

ORIGINAL PAPER

A ROBUST AND AUTOMATED CELL COUNTING METHOD IN QUANTIFICATION OF DIGITAL BREAST CANCER IMMUNOHISTOCHEMISTRY IMAGES

LU CHEN¹, JI BAO¹, QIANG HUANG², HUAIQIANG SUN²¹Department of Pathology, West China Hospital, Sichuan University²Imaging Research Core Facilities, West China Hospital, Sichuan University

Quantitative analysis of immunohistochemically stained breast cancer specimens by cell counting is important for prognosis and treatment planning. This paper presents a robust, accurate, and novel method to label immunopositive and immunonegative cells automatically. During preprocessing, we developed an adaptive method to correct the colour aberration caused by imaging conditions. Next, a pixel-level segmentation was performed on preprocessed images using a support vector machine with a radial basis function kernel in HSV colour space. The segmentation result was processed by mathematical morphology operations to correct error-segmented regions and extract the marker for each cell. Validation studies showed that the automated cell-counting method had divergences varying from -5.05% to 3.99% compared with manual counting by a pathologist, indicating considerable agreement of the present automated cell counting method with manual counting. Thus, this method can free pathologists from laborious work and can potentially improve the accuracy and the reproducibility of diagnosis.

Key words: breast cancer, cell counting, immunohistochemistry, support vector machine (SVM), mathematical morphology.

Introduction

Being the most commonly diagnosed type of cancer among women, breast cancer is expected to account for 29% of all new cancers among women [1]. It is well known that breast cancer exhibits an exceptionally heterogeneous phenotype in histopathology [2], which can be exploited for precision medicine therapies. Many attempts have been made in the past to establish reliable and reproducible prognostic markers. For example, oestrogen receptor (OR) and progesterone receptor (PR) expression levels have been used to predict the response to endocrine therapy [3, 4], and Ki-67 is used to evaluate the proliferation activity of cancer cells [5]. The level of these markers is routinely evaluated using immunohistochemistry

(IHC). Immunohistochemistry is a technique to localise antigens (e.g. proteins) in the cells of a tissue section using an antibody-antigen interaction. Immunohistochemistry staining is widely used in the practice of diagnostic pathology and basic research to visualise the distribution and the localisation of various marker proteins. In routine IHC protocol, cells expressing corresponding protein are stained brown or red brown. The negative cells that do not express the specific protein are consequently stained blue, purple, or cyan according to the acidity of the staining solution.

Quantitative analysis of the immunohistochemically stained images is required in breast cancer diagnosis [6, 7, 8]. The analysis is usually performed by cell counting and then by calculating the label index (LI). Label index is defined as the percentage of im-

munopositive cells that express the specific protein with regards to the total number of them. However, to identify immunopositive cells and calculate the LI in several images manually is laborious, irreproducible, and time-consuming. Thus, an automated computer-assisted positive cell detection and LI calculation method has the potential to assist pathologists significantly.

Unlike greyscale images from computed tomography (CT) or magnetic resonance imaging (MRI), histopathology images are usually chromatic and thus more complicated. The quality of IHC images is significantly affected by many procedures, such as specimen preparation, staining, and the imaging process. There are several issues that make precise cell segmentation and counting a very challenging task:

- Colour aberration caused by the staining process and imaging devices.
- Unpredictable shape and size variations of tumour cells.
- The presence of clustered or touching cells.
- The presence of stroma cells and lymphocytes, which are not subject to counting.
- Intensity variation within nuclei due to chromatin texture.
- Poor contrast between faintly stained cells and background in some slides.

To solve the colour constancy problem in digital microscope images, Levenson *et al.* introduced a multispectral imaging technique [9], which used precise measurement of optical spectra at every pixel of an image to overcome the deficiencies of traditional imaging methods. The spectral information of each pixel contributes to a more precise discrimination of slight colour differences. Wu *et al.* achieved good results in the segmentation of white blood cells by using multispectral imaging techniques [10]. To capture a multispectral image, an additional light source and camera are required. Bell *et al.* employed a high dynamic range (HDR) technique in histopathology image analysis [11, 12]. The HDR images can be acquired by combining a set of traditional digital images that differ in exposure time. The HDR technique can reduce noise and improve the measurement of colours. However, many small hospitals may not be able to afford the cost of a multispectral camera, and it is also impractical to capture several images of the same view field with different bands and exposure time in routine diagnostic work.

Automated cell-counting method has been extensively investigated with the advent of digital imaging and automatic image analysis. Several algorithms have been developed, and newer methods continue to emerge [13, 14, 15, 16, 17, 18, 19, 20]. Many of these approaches have achieved promising results on some samples. However, there has been relatively

limited consideration given to optimising approaches for handling the deviation in random samples. In addition, in the development efforts of our methods, we found it difficult to distinguish between de facto immunopositive cells that expressed low-level antigen and immunonegative cells using unsupervised methods, such as expectation-maximisation algorithm or k-mean clustering. In this paper, we propose a new method based on a support vector machine (SVM) and mathematical morphology to label the immunopositive and immunonegative cells and calculate the LI automatically in breast cancer IHC images without manual input. It can be applied to conventional digital microscope images, and no extra hardware is required. The result shows considerable agreement with judgments made manually by an experienced clinical pathologist.

Material and methods

Image dataset

This retrospective study was approved by the institutional Medical Ethics Committee, which waived the requirement for informed consent. A total of 60 IHC-stained slides from post-operative surgical resection specimens of originally diagnosed breast cancer between 2010 and 2015 were obtained from the Department of Pathology, West China Hospital. Twenty slides (12 Luminal A cases and eight Luminal B cases) were stained with anti-ER rabbit monoclonal antibody (mAb) (Ventana, SP1), 20 slides (10 Luminal A cases and 10 Luminal B cases) with anti-PR rabbit mAb (Ventana, 1E2), and 20 slides (12 Luminal A cases and eight Luminal B cases) with anti-Ki-67 rabbit mAb (Ventana, 30-9). Specimen fixation and paraffin embedding was performed using standard histological techniques. Serial sections at 5 μm were cut from formalin-fixed paraffin-embedded tissue blocks then mounted onto adhesive slides. Deparaffinisation, antigen retrieval, endogenous peroxidase blocking, and antigen staining were performed on the BenchMark XT automated slide preparation and staining system (Ventana, Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions and using Ventana reagents for the entire procedure. All slides used diaminobenzidine (DAB) as the chromogen and haematoxylin for the counterstaining. A representative field of view was captured from each slide in the carcinomas area at 400 \times magnification using a Leica DM4000B microscope with a Leica DP-495 CCD. Thus, a total of 60 digital images were investigated. Half of the images were randomly selected from each stained antigen to form the training set, and the rest were defined as the test set.

