A Grading Strategy for Nuclear Pleomorphism in Histopathological Breast Cancer Images Using a Bag of Features (BOF)

Ricardo Moncayo, David Romo-Bucheli, and Eduardo Romero(✉)

CIM@LAB, Universidad Nacional de Colombia,
Carrera 45 No 26-85, Bogotá, Colombia
{ramoncayomar,deromob,edromero}@unal.edu.co
http://cimlaboratory.com/

Abstract. Nuclear pleomorphism is an early breast cancer (BCa) indicator that assesses any nuclear size, shape or chromatin appearance variations. Research involving the ranking by several experts shows that kappa coefficient ranges from 0.3(low) to 0.5 (moderate)[12]. In this work, an automatic grading approach for nuclear pleomorphism is proposed. First, a large nuclei sample is characterized by a multi-scale descriptor that is then assigned to the most similar atom of a previously learned dictionary. An occurrence histogram represents then any Field of View (FoV) in terms of the occurrence of the descriptors with respect to the learned atoms of the dictionary. Finally, a SVM classifier assigns a full pleomorphism grading, between 1 and 3, using the previous histogram. The strategy was evaluated extracting 134 FoV (×20), graded by a pathologist, from 14 BCa slides of 'The Cancer Genome Atlas' (TCGA) database.The obtained precision and recall measures were 0.67 and 0.67.

Keywords: Breast cancer · Histopathology · Biomedical · Nuclear pleomorphism

1 Introduction

Worldwide breast cancer(Bca) is a major cause of women death. In 2012, approximately 522.000 deaths and 1.677.000 new Bca cases were reported [5]. Bca is frequently diagnosed after a suspicious breast mass is found in radiologic studies, by extracting a tissue sample using a fine needle. This sample is analyzed and information about the type of tumor, aggressiveness and receptor status is obtained. The cancer aggressiveness is determined by using one of the available scoring systems. The World Health Organization and the College of American Pathologists endorse the Notthingham grading system. This system assigns and correlates the scores of three features, namely nuclear pleomorphism, mitotic count and tubule formation [4]. This classification provides prognostic and diagnostic information. Both, the biological variability and the pathologist expertise determine the accuracy and reliability of the evaluation performed when assigning the

A. Pardo and J. Kittler (Eds.): CIARP 2015, LNCS 9423, pp. 75–82, 2015.
DOI: 10.1007/978-3-319-25751-8_10

three aforementioned features. There exist many studies reporting low or moderated grading reproducibility: mitotic counts with κ from 0.45 to 0.64 or tubule formation with κ between 0.57 and 0.83. In nuclear pleomorphism, the index agreement (κ) has been reported to be between 0.3 (low) and 0.5($medium$), a very low figure for an important feature in terms of the prognosis [12]. Nuclear pleomorphism serves as an indicator of the cancer evolution and is part of the Nottingham prognostic index [7,8]. Some authors suggest that nuclear morphometry may improve the grading task, but manual quantification is time consuming and impractical in routine diagnostic workflow [14,17]. The expert made a quantitative and qualitative judgment with the epithelial cells, features to be evaluated are: size of nuclei, size of nucleoli, density of chromatin, thickness of nuclear membrane, regularity of nuclear contour, anisonucleosis, then a nuclear pleomorphism grading is give, score ranges from 1 to 3 , in the figure 1 examples the histological images of nuclear pleomorphism are show.

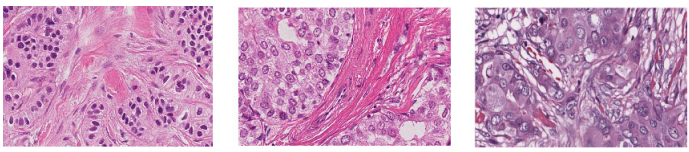

Fig. 1. Nuclear pleomorphism scores. **Left:** Nuclear atypia grade 1. **Center:**, grade 2, **Right:** grade 3

In this paper, an automatic methodology for nuclear pleomorphism grading of breast cancer images is proposed. Unlike other approaches, the proposed method is not based on histomorphometry information such as area, roundness, or texture. Instead, a nuclei detector, a bag of features (BoF) and a multi-scale descriptor are used to characterize the differences among the three nuclear atypia grades. The method is able to learn from a whole FoV, not requiring individual cell annotations. Finally, the BoF representation is used to train a bank of support vector machines (SVM) that assign a pleomorphism grade to a test FoV.

