A Robust Computer Aided Approach for Scoring and Quantitative Evaluation of Whole Slide Digital Images of HER2 Immunohistochemistry in Breast Cancer

Fariba D. Khameneh*, Ercan Alp Serteli§, Salar Razavi†
*Informatics Institute, Istanbul Technical University, Turkey
†Department of Electronics and Communication, Istanbul Technical University, Turkey
§Virasoft, ARI 1 Teknokent, Istanbul, Turkey

***Abstract*—Human epidermal growth factor receptor 2 (HER2) is recognized as an essential prognostic parameter in breast cancer. In a traditional way, the key part in predicting invasive breast cancer was evaluating HER2 amplification analyzed by visual experiments. However, conventionally visual scoring of HER2 (into 0/1+, 2+ or 3+) is subjective to the observers. An objective method is required to satisfy pathologists' demand to confidently evaluate IHC. Digital pathology has led to improvement in diagnosis of cancer tissues. This article describes integration of a machine learning approach and morphological method for detecting region of interest and scoring the Immune histochemical images in whole slide images. Our method gets benefits of feature engineering, unsupervised learning and supervised learning with support vector machine (SVM) and also using morphological operators. Here, the application of simple linear iterative clustering (SLIC) superpixel for segmentation of the whole slide image is evaluated. Some texture and color features are extracted from the superpixels to be classified by support vector machine (SVM). The result of classification method is utilized in breast cancer epithelium detection from Stroma parts. In order to improve the classification of the cases, we presented an automated approach for scoring HER2 in whole slide images (WSI). This part of the algorithm is based on morphological techniques which discriminate between amplified and non-amplified cases with high accuracy. Experimental results on real datasets were compared with the pathologist visual scores that demonstrate 92% agreements and 0.867 kappa value which confirm the wide applicability and high accuracy of our approach in validating of IHC analysis in WSI.**

***Index Terms*—Digital pathology, HER2, Immunohistology, Computer-aided diagnosis, Whole-slide imaging, SVM, Tissue segmentation.**

I. Introduction

Breast cancer is the second most common form of cancer among women in the US that leads to death [1]. Identification of biomarkers in tissues carry significant biological information. Evaluating the expression level in some biomarkers play an essential role in cancer diagnosis. Visual examination is one of the traditional methods which is considered veracious in some areas, however it depends on many factors which makes this process untrustworthy. Furthermore, subjectively scoring of histological biomarkers is not comprehensive [2].

Digital pathology proposes an appreciable alternative way to prevail the non-objectivity by analysing the biological images. In particular, activity measurements of special proteins through the analysis of tissue images afford critical information in identification of such biological slides in the cancer diagnosis. Through digital slide scanners, significant attention for validation of IHC analysis in WSI is achieved. Digital analysis of WSI helps improve the accuracy and reliability that enables precise diagnosis in image analysis of digital pathology. IHC analysis is a method for demonstrating the presence and location of proteins in tissue sections which is placing new demands on the reproducibility, accuracy, and specificity of the extracted information [3]. The automated analysis in WSI has recently achieved considerable attention because of the accessibility of digital slide scanners and the increasing importance of tissue-based biomarkers of stratified medicine [4].

Several biomarkers have been identified for breast cancer [5]. The most prevalent breast cancer types are commonly recognized using HER2-positive. The HER2 biomarker is over-expressed, amplified, or both, in 15%-20% of high-grade invasive breast cancers [6] and has been associated with fast tumor growth, increased risk of recurrence after surgery, and poor response to shortened survival [7].

The evaluation of HER2 with IHC involves the visual examination of cell membrane staining in categories of {0,1+,2+,3+} corresponding to no, barely or weak, moderate, and strong staining respectively. According to the ASCO/-CAP guidelines, cases scored as (3+) are recommended for trastuzumab therapy, whereas (2+) ones are subject to further testing with FISH [8]. If cells with complete, intense, and uniform staining are more than 10%, this is considered as positive (3+) cases of invasive tumor. In addition, if less than 10% of cells with complete membrane staining are either weak or non-uniform in intensity, they are considered as equivocal cases, and the cells showing no staining or weak, incomplete membrane staining in any proportion of tumor cells are determined as a negative result (0,1+) case [9]. In general, evaluation criteria such as intensity and uniformity of staining, and estimation of the stained cell percentage is a subjective