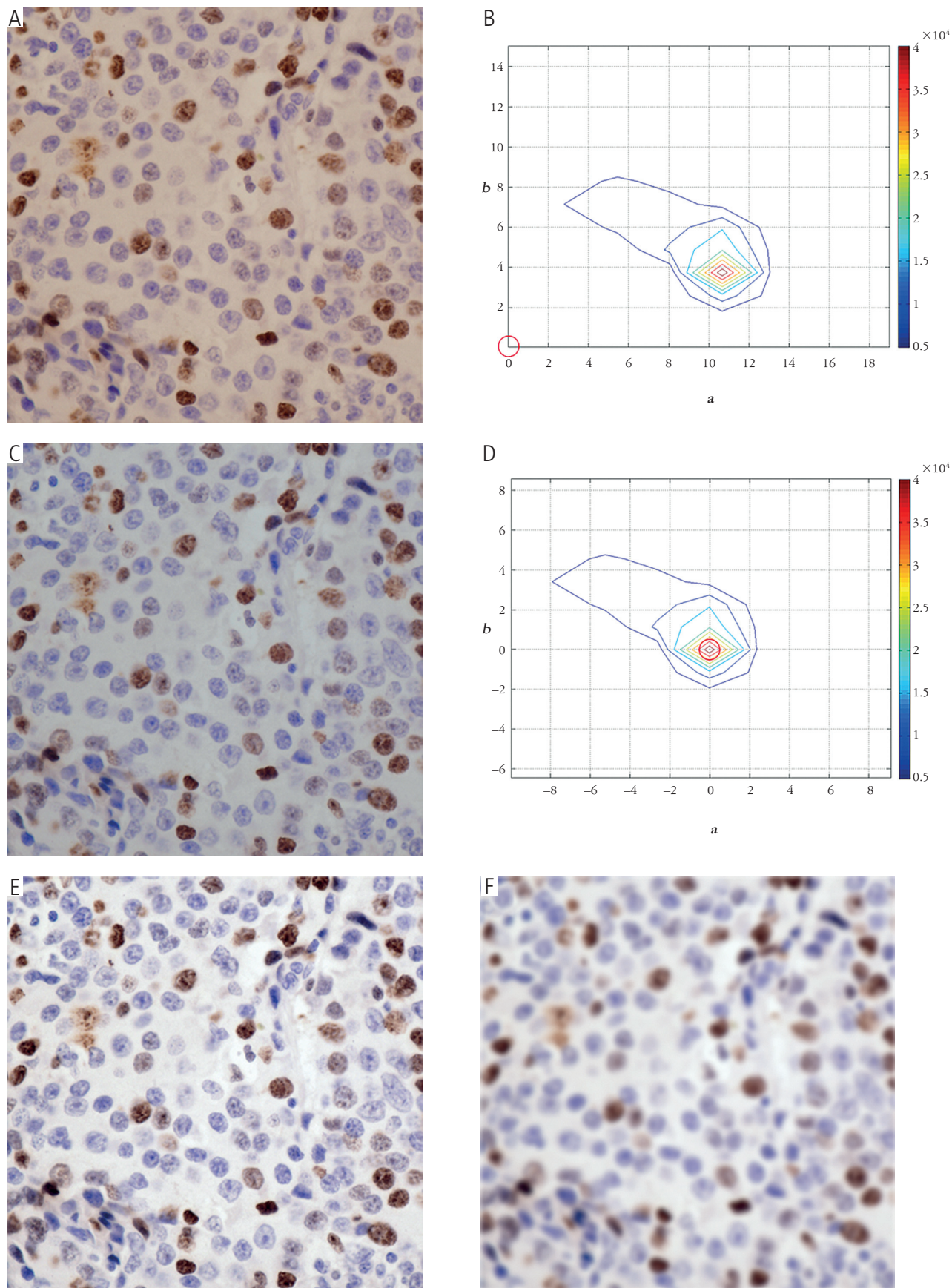


Fig. 1. The image preprocessing. A) The raw image directly from the microscope. B) The ab 2-D histogram of (A). The red circle indicates the original point of the ab panel. C) Image after colour adjustment. D) The ab 2-D histogram of (A). E) Image after contrast stretch on L^* coordinate. F) Image after Gaussian filter

Image preprocessing

Immunohistochemistry images captured from a conventional bright-field microscope may have several variations that can affect the observed colour of stained objects. The variations may be due to the microscope component, such as the different colour characteristics of the light source, intensity adjustment according to personal custom, and the white balance settings. Handling these variations will facilitate the segmentation process and contribute to the robustness of the method. In the experiment, we developed an adaptive method to detect and remove colour cast automatically in breast cancer IHC images.

The colour gamut of the image is first mapped into the CIEL*a*b* colour space; the conversion method was described previously [21]. In this model, the L* coordinate, which varies from 1 to 100, represents the lightness of the colour. On the a* axis, positive values indicate the amounts of red while negative values indicate the amounts of green. On the b* axis, yellow is positive and blue is negative. For both axes, the value of zero indicates neutral grey. Thus, the chromatic component and the lightness are separated. We can decompose and analyse the colour space in a one-dimensional L* coordinate and in two-dimensional ab panel, respectively. Intuitively, the two-dimensional colour histogram $F(a, b)$ only depends on the colour properties of each pixel.

As we observed, most areas in breast cancer IHC images are stroma and background, which should be represented closely to white or grey in ideal conditions. So the point (x, y) that contributes to the maximum of $F(a, b)$ should be at the original point of the ab chromatic panel. The farther the highest peak of $F(a, b)$ is from the neutral axis of the ab panel, the stronger the colour cast that exists. To correct the colour cast, we transformed $F(a, b)$ to $F((a-x), (b-y))$, to make the highest peak of the 3-D histogram align with the neutral axis of the ab chromatic panel (Fig. 1A-D).

