



A new automatic image analysis method for assessing estrogen receptors' status in breast tissue specimens

Aymen Mouelhi ^{a,*}, Mounir Sayadi ^a, Farhat Fnaiech ^a, Karima Mrad ^b, Khaled Ben Romdhane ^b

^a ESSTT-SICISI Laboratory, University of Tunis, 5 Av. Taha Hussein, 1008 Tunis, Tunisia

^b Salah Azaiez Institute of Oncology, Morbid Anatomy Service, bd du 9 avril, Bab Saadoun, 1006 Tunis, Tunisia



ARTICLE INFO

Article history:

Received 26 May 2013

Accepted 19 October 2013

Keywords:

Medical image analysis
Breast tissue images
Nuclear segmentation
Cell recognition
Active contours
Watersheds
Neural networks

ABSTRACT

Manual assessment of estrogen receptors' (ER) status from breast tissue microscopy images is a subjective, time consuming and error prone process. Automatic image analysis methods offer the possibility to obtain consistent, objective and rapid diagnoses of histopathology specimens. In breast cancer biopsies immunohistochemically (IHC) stained for ER, cancer cell nuclei present a large variety in their characteristics that bring various difficulties for traditional image analysis methods. In this paper, we propose a new automatic method to perform both segmentation and classification of breast cell nuclei in order to give quantitative assessment and uniform indicators of IHC staining that will help pathologists in their diagnostic. Firstly, a color geometric active contour model incorporating a spatial fuzzy clustering algorithm is proposed to detect the contours of all cell nuclei in the image. Secondly, overlapping and touching nuclei are separated using an improved watershed algorithm based on a concave vertex graph. Finally, to identify positive and negative stained nuclei, all the segmented nuclei are classified into five categories according to their staining intensity and morphological features using a trained multilayer neural network combined with Fisher's linear discriminant preprocessing. The proposed method is tested on a large dataset containing several breast tissue images with different levels of malignancy. The experimental results show high agreement between the results of the method and ground-truth from the pathologist panel. Furthermore, a comparative study versus existing techniques is presented in order to demonstrate the efficiency and the superiority of the proposed method.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Millions of people die each year due to various types of cancer diseases and breast cancer is considered as one of the main causes of death among women. It has been shown that the ER status is an important biologic factor in predicting the response to hormonal therapy of breast cancers [1–6]. In daily clinical practice, IHC staining is mostly done to quantitate expression of ER in breast tissue specimens to further classify the cancer. Evaluation of ER positive status is based on a subjective assessment of cancer nuclei in each stained slide examined under microscope [7]. Therefore, automatic methods that are used to segment and to classify microscopic breast cancer images by level of disease can potentially make a significant contribution to health care. The major aim of these methods is to identify the different types of malignancy and to detect the real number of positive nuclei (ER+ status) in order to

give quantitative and accurate measures to assist pathologists in the cancer diagnosis process [1,5]. In fact, cancer cell nuclei must be segmented and classified in order to get the percentage of the positive nuclei in the whole IHC stained image. It is known that the accuracy of the classification method is mainly relied on the nature of the segmentation technique used for nuclei detection from the image. In other words, the global and local nuclear features (such as: color, shape, size, texture, etc.) provided by the performed segmentation technique must be sufficiently relevant and accurate to describe the nuclei. Indeed, segmentation of cell nuclei is a very hard task because of the high variability in their morphological and chromatic features (see Fig. 1). Nowadays, there are many promising segmentation methods presented in the literature which can overcome histological noise and give pertinent nuclear features to classify breast cancer. Active contours and watershed algorithms are frequently used for microscopic image segmentation [5,9–13].

Active contours, or snakes, have gained significant attention and become powerful tools used for edge detection, medical image segmentation and object tracking [14,15]. Snakes can be classified as either parametric or geometric active contours. Particularly,

* Corresponding author. Tel.: +216 98 920 394.

E-mail addresses: aymen_mouelhi@yahoo.fr (A. Mouelhi), mounir.sayadi@esstt.rnu.tn (M. Sayadi), fnaiech@ieee.org (F. Fnaiech), Karima.mrad@rns.tn (K. Mrad), khaledbenromdhane@rns.tn (K. Ben Romdhane).

