Automatic Breast Cancer Grading of Histopathological Images

Jean-Romain Dalle, Wee Kheng Leow, Daniel Racoceanu, Adina Eunice Tutac,
Thomas C. Putti

***Abstract*— Breast cancer grading of histopathological images is the standard clinical practice for the diagnosis and prognosis of breast cancer development. In a large hospital, a pathologist typically handles 100 grading cases per day, each consisting of about 2000 image frames. It is, therefore, a very tedious and time-consuming task. This paper proposes a method for automatic computer grading to assist pathologists by providing second opinions and reducing their workload. It combines the three criteria in the Nottingham scoring system using a multi-resolution approach. To our best knowledge, there is no existing work that provide complete grading according to the Nottingham criteria.**

I. Introduction

Breast cancer grading of histopathological images is the standard clinical practice for the diagnosis and prognosis of breast cancer development. Pathologists perform grading manually under a microscope. Their experience directly influence the accuracy of grading. Variability among pathologists have been observed in clinical practice [1]. In a large hospital, a pathologist typically handles 100 grading cases per day, each consisting of about 2000 image frames [2]. It is, therefore, a very tedious and time-consuming task. A computer system that performs automatic grading can assist the pathologists by providing second opinions, reducing their workload, and alerting them to cases that require closer attention, allowing them to focus on diagnosis and prognosis.

Breast cancer grading is performed according to the Nottingham scoring system, which combines three criteria: (1) nuclear pleomorphism, (2) tubular formation, and (3) mitotic count. It is also called the modified Bloom-Richardson system, which is preferred by most pathologists as it gives a more objective grading assessment than previous systems.

Existing methods, except [2] and [3], handle only one of these criteria for histological images. This paper presents a multi-resolution method for breast cancer grading that combines the three criteria.

II. Related Work

Automatic grading of prostate cancer have been proposed [4], [5]. They adopt the Gleason system instead of Nottingham system, and uses different grading criteria.

Jean-Romain Dalle and Wee Kheng Leow are with Dept. of Computer Science, National University of Singapore, Computing 1, Singapore 117590. {jeanroma, leowwk}@comp.nus.edu.sg

Daniel Racoceanu is with IPAL-UMI, CNRS 2955, 21 Heng Mui Keng Terrace, Singapore 119613. danielr@comp.nus.edu.sg

Adina Eunice Tutac is with Engineering Dept., Politehnica University of Timisoara, Bd.V.Parvan Nr.2, Timisoara, Romania. tutac@cs.utt.ro

Thomas C. Putti is with Dept. of Pathology, National University Hospital, 5 Lower Kent Ridge Road, Singapore 119074. pattcp@nus.edu.sg

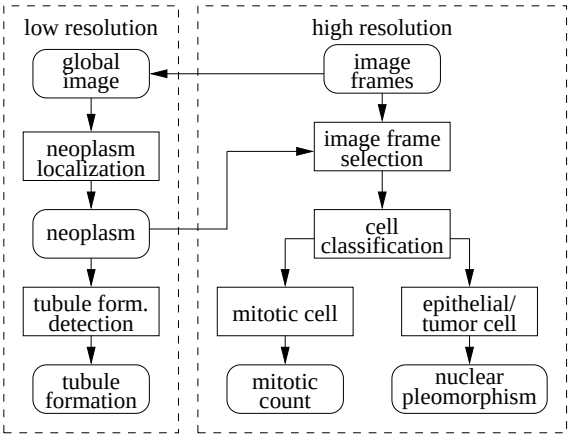


Fig. 1. Overall structure of breast cancer grading system.

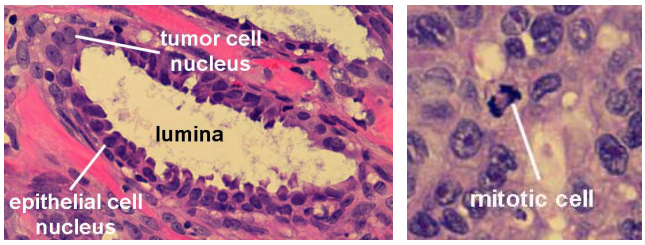
Many techniques have been proposed for the segmentation of cell nuclei in histological images [1], [6], [7], [8], [9], [10], including thresholding, morphological operation, watershed, etc. They are used for nuclear pleomorphism scoring. Petushi et al. [3], [11] proposed a method for tubular formation scoring. They modeled the microstructure present in histological images and used them to segment and classify the cells and assess localized tubular formation. Baak et al. [12], [13] proposed a method for counting mitotic cells in histopathological images. Mitotic cells are segmented using thresholding and region growing techniques. They are then classified according to features such as elongation and roundness before they are counted.