process that affects the accuracy of IHC assessment. There is clearly a demand for quantitative methods to improve the accuracy and reproducibility of scoring HER2 proteins. As a FISH test is an expensive and time-consuming test, some studies are conducted to reduce the number of equivocal (2+) cases. The recent studies report a wide range of automated methods for HER2 scoring algorithms. The ImmunoMembrane application is an open source software for digital image analysis of HER2 IHC. ImmunoMembrane analysis the completeness and intensity of the cell membrane staining reaction, based on the IHC interpretation criteria of the ASCO/CAP guidelines [9] [10]. Fernandez et al. [11] proposed a method to deal with the color variation and distortion problem by describing a density tool that has been implemented to measure the positive IHC stain areas in WSI. In order to annotate and arrange histopathology images, we presented an automated approach for scoring HER2 in WSI which is based on morphological techniques and recognizes each case with high accuracy to decrease the number of equivocal cases. These abilities have increased the accuracy, reproducibility, and reliability of HER2 assessment.

The major contribution of this paper is to provide a fully automated and flexible procedure that overcomes the limitations of the existing computer-based techniques. In order to evaluate the scoring performance of the proposed technique, following experiments are carried out using tissue specimens from patients with breast cancer from the tumor bank collection provided by a local hospital.

II. Materials and Methods

The dataset consists of 100 WSI in Nano-zoomer Digital Pathology (NDPI) and multi-file with very complicated proprietary metadata and indexes (MRXS) formats. The proposed methodology consists of different steps that are discussed in the following sections.

A. Tissue Segmentation

Regions of cancer are typically obtained from the epithelium parts so separating epithelium and stroma leads to get better statics through analysis. In [12] they have presented the machine learning based approach for detecting metastatic tissue regions accomplishes in blockwise detection of breast cancer metastases from lymph node tissue sections. It was applied for hemotexolyin and eosin whole slide images.

1) Image Segmentaion into Superpixels: The superpixels are formed by grouping pixels based on colour similarity and spatial distance. In [13] they have applied the Simple Linear Iterative Clustering (SLIC) [14] to decrease the complexity of large histological images. We utilized SLIC for generating superpixels due to its strong performance and better adherence to the boundaries.

2) Manual Sampling: Training superpixels were selected by a pathologist as normal and tumor tissues. These negative and positive samples were given as a training dataset to the classification model. Sufficient amount of samples should be selected from different kinds of regions for each category in order to train a robust model.

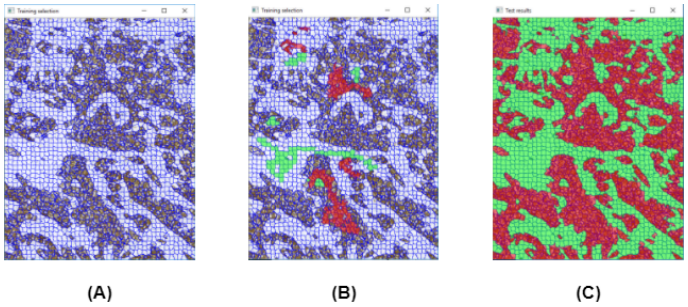

Fig. 1. (A) Image generated by SLIC algorithm. (B) Labelling superpixels as tumor and normal tissues. (C) Result of the SVM classification.

3) Feature Extraction: Texture and colour are considered as the main features of any image. The texture of each tissue was further described using local binary patterns (LBP) [15] and Gray level co-occurrence matrix (GLCM) [16]. In addition to the texture features, mean colour and colour histogram features which consider each channel are used in two colour spaces namely RGB and HSV. The properties of each superpixel are described by 539 features. Normalization is applied in order to eliminate biases created by the range differences between each feature type.

4) Super Vector Machine Model: Classification performance of different models has been compared based on the extracted features. The results produced by the different models using the same training samples are shown in FIG.

B. Morphological Analysis

1) Tile Extraction: One of the challenges in digital pathology is loading excessively large image slides into memory. For example, a MRXS file can have 200000 $\times$ 90000 pixels. An apparent way of handling the WSI is splitting into smaller tiles. A region of interest (ROI) should be selected to apply the method on that area. Since the images have a very high resolution, loading such images are extremely difficult for an application. To overcome the memory usage problem in WSI, the best solution is to keep the images as small as possible and process images in parallel. None overlapping tiles has been extracted as 640 $\times$ 480 images in 0.23 µ / pixels (25$\times$ magnification) to be examined separately as shown in Fig.1. Images from different microscopes with diverse formats have various zoom levels so working on µ / pixels gives us reliable and constant proximity to slides. In general, each slide is divided into about 7k tiles.