The difference in the illumination and contrast of each image can be eliminated by performing contrast stretching on the L* coordinate. Contrast stretching applies a linear transformation to stretch histogram values of L* across the full range of possible values. To prevent a small number of pixels with outlier intensity affecting the result, a certain percentage (2% in this experiment) of pixels were treated as saturated. This essentially winsorises the distribution, setting pixels with an intensity in the first percentile to 0, and those after the 99th percentile to 100. Values in between were linearly scaled in the range 0 to 100 (Fig. 1E).

Variable chromatin texture is another source of segmentation error; a Gaussian filter with appropriate standard deviation (according to the resolu-

tion of the image) was employed to eliminate uneven staining within nuclei, which could influence the integrity of segmentation in the next step (Fig. 1F).

Image segmentation

The segmentation process is a supervised pixel-level classification. This process includes two steps: the training of the classifier and pixel classification.

After preprocessing, the colour aberration caused by imaging condition and capture device was removed. A slight fluctuation in hue still exists due to the type of colour reagent used and differences in the acidity of the staining solution, which are not fully eliminated by preprocessing. Thus, a classifier with good generalisation performance is required for the pixel classification step. Accordingly, SVM was chosen in our experiment.

SVM is a relatively new type of classification algorithm, originally introduced by Vapnik and successively extended by a number of other researchers [22, 23]. The goal of SVM is to find the largest margin between the separating hyperplane and the dataset. In the binary classification problem, given a training set of labelled pairs (\mathbf{x}_i, y_i) , $i = 1 \dots N$, where $\mathbf{x}_i \in \mathbb{R}^n$ corresponds to the feature vector in n-dimensional space and $y \in \{-1, 1\}^N$ denotes its class label, the SVM require the solution of the following optimisation problem

$$\text{minimise } J(\mathbf{w}, b, \boldsymbol{\zeta}) = \frac{1}{2}(\mathbf{w}^T \mathbf{w} + C) \sum_{i=1}^N \zeta_i \quad (1)$$

$$\text{subject to } y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \zeta_i, \zeta_i \geq 0 \quad (2)$$

where the parameter \mathbf{w} is normal to the separating hyperplane, variable $C > 0$ is a user-specified penalty parameter of the error term, and variables ζ_i are known as slack variables that can be used to construct a soft margin hyperplane. Input vectors \mathbf{x}_i are mapped to a higher dimensional space by the function ϕ , while SVM finds a linear separating hyperplane with maximum margin between two classes. Then, an unknown sample \mathbf{x} can be classified according to the following equation

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \phi(\mathbf{x}) + b) = \text{sign} \left(\sum_{i=1}^{N_s} a_i y_i \phi(\mathbf{s}_i) \cdot \phi(\mathbf{x}) + b \right) \quad (3)$$

where variables \mathbf{s}_i are support vectors, and N_s is the number of support vectors. $\phi(\mathbf{s}_i) \cdot \phi(\mathbf{x})$ can be expressed as the kernel function $K(\mathbf{s}_i, \mathbf{x})$. There are two advantages of SVM: the generalisation ability of the SVM method is optimised by maximising the margin distance of classes, and nonlinear classification issues can be solved by mapping data to a higher

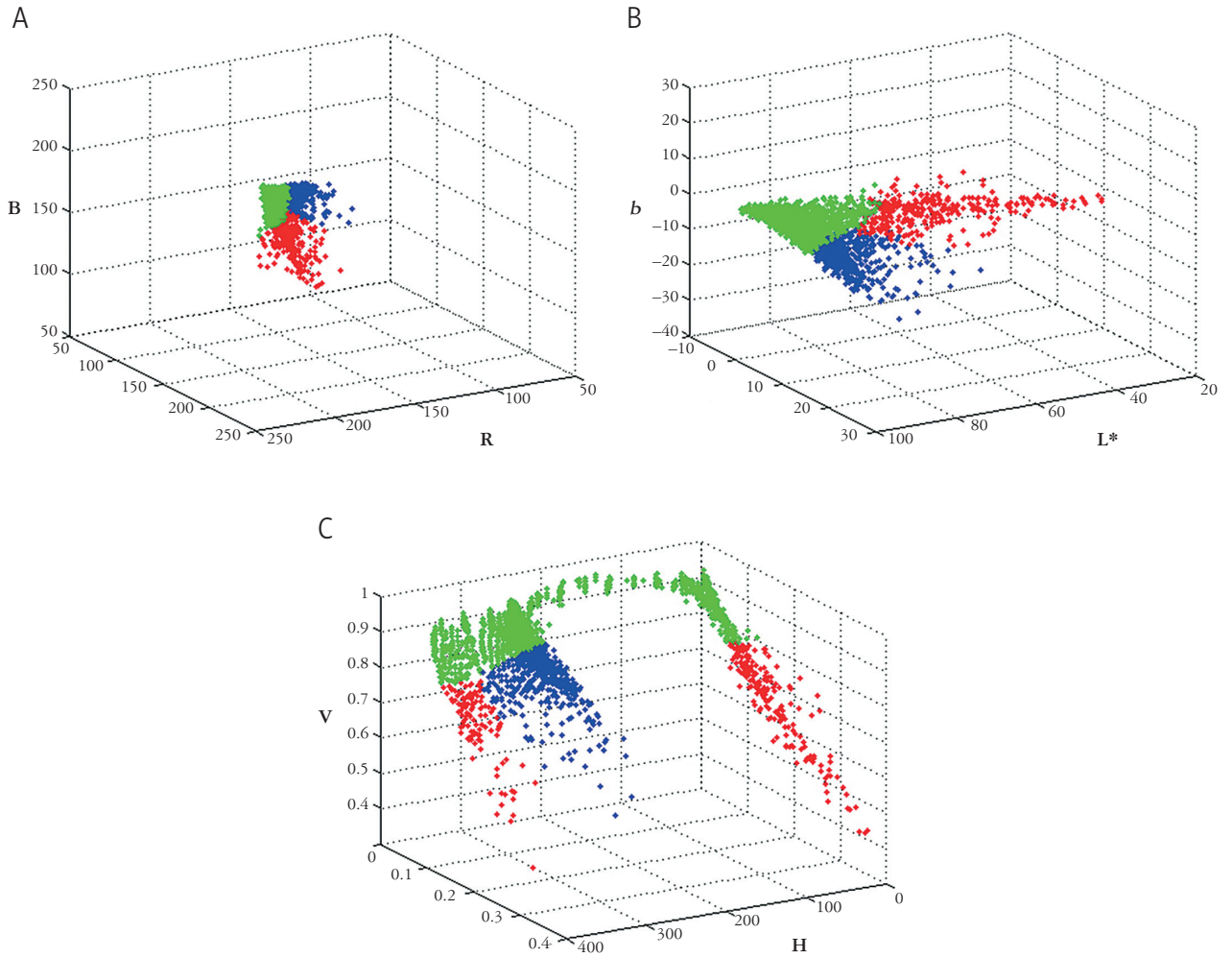


Fig. 2. The distribution of pixels of Fig. 1E in each colour space. A) RGB space. B) CIEL*a*b* space. C) HSV space. The red dots present pixels belonging to immunopositive nuclei, the blue dots present pixels belonging to immunonegative nuclei, and the green dots present pixels belonging to stroma and background