This paper is organized as follows: In Section 2, a brief review of techniques for nuclear atypia (pleomorphism) characterization is presented. Section 3 describes the proposed method. Results of the proposed methodology in the nuclear atypia scoring are presented in Section 4. Finally, we present conclusions and future work on the characterization and feature extraction of nuclear pleomorphism in Section 5.

2 Previous Work

Several algorithms have been proposed to assess nuclear pleomorphism in breast cancer. Automatic nuclear shape and size related measurements are commonly used. Cossato et al. [2] proposed a pleomorphism grading method that starts by

detecting nuclei candidates (related to hematoxylin dye) using a color deconvolution algorithm. Then a 2D difference of gaussians (DoG) is applied to improve nuclei detection. Each nucleus is segmented by applying an active contour algorithm and outliers are discarded by using some statistical measurements. Morphometric features (such as nuclear area, shape, and texture) are computed for each segmentation and then used as an input to a SVM classifier to grading each nucleus between "benign" and "malignant" classes, obtaining a classification performance of 81%. Dalle et al.[3] proposed a similar approach for nuclear pleomorphism grading, but measurements of nucleus roundness(perimeter and area) and texture were included as features. After computing these features, a multivariate gaussian model is estimated for each grading score (only atypia grades 2 and 3 were used). Test images are graded by computing the likelihood of the nuclei grade at a particular magnification frame. After classifying each nucleus, an overall reported accuracy error of 7.8% for the classification task. Veta et al.[18] propose a nuclei segmentation for nuclear pleomorphism method, using a color unmixing process and only the hematoxylin channel was used. Afterward, the nuclear contours are found by a fast radial symmetry transform, followed by a post-processing method to remove regions with not nuclei. Finally the overlapping regions are eliminated. The sensitivity reported by the method was 0.853.

3 Methodology

An overview of the proposed method is presented in figure 2

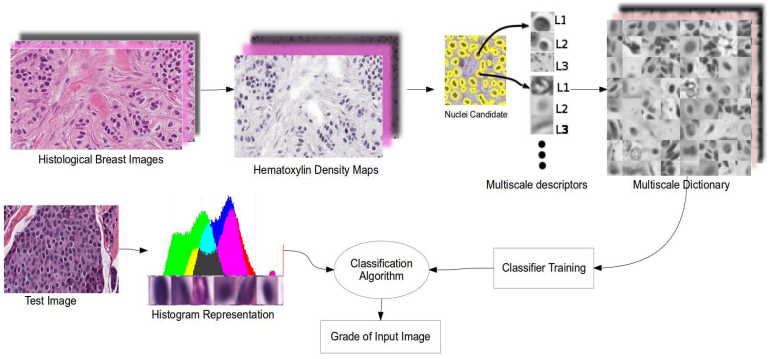

Fig. 2. Diagram of the proposed approach

The first step to detect nuclei candidates is the colour deconvolution in the H&E image[10]. Then, maximally stable extreme regions (MSER) are used to identify blobs[11]. Morphological operations, namely opening and closing, are applied to improve the nuclei detection. A discriminant multi-scale histopathology descriptor [15], centered at each nucleus candidate, is then computed.

The set of selected nuclei descriptors constitutes a dictionary of visual words comprised of samples at different pathological evolution (grading range from 1 to 3 in the nuclear pleomorphism nothingam grading system). The dictionary of visual words is a partition of the space, spanned by the set of collected descriptors or dictionary atoms. Any point of that space may then be represented by a frequential distribution of the basis defined by the partition. The resultant histograms are used to train a bank of binary classifiers or support vector machines (SVM), using the annotated labels associated to the pathological grade of each atom in the dictionary. A new FoV is scored by identifying nuclei candidates that are represented in terms of the dictionary of multiscale nuclei descriptors.

