(a), once each pixel was classified to the class with largest posterior probability.

These local posterior probabilities were used to obtain a pair of global features which aimed to formalise the two Quickscore values assigned by pathologists. The QSP was formalised as the fraction of epithelial pixels (both E- and E+) that were E+, and the QSS was formalised as the average posterior probability of E+ computed over all E+ pixel locations.

3.2 Nuclei segmentation-based method

This method was based on a multi-level algorithm for monochromatic images that explicitly segments nuclei [12]. Given an image A, 16 nested binary images A_i (levels) are created such that $A_i(x, y) = 1$ iff $A(x, y) \leq T_i$, where $T_i = (i - 1) \times 16$. Connected components at each level form a tree structure (each component having as its parent a component that contains it). For each component we compute a global shape index, based on elongation and solidity, that tells us if the component has a shape compatible with the nuclear shape. We call "cores" the leaves of the tree that have a regular shape, as they are the darkest regular regions of each component in the original image A. These cores are grown level by level as far as they keep a regular shape and do not join with other components.

This algorithm was applied to four different monochromatic images derived from the original RGB image. The first image is obtained by turning to white all pixels of colour dissimilar to that of fully unstained nuclei, and then converting to monochrome; the second by turning to white all pixels of colour dissimilar to that of fully stained nuclei. Using the algorithm on these two images we detect heavily stained nuclei. This method is not well suited to deal with textured connected components, as is the case with less stained nuclei (both E- and E+). These are recognised using the third

Table 1. Confusion matrices for some of the experiments.

(a) QSP, W/o ROIs, patch-based

Test	Predicted						
	0	1	2	3	4	5	6
0	54	00	06	00	01	00	00
1	16	00	02	00	00	00	00
2	10	00	10	01	02	00	01
3	04	00	06	00	07	01	00
4	04	00	04	00	04	04	02
5	01	00	03	00	03	00	08
6	00	00	01	00	00	04	16

(b) QSP, With ROIs, segmentation-based

Test	Predicted						
	0	1	2	3	4	5	6
0	60	00	01	00	00	00	00
1	11	00	07	00	00	00	00
2	04	02	13	05	00	00	00
3	01	00	05	05	07	00	00
4	00	00	04	02	09	02	01
5	00	00	00	02	00	04	09
6	00	00	00	00	01	01	19

and the fourth images, that are created using two pseudo-hue functions ($2 \times B - R - G$ and its opposite $-2 \times B + R + G$) and rescaling the results in order to have values in the [0, 255] range. So, in the third image, stained pixels (regardless of their intensity) will have values near to zero, and unstained pixels near to 255, while in the fourth image the opposite holds. In both cases we previously discarded pixels that can safely be classified as background. Figure 1(c) shows the segmentation result for the same region shown previously in Figure 1(a). Most epithelial nuclei are segmented.

To formalise the QSP value we first classify each detected nucleus as either E+ or E-. If there are more than 20% brown pixels, the nucleus is marked as E+, otherwise as E-. Nuclei near the boundary are not marked. Our formalised Quickscore predictor is then the proportion of epithelial nuclei (both E- and E+) that are E+. In turn, the QSS value is first estimated for each individual E+ nucleus, and then for the whole spot from individual values. The score of each nucleus is the intensity of the 20%-quantile darkest pixel, and the whole spot is scored as the 20%-quantile darkest nucleus, in order to approximate what pathologists do when observing these images (i.e. darker pixels and nuclei are given more relevance).