dimensional feature space using the kernel function. Because there are three types of components: immunopositive nuclei, immunonegative nuclei, and background in the images, we constructed a multi-category support vector machine according to the literature [24].

The CIEL*a*b* colour space shows its superiority in colour adjustment. However, this colour space is not always the most appropriate for all the image processing problems, especially for pixel classification in IHC images. We tested several frequently used colour spaces and found that HSV colour space is the most suitable for the IHC image segmentation task. HSV stands for hue, saturation, and value. The pixels from different stained objects are better separated in HSV space than in any other colour space (Fig. 2).

For the training step, preprocessed images from the training set were manually segmented under the direction of a pathologist. Pixels in the region

of positive, negative, and background were converted into HSV colour space. For each pixel in the image, the feature vector can be expressed as $\mathbf{p}_{i,j} = (H_{i,j}, S_{i,j}, V_{i,j})^T$. After removing duplicate entries, the colour feature and the class of pixels were saved as the training dataset.

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (4)$$

In training SVM, appropriate kernel function and its parameters are also significant, as well as the training data. Because the minimum number of parameters should be determined and the number of features is small (H,S,V), the radial basis function (RBF) kernel is a reasonable choice [25]. Thus, two parameters for RBF kernel SVM, C and γ , should be determined. A good (C, γ) could contribute to the prediction of unknown data accurately. A common strategy to identify the best (C, γ) is known as cross-validation.

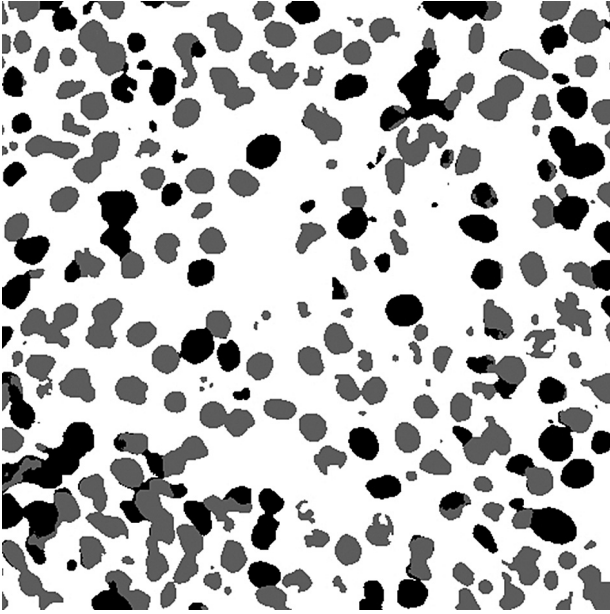


Fig. 3. The segmentation result of Fig. 1F

The training set was divided into n subsets of equal size. Each subset is tested by the classifier trained by the remaining $n-1$ subsets, and the accuracy is the percentage of data correctly classified. Various pairs of (C, γ) are tried, and the pair with the best cross-validation accuracy is obtained. Then the best (C, γ) ($C = 512, \gamma = 0.125$ in our experiment) and the whole dataset were used to generate the final classifier.

For the pixel classification step, the test images were first mapped into HSV colour space, then the colour feature of each pixel was fed into the trained classifier. The output is the class label of each pixel, which can be illustrated as a ternary image (Fig. 3).

Mathematical morphology processing and cell labelling

Mathematical morphology is a very powerful tool for processing geometric structures, which was established based on set theory [26]. It had been widely applied to many medical image analysis tasks to decompose complex shapes into more meaningful representations and separate them from undesirable parts [27, 28, 29]. In our experiment, both binary and greyscale mathematical morphologies were employed to correct error-classified regions and to extract the marker of each cell.

Basic mathematical morphology operations include dilation, erosion, opening, and closing. A detailed definition of these operations is available in the literature [26]. Firstly, a binary opening operation with a disk shape element of appropriate size was employed in the segmented image to remove undesirable parts, like error-classified region,

non-specific staining, and endothelial cells. It also makes the boundary of cells smooth and less clustered (Fig. 4B).

Before extracting the marker of each cell, a distance transform was performed on the binary image containing positive regions and negative regions, respectively. The distance $dist_x$ associated with a set X is defined as:

$$\forall p \in X, dist_x(p) = \min\{n \in \mathbb{N} | p \in X \ominus nB\} \quad (5)$$

where \ominus denotes the erosion operation and B is a unit element. For each pixel p belonging to X , $dist_x(p)$ is the distance between p and the background. The Euclidean distance map of the positive region binary image is shown in Fig. 4C. The centre of each nucleus may have the regional maximum distance. By definition, the regional maximum at altitude b of the distance map is a connected component at altitude b where the altitude of every neighbouring pixel is strictly smaller than b . Thus, the regional maxima in the distance map can be extracted as the marker of each cell. The result is shown in Fig. 4D.

However, the marking is still not perfect, some cells are multiple marked and some markers emerge at the joints of clustered cells. These problems are caused by the discrete workspace as well as small contour irregularities of some cells. To solve these problems, a greyscale opening operation with a disk shape element of appropriate size was performed on the distance map before calculating the regional maxima (Fig. 4E). The greyscale opening on the distance map could merge adjacent regional maxima and remove small-sized regional maxima (Fig. 4F), finally using the centroid of each regional maxima as the cell marker.