3.1 Candidate Detection Using MSER Descriptors

First, the hematoxylin stain is estimated by using a color deconvolution approach [10]. The process starts by mapping the usual RGB color image to the Optical density space (O_d) using equation 1

$$O_d = -log(I); \qquad O_d = VS \rightarrow S = V^{-1}O_d \tag{1}$$

Where I is the RGB color space, O_d is the computed optical density and V, S are matrices of the stain vectors and their saturation, respectively. In the optical density space, the hematoxylin and eosin estimations are modeled by a linear generative model, where two sources are computed.

The hematoxylin contribution map is used in the nuclei detection and the subsequent processing. Nuclei candidates are found by detecting Maximally Stable Extreme Regions (MSER) [11] defined as those areas that change very little after applying multiple thresholds. Finally two morphological operators are applied, opening to remove small objects and closing to fill small holes and gaps in the image.

3.2 Multi-scale Feature Extraction

The characterization of each nucleus candidate was performed by analyzing multiple scales of a pyramidal representation [15]. A nucleus is simply represented by a series of different scales around the nucleus center, i.e., given a point X, a nucleus center, a window of a given size is extracted. In this work a windows size of 16×16 pixels was selected. Afterwards, patches with the same size at lower resolution are obtained by dyadically subsampling the original image after a blurring operation between each scale. These patches are then concatenated for obtaining the feature vector, as depicted in Figure 3.

After collecting a large number of descriptors (equal to the number of nuclei detected in the training set), the space is partitioned with a k-means algorithm [9] using the k-atoms of the dictionary (k was set to 1600 visual words). This dictionary is created with random samples for each grade. The dictionary is thus used to represent each nucleus from any breast cancer at a high magnification

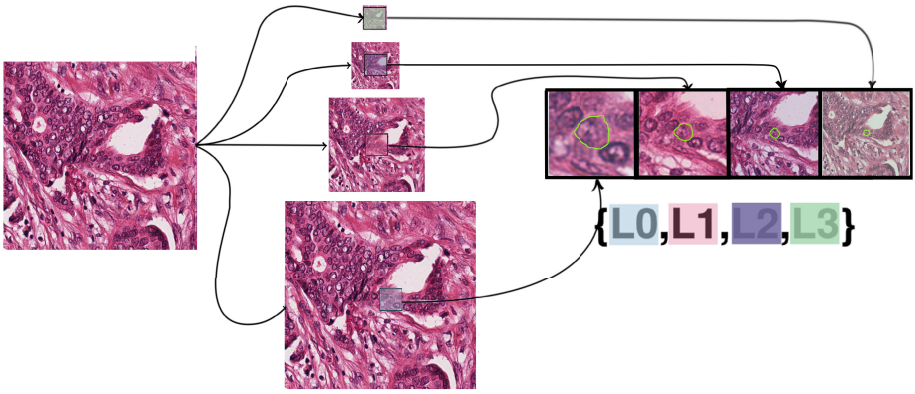

Fig. 3. Multi-Scale Feature Extraction

(× 40), by choosing the most similar atom to a particular nucleus candidate. Then, a feature vector is built by computing a histogram of occurrences, for which each bin stands for each atom of the dictionary. A SVM classifier (with a chi square kernel) is then trained to assign a nuclear pleomorphism grade for a new FoV. SVM classifier with chi square kernel has been extensively used on BoF applications [16].

4 Experimental Setup

4.1 Breast Cancer Dataset

The dataset is comnposed of fourteen(14) breast cancer slides from the TCGA database [13]. The TCGA provides clinical information and the cases were evaluated by different experts since images are provided by 27 different institutions. The associated Nottingham grading to each evaluation is provided (tubule formation, mitotic count, nuclear pleomorphism). Additionally, for each of the database samples, high magnification FoV (× 20) were selected and graded by an independent expert pathologist. A total of 134 high magnification frames (around 10 FoV per slide) were digitized and the associated nuclear pleomorphism grading was recorded. The Figure 4 shows an example of the dataset: a whole breast cancer Virtual Slide, the manually selected high magnification FoV and the grading assigned by the pathologist.