The methods reviewed above handle only one of the three criteria in the Nottingham system. The first attempt at combining all the three criteria is proposed in our previous paper [2]. In this paper, we present an improvement over [2] based on a multi-resolution approach.

III. Breast Cancer Grading System

Our breast cancer grading system adopts a multi-resolution approach (Fig. 1). First, a low-resolution global image of the whole histological slide is reconstructed by scaling down and tiling high-resolution micro-image frames. These frames are captured at $40\times$ magnification from a patient's histological sample stained with H&E marker. Typically, as many as 2000 image frames are generated from a patient's sample.

Next, neoplasm localization and detection of tubular formations are performed on the low-resolution global image. High-resolution image frames that contain the neoplasm are selected. Within these image frames, cells are segmented



(a) (b)

Fig. 2. Sample micro-image frames. (a) Epithelial and tumor cells forming tubular formation. (b) Mitotic cell.

using Gaussian color models. Then, they are classified as epithelial/tumor cells or candidate mitotic cells based on similarity of Gaussian distributions (similar to Eq. 2). The candidate mitotic cells are further classified as mitotic or non-mitotic (Section IIID). Epithelial/tumor cells are used for nuclear pleomorphism scoring, and mitotic cells are detected for scoring of mitotic count. Finally, a global pathological grading is computed from the three criteria of nuclear pleomorphism, tubular formation, and mitotic counts. These stages are described in the following sections.

A. Localization of the Neoplasm

In the input images, cell nuclei appear blue-purple in color due to staining by H&E marker. They can be distinguished from the fat and lumina, which are white, and the stroma, which is red (Fig. 2). Given the low-resolution global image, red regions are segmented from other structures such as the stroma, fat, and lumina by a color thresholding technique based on Otsu's method [14]. Next, morphological opening and closing operations are performed on the segmented regions to fill up small holes and connect nearby cells, yielding compact areas corresponding to neoplasm. High-resolution image frames that contain neoplasm are selected for further analysis.

B. Detection and Scoring of Tubule Formation

Tubule formation is a structure that consists of at least one lumina (i.e., white region) surrounded by tumor cells (Fig. 2(a)) [3]. To detect tubule formation, a morphological closing operation is applied to the segmented neoplasm in the low-resolution global image to connect nearby cells. Next, a morphological filling operation is applied to obtain blob structures. If a blob structure contains fat/lumina regions, then it is regarded as a tubule formation.

Tubule formation is scored as follows. First, a measure R of the amount of tubule formations is computed. as the ratio of the total area occupied by the tubule formations over the total area of the cells in all the image frames. Then, the score S_T of tubule formation is computed as:

$$S_T = \begin{cases} 1 & \text{if } R > 0.75 \\ 2 & \text{if } 0.1 \leq R \leq 0.75 \\ 3 & \text{if } R < 0.1 \end{cases} \quad (1)$$

C. Scoring of Nuclear Pleomorphism

Given the detected epithelial/tumor cells, nuclear pleomorphism is scored as follows. Gaussian functions are constructed from training samples to model the probability distributions of the colors c in the three types of cells used for nuclear pleomorphism scoring. The Gaussian model $G_i(c|\mu_i, \Sigma_i)$ of type i cells is defined in terms of the mean μ and covariance matrix Σ of the colors in type i cells.

For a cell detected in a high-resolution image, the distribution of colors in the cell is also modeled using a Gaussian model $G(c|\mu, \Sigma)$. Then, the differences $d(G, G_i)$ between the color distributions of the detected cell and the three cell types are computed based on χ^2 test:

$$d(G, G_i) = \frac{1}{4} \sum_{c \in N(\mu,\Sigma)} \frac{(G(c|\mu, \Sigma) - G_i(c|\mu_i, \Sigma_i))^2}{m_c} \quad (2)$$

where m_c is the mean of $G(c|\mu, \Sigma)$ and $G_i(c|\mu_i, \Sigma_i)$, and $N(\mu, \Sigma)$ is a neighborhood of color samples within one standard deviation of μ.