Results

The images from the previously defined test set were used to evaluate the accuracy of the proposed method. The proposed method can automatically separate nuclei from their background and label them as either positively or negatively stained in a visually acceptable way. The influence of colour cast caused by imaging conditions and device characteristics can be successfully removed by our breast tissue-specific preprocessing procedure. It is noteworthy that some immunopositive cells with a very low concentration antigen are also correctly labelled. Part of the results are shown in Fig. 5. To quantitatively measure the performance of the automated method, we also carried out a comparison between the result achieved from our automated method and manual counting done by an experienced pathologist in 30 cases selected from each type of staining (results are presented in Table I).

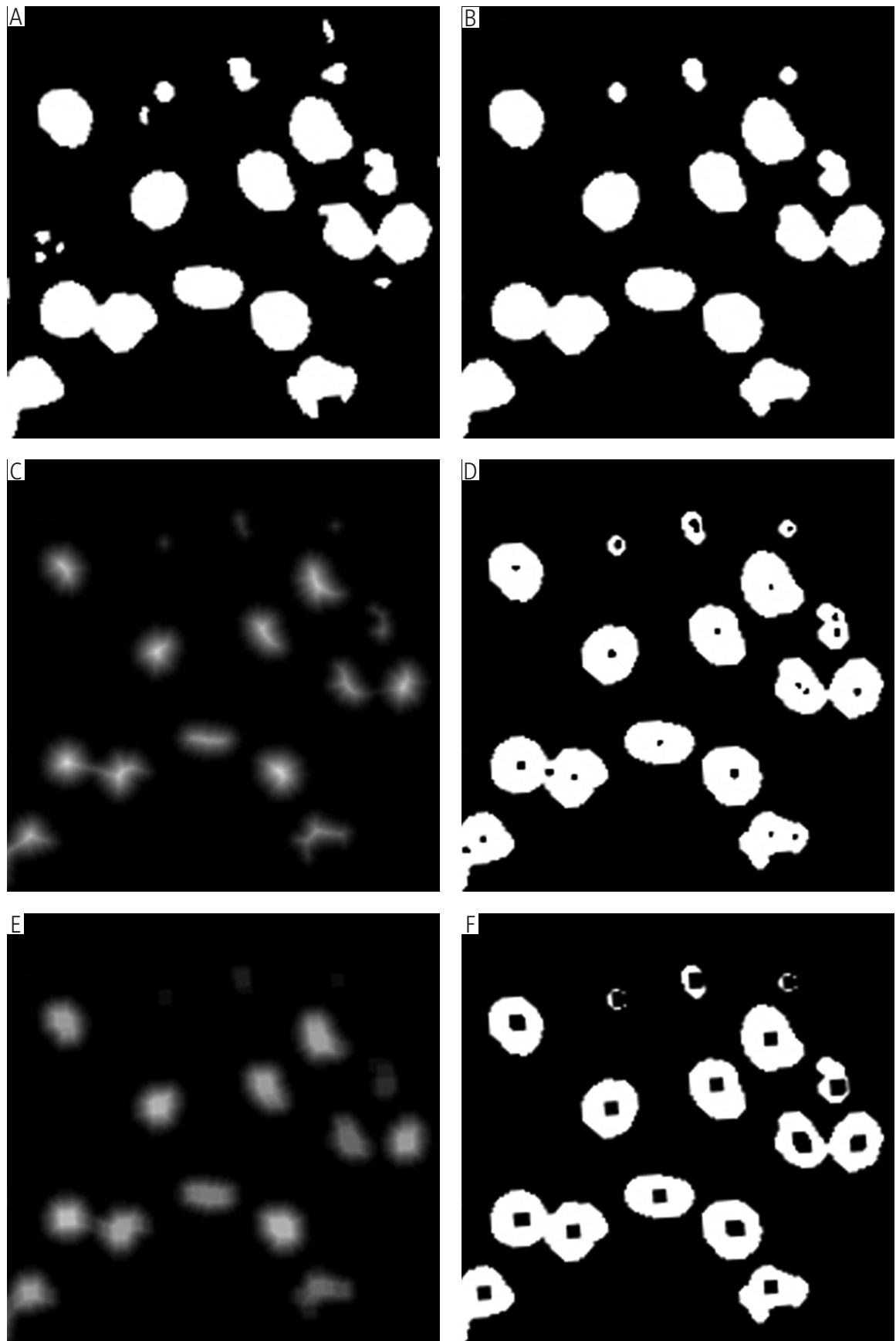


Fig. 4. The mathematical morphology processing. A) The raw segmentation result. B) After binary opening operation. C) Distance transform of (B). D) The regional maxima from (C). E) Greyscale opening operation work on (C). F) The regional maxima of (E)