4.2 Evaluation and Dictionary Setup

The evaluation was carried out using a leave one out scheme, i.e., for a total of 14 cases, 13 were used for training and the remaining one for testing. In this experimentation, the dictionary was initialized by randomly sampling 18 FoV per each grade. After the learning process, the dictionary contains 1600 atoms

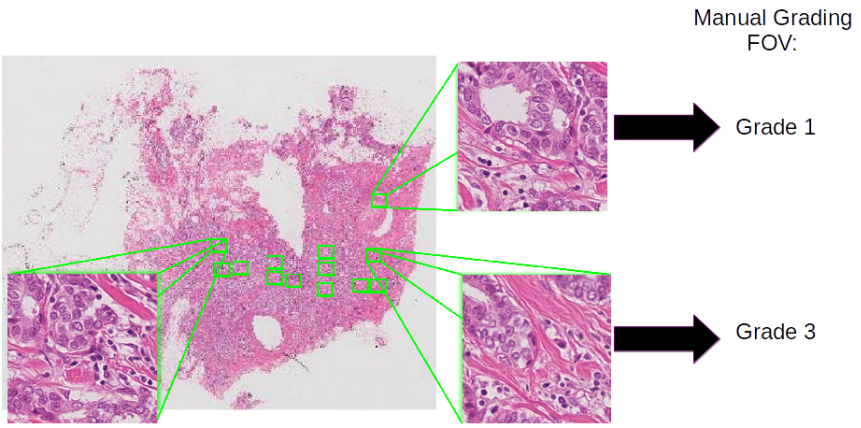

Fig. 4. Breast Cancer Dataset

with 768 dimensions (size patch × scales : 16 × 16 × 3). The SVM classifier was trained by using the multi-scale features extracted from 90 random training images (30 for each grade).

5 Results

The nuclei candidates were processed by extracting three (3) scale levels on a patch size of 16×16 pixels. Only the hematoxylin contribution map was used. A set of dictionaries for descriptors of each grade is built and then concatenated. Afterwards a bank of three (3) SVM classifiers was trained. The evaluation was performed as described in section 3.2. The precision, recall and F-measure results with standard deviation (SD) are shown in table 1.

Table 1. Results of Proposed Method. Mean ± standard deviation figures for the precision, recall and f-measure of the nuclear atypia grades are presented.

	grade 1	*grade 2*	*grade 3*	*Mean*
Precision	0.74 ± 0.04	0.58 ± 0.06	0.69 ± 0.03	0.67 ± 0.03
Recall	0.55 ± 0.09	0.62 ± 0.03	0.85 ± 0.04	0.67 ± 0.04
F-measure	0.63 ± 0.06	0.60 ± 0.04	0.76 ± 0.02	0.66 ± 0.04

Atypical grade three (3) obtained larger F-measures. The nuclei related with these grading shows a significant deviation from the shape and size of grade two

(2) and one(1), that is adequately assessed by the proposed method, the intermediate class grade two(2) have the low f-measure.

The Cohen's kappa coefficient (agreement between 2 rater) [1] and the Fleiss' kappa coefficient (multiple rater)[6] were also used to obtain a statistical measure of the agreement between a pathologist, TCGA record diagnosis and the proposed method. The Cohen's kappa coefficient between the TCGA's reported grading and our pathologist was $\kappa = \mathbf{0.43}$(moderate agreement). The kappa coefficient between the proposed method and TCGA diagnosis was $\kappa = \mathbf{0.51}$ (moderate agreement) while the agreement of the proposed method with our pathologist corresponds to $\kappa = \mathbf{0.31}$(fair agreement). Finally, the Fleiss' Kappa was used to establish an agreement among the three raters: the pathologist, the TCGA record and the proposed method. The obtained value was $\kappa = \mathbf{0.46}$ which amounts to a moderate agreement.