Note that there are three standard deviations in a 3D color space, one for each of the three dimensions. These standard deviations are computed from the square roots of the diagonal elements in the covariance matrix Σ. In the current implementation, 3 color samples are considered along each color dimension. Thus, a total of 27 colors are sampled from $N(\mu, \Sigma)$. Finally, the cell is classified as a type i cell if its color distribution G is closest to the Gaussian model G_i.

All the detected cells in the high-resolution images are classified into one of the three types. The proportion p_i of cells in each type is then computed. Based on the proportion, the score S_N of nuclear pleomorphism is computed as:

$$S_N = \begin{cases} 3 & \text{if } p_3 \geq T_3 \\ 2 & \text{if } p_3 < T_3 \text{ and } p_2 > T_2 \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

where T_2 and T_3 are appropriate parameters.

D. Detection and Scoring of Mitotic Cells

Candidate mitotic cells are classified into mitotic and non-mitotic using five kinds of features, namely solidity (i.e., roundness), eccentricity, area, and mean and standard deviation of the intensity in the region. Solidity is measured by fitting a standard polygon to the boundary of a cell region and computing the ratio $S = (A_p - A_c)/A_c$, where A_p is the area (i.e., number of pixels) of the fitted polygon and A_c is the area of the cell region. If the cell region is close to a circular disk, its R will be close to 1, and the cell region is classified as epithelial/tumor cell. Eccentricity E is measured by fitting an ellipse to the boundary of a cell region and computing the ratio $E = D_f/D_M$ where D_f is the distance between the foci of the ellipse and D_M is the length of its major axis. Area A is the number of pixels in a cell region.

Two Gaussian models are constructed from training samples, one for mitotic cells P_M and one for non-mitotic cells P_N. Then, a candidate mitotic cells is classified as follows:

If $P_M > C_0 P_N$, then the cell is a mitotic cell.

P_M and P_N are the probabilities that the candidate is and is not a mitotic cell. C_0 is a weighting factor.

The score for mitotic count is computed as follows. First, a mean count C is computed as the mean mitotic count over all image frames multiplied by a factor of 10. The factor of 10 is included because the Nottingham scoring system for mitotic count is computed as the total count over 10 randomly selected image frames. In our automatic grading system, this randomness is replaced by averaging over all the image frames. Therefore, the factor of 10 is included to be consistent with the Nottingham system. Given the mean count C, the score S_M for mitotic count is computed as:

$$S_M = \begin{cases} 1 & \text{if } C \leq 9 \\ 2 & \text{if } 10 \leq C \leq 19 \\ 3 & \text{if } C \geq 20 \end{cases} \quad (4)$$

E. Overall Grading

The overall grade of the patient is defined in terms of the scores for tubule formation S_T, nuclear pleomorphism S_N, and mitotic cells S_M. Compute G as the sum of S_T, S_N, and S_M. Then, the overall grade is assigned as follows:

- Grade I (low grade): $G = 3, 4, 5$
- Grade II (intermediate grade): $G = 6, 7$
- Grade III (high grade): $G = 8, 9$

IV. Experiments and Discussion

The biopsy samples of six patients were obtained from a local hospital. Each sample was stained with the H&E marker and digitized into a set of histological image frames of $40\times$ magnification and 1024×1280 pixel resolution. There were altogether 16565 image frames.

Figure 3 shows a sample low-resolution global image and a segmented neoplasm. Tubule formations are found within the neoplasm. Figure 4 illustrates detected cells for nuclear pleomorphism scoring. It is noted that not all the cells in the images are detected. In particular, indistinct cells are not detected. Nevertheless, scoring is still reliable as long as enough type-2 and type-3 cells are detected. This is analogous to clinical practice in which a pathologist just performs a quick scan of at most 10 image frames to pick up enough type-2 and type-3 cells to make an assessment.

Table I compares the grading results of the algorithm and a pathologist. It can be seen the system's scores are very close to the pathologist's scores. The system's scores tend to be slightly lower than the pathologist's scores. This could be due to the slightly more stringent criteria taken by the system. Given these encouraging results, we are confident that an automatic grading system can be developed to assist the pathologists by providing second opinions and alerting them to cases that require further attention. Nevertheless, more comprehensive tests are needed to provide better evaluation of the algorithm's performance.