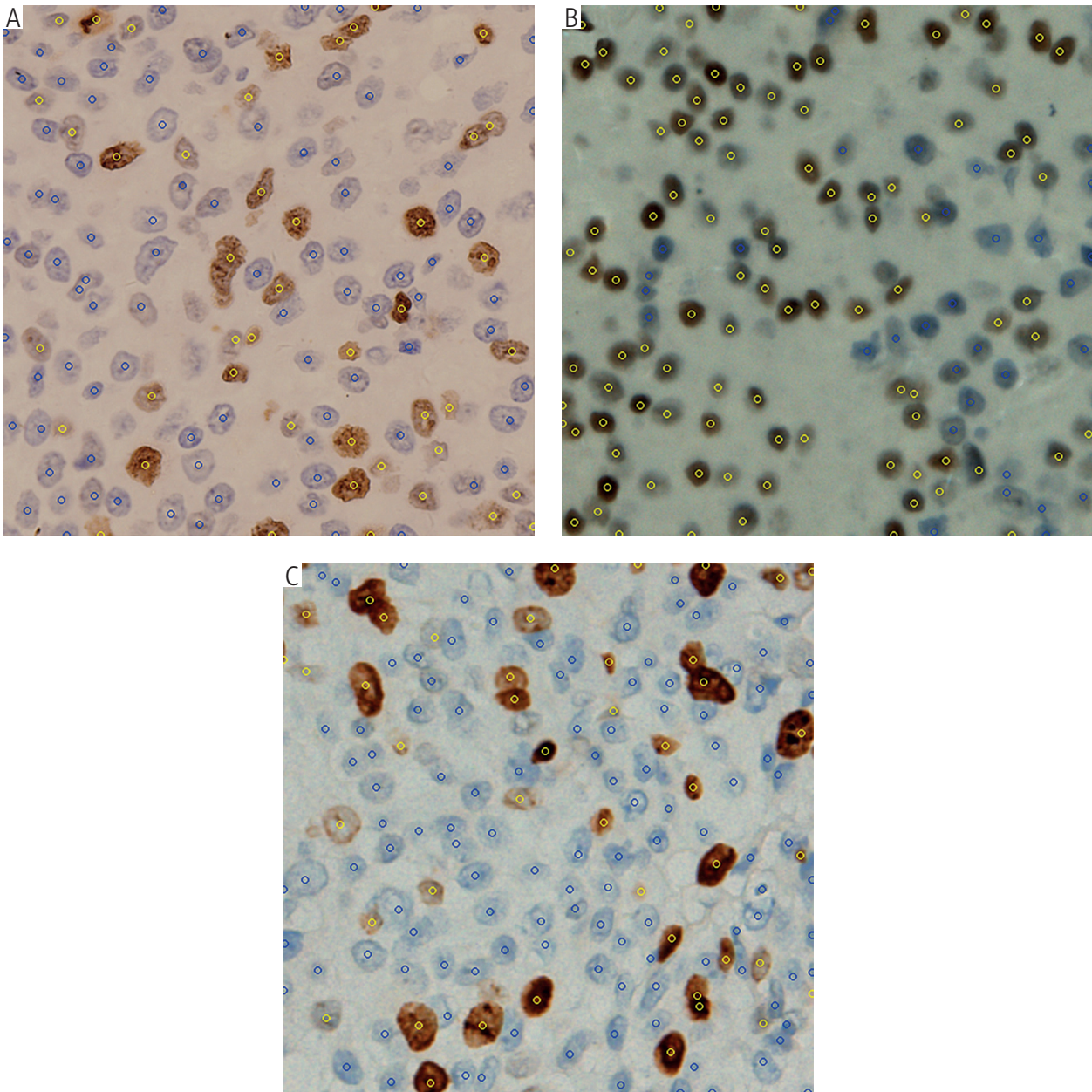


Fig. 5. The final result achieved from the proposed automated method on PR (A), ER (B), and Ki67 (C) – stained images

A Bland-Altman plot [30] was employed to assess the agreement between automated counting and manual counting. Bland Altman plot was introduced by J. Martin Bland and Douglas G. Altman to describe the agreement between two quantitative measurements. It uses a “quality control” concept instead of a p-value to describe the agreement. The difference between the paired measurements is plotted against the mean of the two measurements, and it is recommended that 95% of the data points lying within the ± 2 -fold standard deviation (SD) of the mean difference can be interpreted as good agreement. Bland-Altman plots were performed on

the counting result of immunopositive cells, immunonegative cells, and LI.

Figure 6A shows a comparison of the number of immunopositive cells achieved from manual counting and automated counting in all 30 test cases. Here the mean (SD) of the differences is $-2.97 (5.25)$, thus the mean $\pm 2SD$ is $(-13.47, 7.53)$. We observed that 2/30 (6.67%) of the points are beyond the mean $\pm 2SD$. The automated method tends to give a higher reading. Figure 6B presents the comparison of the number of immunonegative cells achieved from both methods. The mean (SD) is $-1.73 (7.03)$, and nearly all the points are in the 95% confidence interval. The result from automated counting is

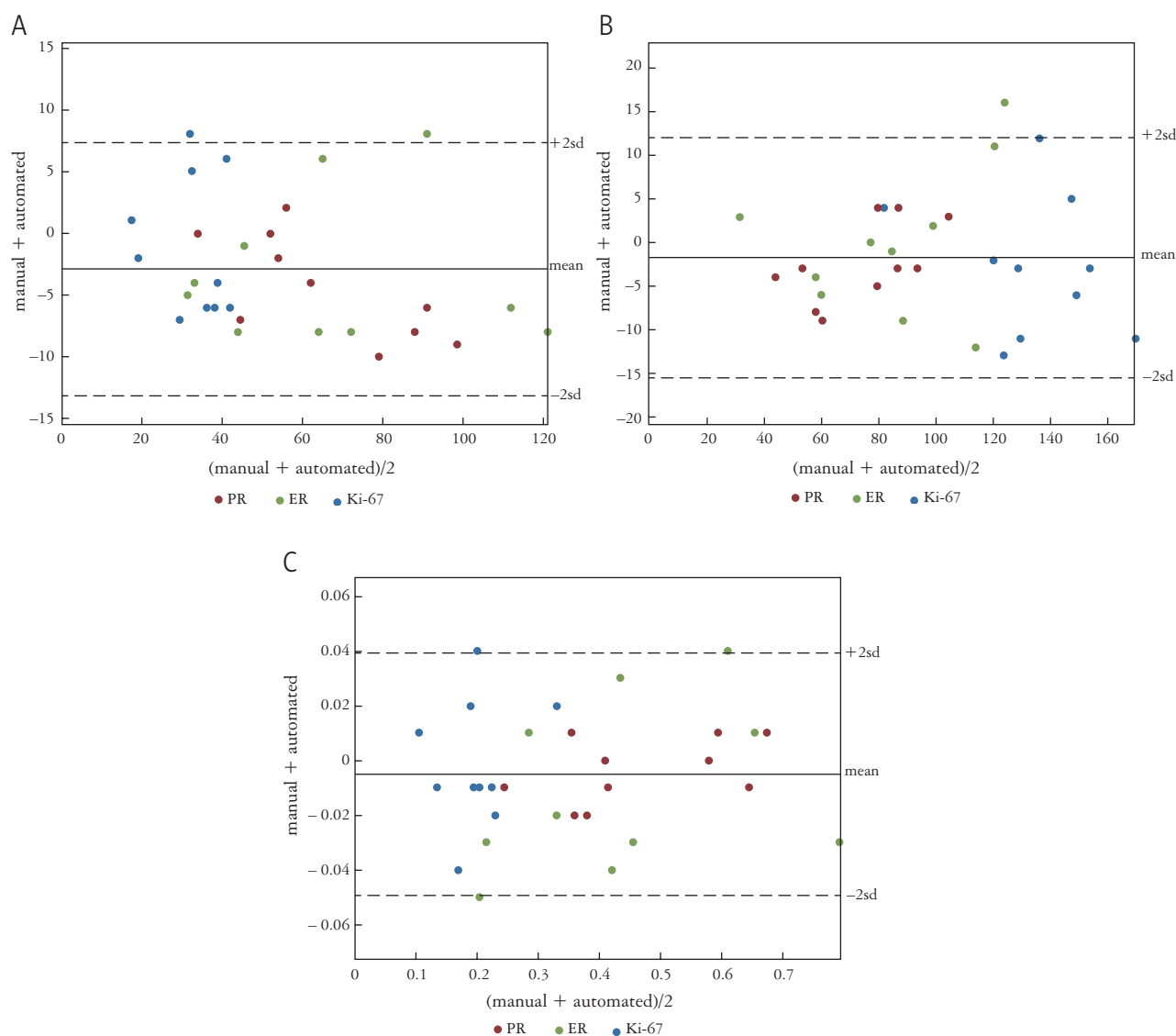


Fig. 6. Bland-Altman plot for comparison between manual counting and automated counting on immunopositive cells (A), immunonegative cells (B), and LI (C)