6 Conclusion

In this work an automatic method for a complete grading of nuclear pleomorphism in breast cancer images, in accordance to the Nothingham grading system, was proposed. The method does not use morphometric information from manually segmented cells, but builds up a visual dictionary of nuclei that implicitly captures differences in nuclei size and shape, among the different nuclear atypia grades. The descriptor also includes the neighborhood of the nuclei at subsequent lower scales, including context information of the nuclei that is known to be important. Performance measures show that the method is suitable for automatic pleomorphism grading of microscopical FoV.

Future work includes the improvement of candidate nuclei detection and the use of nuclei density information to improve the grading task. Another important topic is the exploration of this representation in the analysis of another potential or recognized diagnostic/prognostic indicators.

References

1. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement **20**(1), 37 (1960)
2. Cosatto, E., Miller, M., Graf, H., Meyer, J.: Grading nuclear pleomorphism on histological micrographs, pp. 1–4, December 2008
3. Dalle, J.R., Li, H., Huang, C.H., Leow, W.K., Racoceanu, D., Putti, T.C.: Nuclear pleomorphism scoring by selective cell nuclei detection. In: IEEE Workshop on Applications of Computer Vision Snowbird, Utah, USA (2009)
4. Elston, C.W., Ellis, I.O.: Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991; 19; 403–410. Histopathology 41(3a), 151 (2002)

5. Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D., Forman, D., Bray, F.: GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 (2012). http://globocan.iarc.fr
6. Fleiss, J.: Measuring nominal scale agreement among many raters. Psychological Bulletin **76**(5), 378–382 (1971)
7. Galea, M., Blamey, R., Elston, C., Ellis, I.: The nottingham prognostic index in primary breast cancer. Breast Cancer Research and Treatment **22**(3), 207–219 (1992). http://dx.doi.org/10.1007/BF01840834
8. Green, A.R., Soria, D., Stephen, J., Powe, D.G., Nolan, C.C., Kunkler, I., Thomas, J., Kerr, G., Jack, W., Camreron, D., Piper, T., Ball, G.R., Garibaldi, J.M., Rakha, E.A., Bartlett, J.M., Ellis, I.O.: Nottingham prognostic index plus (npi+): Validation of the modern clinical decision making tool in breast cancer. Cancer Research **75**(9), P5–09 (2015)
9. Hartigan, J.A., Wong, M.A.: Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) **28**(1), 100–108 (1979). http://www.jstor.org/stable/2346830
10. Macenko, M., Niethammer, M., Marron, J., Borland, D., Woosley, J., Guan, X., Schmitt, C., Thomas, N.: A method for normalizing histology slides for quantitative analysis. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009. ISBI 2009, pp. 1107–1110, June 2009
11. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Proceedings of the British Machine Vision Conference, pp. 36.1–36.10 (2002) doi:10.5244/C.16.36
12. Meyer, J.S., Alvarez, C., Milikowski, C., Olson, N., Russo, I., Russo, J., Glass, A., Zehnbauer, B.A., Lister, K., Parwaresch, R.: Breast carcinoma malignancy grading by Bloom-Richardson system vs proliferation index reproducibility of grade and advantages of proliferation index. Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc **18**(8), 1067–1078 (2005)
13. National Cancer Institute: The Cancer Genoma Atlas. http://cancergenome.nih.gov/
14. Prvulović, I., Kardum-Skelin, I., Sustercić, D., Jakić-Razumović, J., Manojlović, S.: Morphometry of tumor cells in different grades and types of breast cancer. Collegium antropologicum **34**(1), 99–103 (2010)
15. Romo, D., Garcez, P., Romero, E.: A discriminant multi-scale histopathology descriptor using dictionary learning. In: Proc. SPIE 9041, 90410Q–90410Q-6 (2014)
16. Vedaldi, A., Zisserman, A.: Efficient additive kernels via explicit feature maps. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(3), 480–492 (2012)
17. Veta, M., Kornegoor, R., Huisman, A., Verschuur-Maes, A.H.J., Viergever, M.A., Pluim, J.P.W., van Diest, P.J.: Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **25**(12), 1559–15565 (2012)
18. Veta, M., Van Diest, P.J., Kornegoor, R., Huisman, A., Viergever, M.A., Pluim, J.P.W.: Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. PLoS ONE **8**(7), e70221 (2013)