2) Preprocessing: Fig.2 shows the workflow of the proposed method. Firstly, the image is separated into hematoxylin (H-Blue) and diaminobenzidine (DAB-Brown) color channels. Cell nuclei and membranes are separated using the color deconvolution algorithm. This algorithm separates the image into three channels, corresponding to the actual colors of the stains used. This allows the pathologist to accurately measure the area for each stain separately, even when the stains are overlying at the same location.

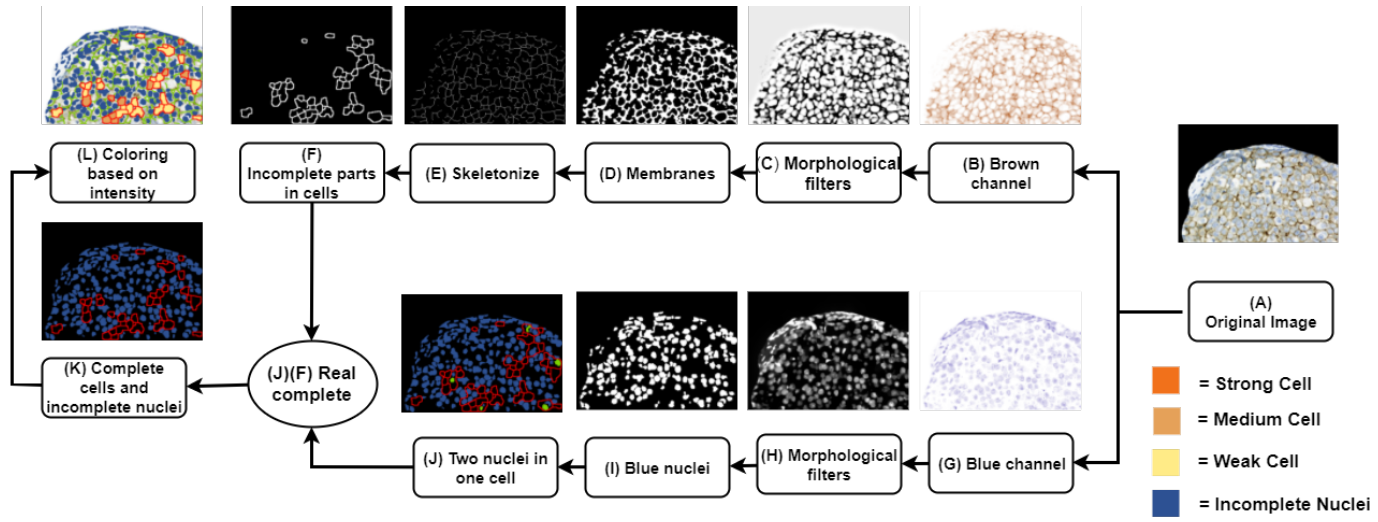

Fig. 2. A flowchart depicts the proposed method after tissue segmentation processes. (A) Original Image which has got from the SVM classification result. (B)(G) Applying color deconvolution for separating diaminobenzidine (DAB) and hematoxylin (H-Blue) components. (C)(H) Using some morphological filters like median, Gaussian, and bilateral filter. (D)(I) Adaptive threshold method is used for separating foreground and background. (E) Implementing skeletonize to preserve connectivity and throw away most of the foreground component. (F) Complete membrane structures include both real complete cells and incomplete cells that are next to each other. If there are incomplete branches inside, it is detected to be multiple incomplete cells. (J) Complete cells with two or more nuclei inside are excluded as joined incomplete cells. (K) Real complete cells have been detected by applying (J) and (F) conditions. (L) Coloring the cells based on intensity. Finally, the result of each tile is aggregated to calculate the total number of complete and incomplete cells in WSI.