also slightly higher than manual counting. Figure 6C shows the comparison of LI achieved from both methods. The mean (SD) of differences is -0.53% (2.26%). Defining the manual counting as the “gold standard”, the final results from the proposed automated method have differences varying from -5.05% to 3.99% , which is acceptable in clinical diagnosis. The result from the Bland-Altman plot indicates that the proposed automated method has excellent agreement with manual counting.

Discussion

The expression of oestrogen/progesterone receptor as well as Ki-67 in breast carcinoma is highly associated with chemoin sensitivity and prognosis. The predictive and prognostic capability of these markers

is enhanced by quantitative immunohistochemistry. The American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) has recommended the use of image analysis for estimating the percentage of immunopositive cells [31]. However, the conventional manual method for quantitating the marker expression is not only labour-intensive and time-consuming but also prone to introduce inter-observer differences in the interpretation of results. Automated analysis may thus increase the speed and reproducibility of evaluation. In addition, recent studies also indicated that continuous variable output from automated analysis methods may allow identification of IHC cut-off values of prognostic relevance that are barely detected by visual evaluation by a pathologist [32].

Table I. Comparison of results achieved by automated counting and by manual counting

| | CASE No. | MANUAL | | | AUTOMATED | | |
|-------|----------|------------------|------------------|--------|-----------|-----|--------|
| | | IPC ^a | INC ^b | LI | IPC | INC | LI |
| PR | case 1 | 84 | 56 | 60.00% | 92 | 65 | 58.60% |
| | case 2 | 41 | 77 | 34.75% | 48 | 82 | 36.92% |
| | case 3 | 53 | 89 | 37.32% | 55 | 85 | 39.29% |
| | case 4 | 34 | 106 | 24.29% | 34 | 103 | 24.82% |
| | case 5 | 74 | 54 | 57.81% | 84 | 62 | 57.53% |
| | case 6 | 52 | 92 | 36.11% | 52 | 95 | 35.37% |
| | case 7 | 60 | 85 | 41.38% | 64 | 88 | 42.11% |
| | case 8 | 88 | 42 | 67.69% | 94 | 46 | 67.14% |
| | case 9 | 94 | 52 | 64.38% | 103 | 55 | 65.19% |
| | case 10 | 57 | 82 | 41.01% | 55 | 78 | 41.35% |
| ER | case 1 | 95 | 56 | 62.91% | 87 | 60 | 59.18% |
| | case 2 | 68 | 84 | 44.74% | 62 | 85 | 42.18% |
| | case 3 | 109 | 57 | 65.66% | 115 | 63 | 64.61% |
| | case 4 | 68 | 100 | 40.48% | 76 | 98 | 43.68% |
| | case 5 | 29 | 132 | 18.01% | 34 | 116 | 22.67% |
| | case 6 | 117 | 33 | 78.00% | 125 | 30 | 80.65% |
| | case 7 | 40 | 84 | 32.26% | 48 | 93 | 34.04% |
| | case 8 | 60 | 77 | 43.80% | 68 | 77 | 46.90% |
| | case 9 | 45 | 108 | 29.41% | 46 | 120 | 27.71% |
| | case 10 | 31 | 126 | 19.75% | 35 | 115 | 23.33% |
| Ki-67 | case 1 | 37 | 152 | 19.58% | 41 | 155 | 20.92% |
| | case 2 | 18 | 119 | 13.14% | 20 | 121 | 14.18% |
| | case 3 | 35 | 124 | 22.01% | 30 | 135 | 18.18% |
| | case 4 | 36 | 142 | 20.22% | 28 | 130 | 17.72% |
| | case 5 | 26 | 150 | 14.77% | 33 | 145 | 18.54% |
| | case 6 | 35 | 127 | 21.60% | 41 | 130 | 23.98% |
| | case 7 | 44 | 84 | 34.38% | 38 | 80 | 32.20% |
| | case 8 | 39 | 164 | 19.21% | 45 | 175 | 20.45% |
| | case 9 | 33 | 117 | 22.00% | 39 | 130 | 23.08% |
| | case 10 | 18 | 146 | 10.98% | 17 | 152 | 10.06% |

^a *ipc* stands for immunopositive cells

^b *inc* stands for immunonegative cells

The present work, combined with colour adjustment, SVM, and mathematical morphology, has proven to be robust and accurate. It would free pathologists from laborious work and improve the accuracy and the reproducibility of diagnosis. It can be directly used on traditional digital images captured from a camera mounted on a microscope. For the virtual pathology slide generated by a slide scanner, the pathologist can manually create a mask to define the area to perform cell counting.

Underestimation and overestimation are the two main errors in the proposed automated method.

Several known sources of errors are analysed in our experiment. Overestimation usually happens when a nucleus chromatin is highly textured or when the nucleus shape is extremely elongated. It is particularly common with nuclei that deviate significantly from regular shape if the specimen was not correctly fixed. Another cause of overestimation is the interference of some vascular endothelial cells. Underestimation usually occurs when the nuclei are densely clustered or when the nuclei are very weak stained, the latter case is usually found in immunonegative nuclei.

A traditional difficulty with automated algorithms is the effort required to tune the parameters when encountering new images. Because our automated method is specifically designed for breast tissue and the cell counting work is set to perform under $400\times$ magnification, the size of the element in mathematical morphology processes depends only on the resolution of the captured image. These parameters can be determined automatically according to the resolution of the captured image. Also, one would only need to retrain the classifier with sample images if the staining agent changed.