TABLE I

Grading results. S_T: tubule formation score, S_N: nuclear pleomorphism score, S_M: mitotic count score, G: overall grade. P: pathologist, S: system.

patient		S_T	S_N	S_M	G
B15	P	1	2	1	1
	S	1	2	2	1
B10	P	1	2	1	1
	S	1	3	1	1
C10	P	3	3	3	3
	S	2	2	2	2
B11	P	3	3	2	3
	S	2	2	2	2
5075	P	3	2	1	2
	S	2	2	1	1
5042	P	3	3	2	3
	S	2	2	2	2

V. Conclusion

This paper presented a multi-resolution method for automatic breast cancer grading of histopathological images. Tubule formations are detected in the low-resolution global image whereas the individual cells are detected and classified in the high-resolution image frames. They are then scored according to the three criteria of the Nottingham system. Test results show that the overall grades computed by the system match those given by the pathologist very well. Given the encouraging test results, we are confident that an automatic grading system can be developed to assist the pathologists by providing second opinions and alerting them to cases that require further attention.

References

[1] C. Demir and B.Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," Rensselaer Polytechnic Institute, Tech. Rep., 2005.

[2] A. Tutac, D. Racoceanu, T. Putti, W. Xong, W.-K. Leow, and V. Cretu, "Knowledge-guided semantic indexing of breast cancer histopathology images," in *Proc. Int. Conf. on Biomedical Engineering and Informatics*, 2008.

[3] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC Medical Imaging*, vol. 6, no. 14, 2006.

[4] S. Doyle, M. Hwang, M. Feldman, and J. Tomaszeweski, "Automated grading of prostate cancer using architectural and textural image features," in *Proc. of 4th IEEE Int. Symp. on Biomedical Imaging*, 2007, pp. 1284 – 1287.

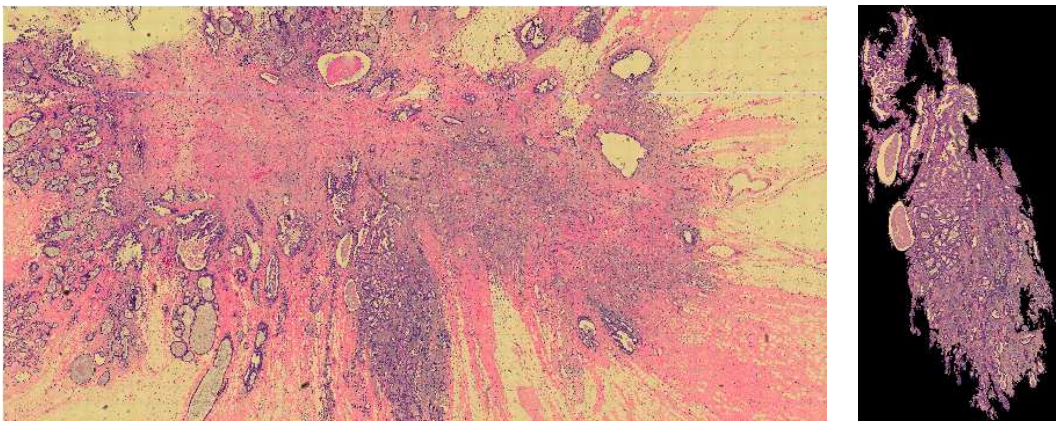
[5] H. Soltanian-Zadeh and K. Jafari-Khouzani, "Multiwavelet grading of prostate pathological images," *IEEE Trans. on Biomedical Engineering*, vol. 50, pp. 697–704, 2003.

[6] A. Nedzved, S. Ablameyko, and I.Pitas, "Morphological segmentation of histology cell images," in *Proc. Int. Conf. Pattern Recognition*, 2000, pp. 1500–1503.

[7] H. Jeong, T.-Y. Kim, H.-G. Hwang, and H.-J. Choi, "Comparison of thresholding methods for breast tumor cell segmentation," in *Proc. of 7th Int. Workshop on Enterprise networking and Computing in Healthcare Industry*, 2005, pp. 392–395.