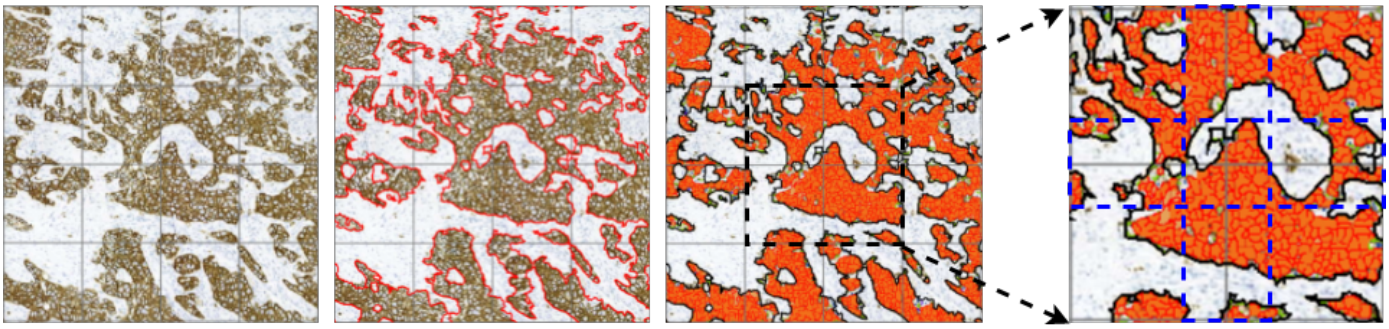

Fig. 3. The proposed merging method by considering neighboring tiles. Four tiles are shown in (D). Each tile considers its right and below edge tiles. (A) The original image. (B) Result after applying Tissue segmentation. (C) Applying merging method by considering neiboring tiles. (D) Four tiles are shown demonstrating the connectivity of the complete and incomplete membrane staining in the edge parts of each tile.

To improve image quality, some preprocessing steps are implemented. For image enhancement, a median filter and unsharp mask are used. The Gaussian filter is also applied to smooth the image for each channel separately.

The usual way of segmenting the image into different regions is thresholding. Adaptive threshold calculates a threshold for each pixel considering its local neighborhood, whereas global thresholding calculates a single value for all pixels. Therefore, adaptive threshold changes the threshold dynamically over the image. This thresholding method becomes more effective in finding more details. In this paper, we utilize local adaptive thresholding for binarization in order to preserve as much information as possible. Furthermore, for finding the membranes each component of the DAB channel is skeletonized to preserve the connectivity of original region and throw away most of the foreground pixels.

There are also some conditions that should be considered. If a complete invasive tumor cell contains incomplete and weak circumferential membrane staining or consists of two or more blue cells, the complete cells would be accepted as multiple incomplete membrane stainings.

3) Merging Tiles: After processing, all tiles should be merged to get the total result. In order to merge tiles, parts from the edges of each tile are extracted and combined together. During a horizontal merge, the left-most part of the right image is combined with the right-most part of the left image. Similarly, during a vertical merge, the bottom-most part of the top image is combined with the top-most part of the bottom image. Extracted parts are wide enough to encompass the largest possible cell size so that no cell will

be cut in half. The algorithm is run in this newly created region, but only the particles (complete cells and negative nuclei) that intersect with the center line are considered for calculations and overlays. The center line in the combined image corresponds to the merging edges of the original tiles. While processing each tile, particles that are found to intersect the edges are excluded from the calculations, thus making sure that no particle will be considered twice. Each will be considered only once, either in the original tiles or in one of the generated merging tiles. In Fig.3, complete membrane stainings intersecting with borders can be observed. Fig.3 (C)(D), show the merging results that consider the edge parts of right and bottom tiles.

III. Results

The detailed results correspond to an almost perfect agreement between visual IHC and automated image analysis which allows us to simplify the comparison as indicated in TABLE 1. Among the 300 ROI, the 140 (46%) slides were evaluated as IHC negative (score 0/1+) by the proposed method, however, 127 (42%) slides evaluated as IHC negative by visual IHC whereas all FISH tests were negative. Among 50 equivocal cases which are scored 2+ by visual IHC, 24 of them are categorized as 0/1+ and 3+ by the proposed method. In 26 (8%) cases both approaches agreed in equivocal cases (score 2+). All the 123 (41%) cases which considered as 3+ cases by the pathologist are also positive in our method and the FISH results were also positive. Furthermore, 11 (3%) slides with positive FISH scores are counted as positive cases which are evaluated as equivocal cases in pathologist visual scoring. By the presented method only 26 slides should be further analyzed by the FISH test. Cohon's kappa is 0.86, that shows a good agreement between visual IHC and the proposed method.

TABLE 1 Comparison of the proposed method and visual IHC assessment

		Visual IHC			
		0/+1	2+	3+	Total
Proposed method	0/1+	127	13	0	140
	2+	0	26	0	26
	3+	0	11	123	134
	Total	127	50	123	300
Agreement					92%
kappa					0.867

IV. Conclution

This paper presents an automated method for HER2 segmentation which is applied on WSI and it provides better measurements of HER2 biomarker than the ones that were ordinarily executed by visual inspection over the past years. Our method overwhelms this error-prone and time-consuming process. SVM classification is implemented for tissue segmentation. Furthermore, the WSI's result would be more satisfactory since the slides consist of different color distributions. The merging technique is proposed to identify edge parts of tiles according to its neighboring. In particular, our technique overcomes the limitations of traditional segmentation techniques based on local or spatial intensity information.