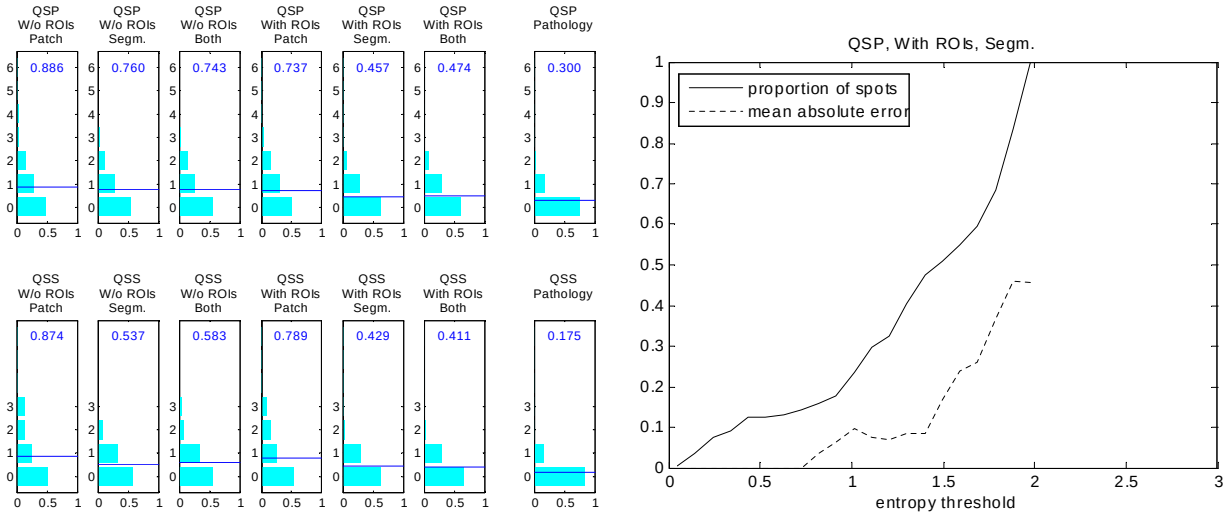
4 Ordinal regression

For the prediction of QSP and QSS values through ordinal regression, we employed the Gaussian process techniques reported by Chu & Ghahramani [13, 14], briefly summarised in the following. Considering a data set composed of *n* samples, where the *i*th sample is a pair of input vector $x_i \in R^d$ and target $y_i \in \{1, 2, ..., r\}$ (without loss of generality). Gaussian processes assume each x_i to be associated with an unobservable latent function $f(x_i) \in R$ (a zero-mean random variable), on which the ordinal variable y_i in turn depends. The process is specified by the covariance matrix for the set of latent functions, and the elements of this matrix can be defined by Mercer kernel functions. In this work, we used a Gaussian kernel. Every Gaussian process has a number of hyper-parameters that need to be optimised. In this work, the maximum *a posteriori* approach with Laplace approximation was used for hyper-parameter learning. The automatic relevance determination method proposed by MacKay [15] and Neal [16] is embedded into the covariance function.

Along with the predicted score for each spot, the ordinal regression algorithm outputs *r* real values that can be interpreted as the posterior probabilities of the possible score values. As discussed later in sections 5 and 6, this type of output proved to be useful. We used the Gaussian process code developed and made publicly available by Chu & Ghahramani [13].