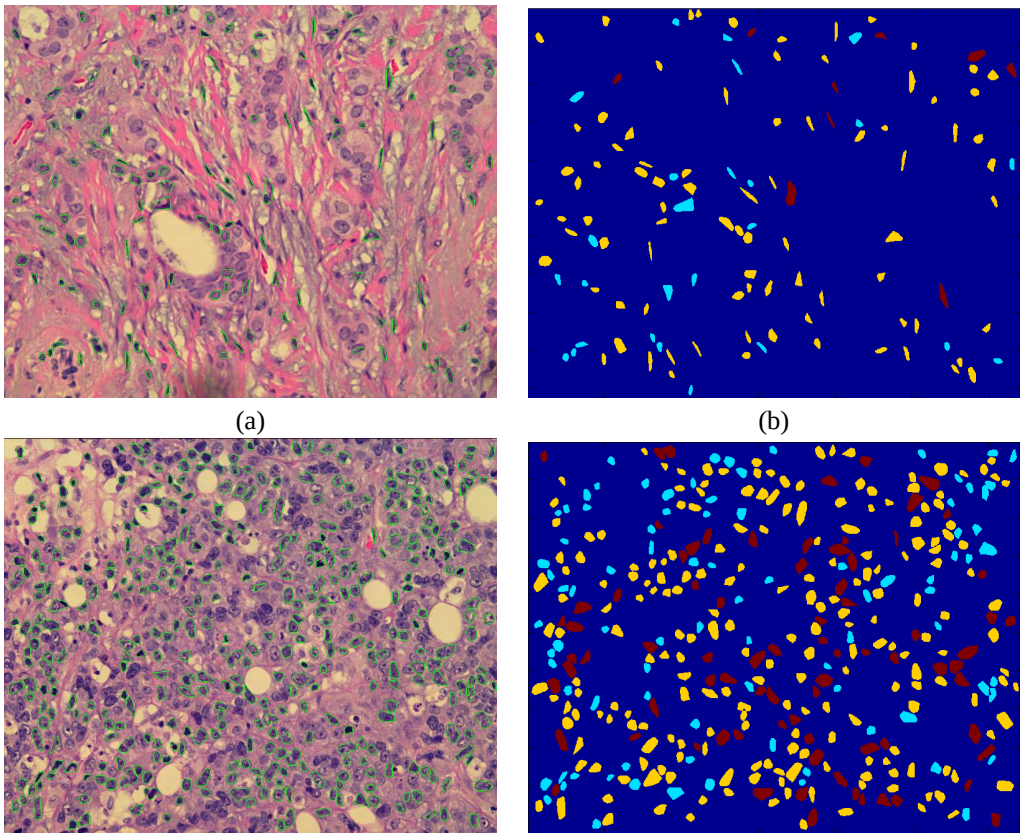
[8] F. Schnorrenberg, "Comparison of manual and computer-aided breast cancer biopsy grading," in *Proc. of IEEE EMBS*, 1996.

[9] M. E. Adawi, Z. Shehab, H. Keshk, and M. E. Shourbagy, "A fast algorithm for segmentation of microscopic cell images," in *Proc. of 4th Int. conf. on Information & Communications Technology*, 2006.



(a) (b)

Fig. 3. Segmentation of neoplasm. (a) Low-resolution global image with neoplasm. (b) Segmented neoplasm.



(a) (b)

(c) (d)

Fig. 4. Sample results for nuclear pleomorphism scoring. (a, c) Detected cells are marked with green boundaries. A score-3 image frame (d) has more type-3 cells than a score-2 image frame. Blue: type 1, yellow: type 2, red: type 3.

[10] A. S. Jadhav, P. Banarjee, K. Chaudhuri, and J. Chatterjee, "Quantitative analysis of histopathological features of precancerous lesion and condition using image processing techniques," in *Proc. of 19th IEEE Int. Symp. on Computer-Based Medical Systems*, 2006, pp. 231–236.

[11] S. Petushi, C. Katsinis, C. Coward, F. Garcia, and A. Tozeren, "Automated identification of microstructures on histology slides," in *Proc. of IEEE Int. Symp. on Biomedical Imaging: Nano to Macro*, vol. 1, 2004, pp. 424–427.

[12] T. ten Kate, J. Belien, and J. Baak, "Counting mitoses by image processing in feulgen stained breast cancer sections: the influence of resolution," *Cytometry*, vol. 28, pp. 135–40, 1997.

[13] ——, "Method for counting mitoses by image processing in feulgen stained breast cancer sections," *Cytometry*, vol. 14, pp. 241–250, 1993.

[14] O. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. Systems, Men, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.