However, there is still room for improvement in the proposed method. The method can only assess the nuclei-stained marker currently. There are also membrane-staining markers, like HER-2 and EGFR, and a quantitative analysis is also required for these markers [33, 34]. Other than LI, the staining density and distribution of immunopositive cells are mentioned in the latest breast cancer diagnosis guidelines [31]. Our future work will focus on the assessment of non-nuclei-stained marker and will involve the staining intensity assessment in our automated quantification method.

The authors are grateful to their colleagues at the Department of Pathology for their cooperation and for kindly providing us with the experimental images and directions.

The authors declare no conflict of interest.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016; 66: 7-30.
2. Bertucci F, Birnbaum D. Reasons for breast cancer heterogeneity. *J Biol* 2008; 7: 6.
3. Mohsin SK, Weiss H, Havighurst T, et al. Progesterone receptor by immunohistochemistry and clinical outcome in breast cancer: a validation study. *Mod Pathol* 2004; 17: 1545-1554.
4. Kaufmann M, Pusztai L, Biedenkopf Expert Panel Members. Use of standard markers and incorporation of molecular markers into breast cancer therapy: Consensus recommendations from an International Expert Panel. *Cancer* 2011; 117: 1575-1582.
5. de Azambuja E, Cardoso F, de Castro G, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer* 2007; 96: 1504-1513.
6. Veronese SM, Gambacorta M, Gottardi O, et al. Proliferation index as a prognostic marker in breast cancer. *Cancer* 1993; 71: 3926-3931.
7. Park D, Kåresen R, Noren T, Sauer T. Ki-67 expression in primary breast carcinomas and their axillary lymph node metastases: clinical implications. *Virchows Arch* 2007; 451: 11-18.
8. Chung GG, Zerkowski MP, Ghosh S, et al. Quantitative analysis of estrogen receptor heterogeneity in breast cancer. *Lab Invest* 2007; 87: 662-669.
9. Levenson RM, Mansfield JR. Multispectral imaging in biology and medicine: Slices of life. *Cytom Part A* 2006; 69A: 748-758.
10. Wu Q, Zeng L, Zheng H, Guo N. Precise Segmentation of White Blood Cells by Using Multispectral Imaging Analysis Techniques. 2008 First International Conference on Intelligent Networks and Intelligent Systems. IEEE; 2008; 491-494.
11. Bell AA, Brauers J, Kaftan JN, et al. High Dynamic Range Microscopy for Cytopathological Cancer Diagnosis. *IEEE J Sel Top Signal Process* 2009; 3: 170-184.
12. Bell AA, Meyer-Ebrecht D, Bocking A, Aach T. HDR-Microscopy of Cell Specimens: Imaging and Image Analysis. 2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers. IEEE; 2007; 1303-1307.
13. Ranefall P, Egevad L, Nordin B, Bengtsson E. A New Method for Segmentation of Colour Images Applied to Immunohistochemically Stained Cell Nuclei. *Anal Cell Pathol* 1997; 15: 145-156.
14. Loukas CG, Wilson GD, Vojnovic B, Linney A. An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections. *Cytometry* 2003; 55A: 30-42.
15. Arambula Cosio F, Marquez Flores JA, Padilla Castaneda MA, et al. Automatic counting of immunocytochemically stained cells. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat No03CH37439). IEEE; 790-793.
16. Yang X, Li H, Zhou X. Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy. *IEEE Trans Circuits Syst I Regul Pap* 2006; 53: 2405-2414.
17. Bewes JM, Suchowerska N, McKenzie DR. Automated cell colony counting and analysis using the circular Hough image transform algorithm (CHiTA). *Phys Med Biol* 2008; 53: 5991.
18. Kothari S, Chaudry Q, Wang MD. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, 2009; 795-798.
19. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images. *IEEE Trans Biomed Eng* 2010; 57: 841-852.
20. Grala B, Markiewicz T, Kozlowski W, et al. New automated image analysis method for the assessment of Ki-67 labeling index in meningiomas. *Folia Histochem Cytobiol* 2009; 47: 587-592.
21. Connolly C, Fleiss T. A study of efficiency and accuracy in the transformation from RGB to CIELAB color space. *IEEE Trans Image Process* 1997; 6: 1046-1048.
22. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory – COLT '92. ACM Press, New York 1992; 144-152.
23. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995; 20: 273-297.
24. Lee Y, Lin Y, Wahba G. Multicategory Support Vector Machines. *J Am Stat Assoc* 2004; 99: 67-81.
25. Scholkopf B, Kah-Kay Sung, Burges CJC, et al. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans Signal Process* 1997; 45: 2758-2765.
26. Haralick RM, Sternberg SR, Zhuang X. Image Analysis Using Mathematical Morphology. *IEEE Trans Pattern Anal Mach Intell* 1987; PAMI-9: 532-550.
27. Klingler JW, Vaughan CL, Fraker TD, Andrews LT. Segmentation of echocardiographic images using mathematical morphology. *IEEE Trans Biomed Eng* 1988; 35: 925-934.
28. Zana F, Klein J-C. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Trans Image Process* 2001; 10: 1010-1019.
29. Zhao Yu-qian, Gui Wei-hua, Chen Zhen-cheng, et al. Medical Images Edge Detection Based on Mathematical Morphology. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE 2005; 6492-6495.
30. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-310.
31. Hammond MEH, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists

- Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *Arch Pathol Lab Med* 2010; 134: e48-e72.
32. Rizzardi AE, Johnson AT, Vogel RI, et al. Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol* 2012; 7: 1-10.
 33. Masmoudi H, Hewitt SM, Petrick N, et al. Automated Quantitative Assessment of HER-2/neu Immunohistochemical Expression in Breast Cancer. *IEEE Trans Med Imaging* 2009; 28: 916-925.
 34. Hall BH, Ianosi-Irimie M, Javidian P, et al. Computer-assisted assessment of the Human Epidermal Growth Factor Receptor 2 immunohistochemical assay in imaged histologic sections using a membrane isolation algorithm and quantitative analysis of positive controls. *BMC Med Imaging* 2008; 8: 11.

Address for correspondence

Huaiqiang Sun
Imaging Research Core Facilities
West China Hospital
Sichuan University
37 Guoxue alley
610041, Chengdu, China
e-mail: sunhuaiqiang@scu.edu.cn