Acknowledgment

References

[1] Robert A Smith, Kimberly S Andrews, Durado Brooks, Stacey A Fedewa, Deana Manassaram-Baptiste, Debbie Saslow, Otis W Brawley, and Richard C Wender. Cancer screening in the united states, 2017: a review of current american cancer society guidelines and current issues in cancer screening. *CA: a cancer journal for clinicians*, 67(2):100–121, 2017.

[2] Magali Lacroix-Triki, Simone Mathoulin-Pelissier, Jean-Pierre Ghnassia, Gaetan Macgrogan, Anne Vincent-Salomon, Véronique Brouste, Marie-Christine Mathieu, Pascal Roger, Frédéric Bibeau, Jocelyne Jacquemier, et al. High inter-observer agreement in immunohistochemical evaluation of her-2/neu expression in breast cancer: a multicentre gefpics study. *European Journal of Cancer*, 42(17):2946–2953, 2006.

[3] Melissa Cregger, Aaron J Berger, and David L Rimm. Immunohistochemistry and quantitative analysis of protein expression. *Archives of pathology & laboratory medicine*, 130(7):1026–1030, 2006.

[4] Farzad Ghaznavi, Andrew Evans, Anant Madabhushi, and Michael Feldman. Digital imaging in pathology: whole-slide imaging and beyond. *Annual Review of Pathology: Mechanisms of Disease*, 8:331–359, 2013.

[5] Patrick L Fitzgibbons, David L Page, Donald Weaver, Ann D Thor, D Craig Allred, Gary M Clark, Stephen G Ruby, Frances O'Malley, Jean F Simpson, James L Connolly, et al. Prognostic factors in breast cancer: College of american pathologists consensus statement 1999. *Archives of pathology & laboratory medicine*, 124(7):966–978, 2000.

[6] DJ Slamon, G Clark, S Wong, W Levin, A Ullrich, and W McGuire. Human breast cancer: correlation of relapse and. *science*, 3798106(177):235, 1987.

[7] Jeffrey S Ross and JA Fletcher. Her-2/neu (c-erb-b2) gene and protein in breast cancer. *American journal of clinical pathology*, 112(1 Suppl 1):S53–67, 1999.

[8] Salar Razavi, Gökhan Hatipoğlu, and Hülya Yalçın. Automatically diagnosing her2 amplification status for breast cancer patients using large fish images. In *Signal Processing and Communications Applications Conference (SIU), 2017 25th*, pages 1–4. IEEE, 2017.

[9] Antonio C Wolff, M Elizabeth H Hammond, David G Hicks, Mitch Dowsett, Lisa M McShane, Kimberly H Allison, Donald C Allred, John MS Bartlett, Michael Bilous, Patrick Fitzgibbons, et al. Recommendations for human epidermal growth factor receptor 2 testing in

breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline update. *Journal of clinical oncology*, 31(31):3997–4013, 2013.

[10] Henrik O Helin, Vilppu J Tuominen, Onni Ylinen, Heikki J Helin, and Jorma Isola. Free digital image analysis software helps to resolve equivocal scores in her2 immunohistochemistry. *Virchows Archiv*, 468(2):191–198, 2016.

[11] M Milagro Fernández-Carrobles, Gloria Bueno, Marcial García-Rojo, Lucía González-López, Carlos López, and Oscar Déniz. Automatic quantification of ihc stain in breast tma using colour analysis. *Computerized Medical Imaging and Graphics*, 61:14–27, 2017.

[12] Mira Valkonen, Kimmo Kartasalo, Kaisa Liimatainen, Matti Nykter, Leena Latonen, and Pekka Ruusuvuori. Metastasis detection from whole slide images using local features and random forests. *Cytometry Part A*, 91(6):555–565, 2017.

[13] Jiří Borovec and J Kybic. Fully automatic segmentation of stained histological cuts. In *International Student Conference on Electrical Engineering*, volume 17, pages 1–7, 2013.

[14] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, Sabine Süsstrunk, et al. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[15] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[16] Mryka Hall-Beyer. Glcm texture: a tutorial. *National Council on Geographic Information and Analysis Remote Sensing Core Curriculum*, 2000.