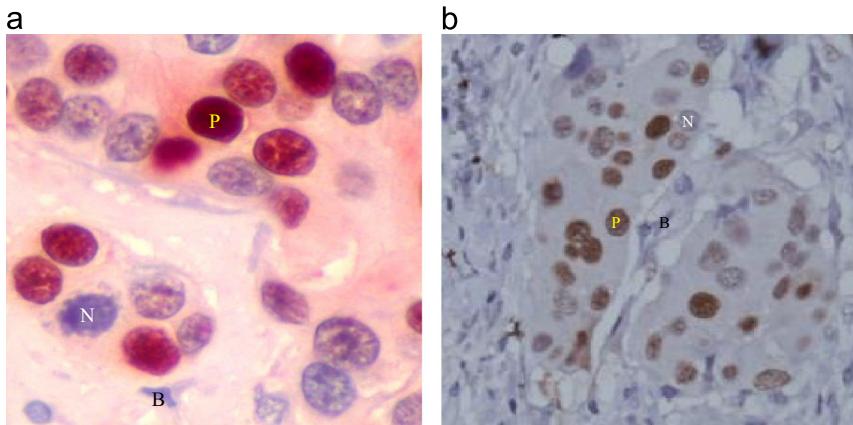


Fig. 1. Example of extracted regions from stained breast cancer tissue images: (a) stained breast cancer tissue with AEC chromogen using $\times 40$ magnification. (b) Stained breast cancer tissue with DAB chromogen using $\times 20$ magnification. P, N and B designate respectively positive staining for ER, negative staining for ER and benign nuclei. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

parametric active contours are explicitly expressed as parameterized curves in Lagrange formulation [16,17]. One deficiency, they have is sensitivity to initialization and lack of ability to handle changes in the topology of the evolving curve. Geometric active contours (GAC) were introduced more recently and are based on the theory of curve evolution and geometric flows [15]. Their numerical implementation is based on the level set method and allows an automatic segmentation of multiple objects simultaneously even in the noisy images [15]. However, convergence speed of GAC is slower than parametric snakes because of their computational complexity [11]. Besides, classical GAC models [16,19,20] need to compute a stopping function based on the gradient of the image, so they can detect only the external boundaries of objects. Other active contour models have been proposed recently to improve the segmentation accuracy and the convergence speed. Chan and Vese proposed an active contour model without edge (CV model) which is able to detect the interior and exterior borders of objects without using an edge stopping function [21]. Other improved active contour models which combine the classical GAC formulation and two-region segmentation model have been presented [22–24]. The Geodesic-Aided CV model is a typical example of hybrid active contours that combines GAC and CV models [22]. An enhanced geometric active contour model for color images was proposed by Ying et al. [23]. It utilizes directional information about contour position joined with an improved GAC including statistical information based two-region segmentation model. Pi et al. proposed a hybrid model combining the classical GAC and a modified CV model to detect the desired objects in color images [24]. A discrimination function based on principal components analysis (PCA) and interval estimation was applied to estimate desired object regions and it is used to compute the initial curve and the edge function. This guarantees that the deformed curve will stop near object boarders in reasonable time. There are other works that increase considerably the computational speed of the model and facilitate the level set manipulation [12,13]. They employ fuzzy clustering as a pre-segmenting step in order to relieve manual intervention and accelerate the level set optimization.