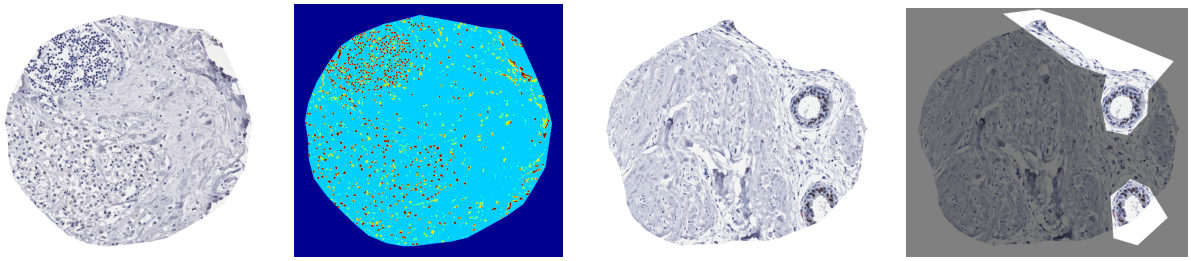
5 Results and discussion

Leave-one-out experiments were carried out to predict the QSP and the QSS values of each of the 175 available spots, using as input features: (i) only the formalised Quickscores based on local patch classification (two real values); (ii) only the formalised Quickscores based on nuclei segmentation (two real values); and (iii) both formalised Quickscores (four real values). All the formalised Quickscores were computed first over the whole spots, and then analysing only the ROIs. Figure 2(a) summarises the results, showing the distributions of absolute errors over the 175 predictions (an individual error being the absolute difference between the predicted score for a given spot and its true score). For each experiment, the mean absolute error is also shown, both numerically and by means of a horizontal line. For comparison, Figure 2(b) shows the distributions of absolute differences between two scoring sessions of the same spots, undertaken by the same pathologist. Table 1 shows example confusion matrices for two of the QSP prediction experiments: without ROIs and local patch-based; and with ROIs and nuclei segmentation-based.



(a) (b) (c)

Figure 2. (a) Distributions of absolute prediction errors for the various experiments carried out; (b) distributions of absolute differences between two scoring sessions; and (c) fraction of processed spots and mean absolute error over those spots, versus confidence threshold (lower entropy means higher confidence), for one of the experiments.



(a) (b) (c) (d)

Figure 3. (a-b) A spot correctly predicted after switching from pixel classification to segmentation, along with the pixel classification results; and (c-d) a spot correctly predicted after its ROI was taken into account, along with a depiction of the ROI.

Figure 3 shows two interesting examples of spots. Spot 3(a) has true QSP and QSS of 0 and 0, respectively; without using ROIs, formalised Quickscores based on local patch classification led to predictions of 4 and 3, respectively; but, keeping the non-use of ROIs and switching to nuclei segmentation, correct predictions were achieved, for both QSP and QSS. The top-left region of this spot contains a large number of inflammatory cells. The patch-based approach visibly classified many of these cells' pixels as E+ pixels, which led to over-estimated formalised QSP and QSS. This misclassification was probably due to the fact that the training set of manually annotated nuclei contains a certain proportion of nuclei that are neither fully non-stained nor fully stained. In turn, even though the segmentation-based approach, too, wrongly detected the inflammatory cells in this spot as epithelial nuclei, they were assigned to the E- class, thus having no harmful effect on the computed formalised Quickscore. Figure 3(b) shows the (largely erroneous) result of pixel classification for the same spot.

In turn, spot 3(c) has true QSP and QSS of 2 and 1, respectively; without using ROIs, formalised Quickscores based on segmentation led to predictions of 0 and 0, respectively; but, using the spot's ROI while keeping the segmentation-based approach, correct predictions were achieved, again for both QSP and QSS. Figure 3(d) shows the same spot with its ROI highlighted. Within this spot, the segmentation-based method misidentified many of the stromal cells in the region of connective tissue (that is, the two left thirds of the spot) as E- nuclei. This led to under-estimated formalised QSP and QSS. In turn, when the manually annotated ROI was used, the stromal cells fell outside the ROI and therefore were not taken into account.

As mentioned previously in section 4, the ordinal regression algorithm outputs, along with each prediction, a posterior probability distribution over the r targets. The entropy of a posterior distribution can be used as a simple measure of ordinal regression confidence (the lower the entropy, the higher the confidence). For the experiment based on ROIs and segmentation, figure 2(c) shows the fraction of test spots that can be predicted below a given entropy threshold. Also

shown is the mean absolute error computed over each fraction of spots. As the entropy threshold set on the predictions was decreased (i.e., as the minimum confidence threshold was increased), the mean absolute error tended to decrease as well.

6 Conclusions

This paper reported the use of ordinal regression to predict Quickscores of breast TMA spots subjected to a form of nuclear staining, comparing the effects of different approaches to the computation of global features, namely the use vs. non-use of nuclear segmentation, and the use vs. non-use of ROIs similar to those involved in the manual scoring process. The combined use of segmentation and ROIs led to a reduction of 0.43 (48%) in the mean absolute error of QSP predictions, and of 0.45 (51%) in QSS predictions. Thus, this work suggests the need to either rely on accurate segmentation of nuclei for the computation of formalised Quickscores features, or improve the quality of patch-based posteriors if segmentation is to be avoided (possibly considering pixels belonging to inflammatory cells as a class in its own right). Furthermore, the need for the development of methods to automatically segment ROIs within spots has been highlighted.

From Figure 3(c), it can be seen that, below an entropy threshold of about 1.75, over 60% of the spots could still be classified with a mean absolute error equal to the mean average intra-observer disagreement of 0.30. This suggests that the reported methods can be used to automatically process, with acceptable error, significant fractions of spots that are more unequivocal, while identifying the more difficult spots that cannot dispense with human assessment.

Acknowledgements: This research was funded by the Breast Cancer Research Trust[1].

References

1. J. Kononen, L. Bubendorf, A. Kallionimeni et al. "Tissue microarrays for high-throughput molecular profiling of tumor specimens." *Nature Medicine* **4(7)**, pp. 844–847, 1998.
2. S. Detre, G. Saccani Jotti & M. Dowsett. "A "quickscore" method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas." *Journal of Clinical Pathology* **48(9)**, pp. 876–878, September 1995.
3. S. Petushi, F. Garcia, M. Haber et al. "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer." *BMC Medical Imaging* **6**, pp. 14, Oct 2006.
4. D. Axelrod, N. Miller, H. Lickley et al. "Effect of Quantitative Nuclear Image Features on Recurrence of Ductal Carcinoma In Situ (DCIS) of the Breast." *Cancer Informatics* **4**, pp. 99–109, 2008.
5. S. Doyle, S. Agner, A. Madabhushi et al. "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features." In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging*, pp. 496–499. IEEE, Paris, France, 2008.
6. J. Chapman, N. Miller, H. Lickley et al. "Ductal carcinoma in situ of the breast (DCIS) with heterogeneity of nuclear grade: prognostic effects of quantitative nuclear assessment." *BMC Cancer* **7**, pp. 174, 2007.
7. J. Dalle, W. Leow, D. Racoceanu et al. "Automatic Breast Cancer Grading of Histopathological Images." In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3052–3055. 2008.
8. J. Zhang, S. Petushi, W. Regli et al. "A study of shape distributions for estimating histologic grade." In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1200–1205. 2008.
9. S. Kostopoulos, D. Cavouras, A. Daskalakis et al. "Colour-Texture based image analysis method for assessing the Hormone Receptors status in Breast tissue sections." In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4985–4988. IEEE, 2007.
10. Adjuvant Breast Cancer Trials Collaborative Group. "Polychemotherapy for early breast cancer: Results from the international adjuvant breast cancer chemotherapy randomized trial." *Journal of the National Cancer Institute* **99(7)**, pp. 506–515, 2007.
11. T. Amaral, S. McKenna, K. Robertson et al. "Classification of breast tissue microarray spots using colour and local invariants." In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging*, pp. 999–1002. IEEE, Paris, France, 2008.
12. M. Sciarabba, G. Serrao & N. A. Borghese. "Automatic neurons identification inside large cortical slices." *Journal of Neuroscience Methods* Submitted.
13. W. Chu & Z. Ghahramani. "Gaussian processes for ordinal regression." *Journal of Machine Learning Research* **6**, pp. 1019–1041, July 2005.
14. T. Amaral, S. McKenna, K. Robertson et al. "Scoring of breast tissue microarray spots through ordinal regression." In *Proceedings of the 4th International Conference on Computer Vision Theory and Applications*, volume 2, pp. 243–248. INSTICC, Lisbon, Portugal, 2009.
15. D. J. C. MacKay. *Bayesian methods for backpropagation networks*, pp. 211–254. Springer-Verlag, New York, 1994.
16. R. Neal. *Bayesian learning for neural networks*. Number 118 in Lecture Notes in Statistics. Springer-Verlag, New York, 1996.

[1] http://www.breastcancerresearchtrust.org.uk/