Watershed algorithm deals with overlapping and clustered object separation. It is widely studied and used for microscopic image segmentation. The major drawback of the classical watershed method is over-segmentation which is due to the presence of a multitude of regional minima consequence of noise [25]. Many techniques have been proposed to reduce the noise sensitivity of the algorithm such as marker-controlled watershed [26], region merging-watershed [27,34], and watershed method using prior shape [28]. Other recent techniques are presented in

the literature providing higher accuracy of separation results than traditional methods for diverse types of microscopic images. A combined method based on watershed and balloon snake was proposed by Dagher et al. [29]. Watershed algorithm is applied on the gradient image to construct initial balloon snakes that will deform to detect the object boundaries. For each balloon model, parameter optimization is required to ensure the best boundary localization. A significantly different segmentation technique is proposed in [30] to split clustered lymphocytes automatically. A concave vertex graph is first constructed from concave points and inner edges. Then, the optimal path in the graph, which separates clustered nuclei, is computed by minimizing a cost function based on morphological features. A hybrid image segmentation based on marker-controlled watershed algorithm was proposed by Cheng et al. [31]. The authors use the CV model to get an initial segmentation. Then, a marker-controlled watershed transform is applied on the segmented image using a new marking function based on shape markers.

In this paper, we propose a new automatic image analysis method for quantitative assessment of ER positive status in breast cancer sections. The proposed method includes both segmentation and classification of breast cell nuclei. First, all cell nuclei in the image are segmented using a new fuzzy active contour model and an improved watershed algorithm. Then, a multilayer neural network (MNN) classifier trained by Fisher's linear discriminant (FLD) features is used to recognize positive and negative stained nuclei for subsequent ER-status determination. The contribution of the present work is the proposal of an automatic segmentation scheme and the study of its ability to cooperate with FLD–MNN classifier in evaluating the ER status of breast cancer specimens. Indeed, the nuclear classification procedure proposed in our previous work [3] was based only on the color features attached to a single pixel sample belonging to the considered nuclei staining. Spatial information such as the values of the neighboring pixels and nuclei configurations were completely ignored. However, the nuclear classification process is heavily improved in this work thanks to the use of the proposed segmentation method and its accuracy in providing morphological and textural features of stained nuclei. The nuclear segmentation approach proposed in this work capitalizes on the new fuzzy active contour model in detecting stained nuclei and takes advantages of the improved watershed method [3] for clustered nuclei separation, which was not considered in earlier works. The entire image analysis method is evaluated on several breast cancer images of the database including different levels of malignancy. In addition, the performance of the method is verified when compared to state-of-art

segmentation and classification methods as well as to the manually truths given by experienced pathologists.

2. Material and methods

2.1. Acquisition of the image datasets and manual assessment of ER status

The studied images in this work are paraffin-embedded breast cancer tissue slides stained with an antibody against the ER biomarker in the nuclei. IHC staining is applied according to the LSAB method with 3-amino-9-ethylcarbazole (AEC) or 3,3' diaminobenzidine (DAB) chromogen. After IHC staining, different evaluation criteria may be conducted by pathologists to identify cancer in tissue specimens. These assessment measures vary according to the type of stains and the analyzed protein targets (i.e. nuclei, cellular membranes or cytoplasm) [1,5]. For example, membrane protein activity can be quantified on the basis of the membrane stain continuity [1]. For IHC stained tissue with nuclear activations, as in the case of our study, experts rely on other features such as color and intensity of the staining, shape and size of cell nuclei and nuclear density. In Fig. 1 for example, we show obviously that clustered nuclei and nuclear morphological variations are especially prominent in cancer cases. In addition, tumor cell nucleus can be identified by its large size and dark staining when it is compared to a benign nucleus. The positive nuclei (P) are marked by red color using the AEC chromogen (Fig. 1(a)) whereas they are marked by brown color with DAB chromogen (Fig. 1(b)). For both chromogens, negative nuclei (N) are stained by blue color (see Fig. 1). Benign nuclei (B) such as lymphocytes, plasma cell nuclei are also labeled with blue color but they are identified by their small size and pattern histologic arrangements [9] (Fig. 1). Several microscopic fields are taken with a light microscope NIKON E600 and CCD digital camera NIKON coolpix 995 using a magnifying factor of $\times 40$ (Fig. 1(a)) or $\times 20$ (Fig. 1(b)). A database of 120 breast cancer tissue images (2048×1360 pixels) is used to develop, train, and evaluate the proposed nuclear algorithm. Computer-assisted assessment of ER status in the collected image datasets is verified by three experienced pathologists. In fact, the role of pathologists is essential in finding the tumoral tissue to ensure a correct IHC evaluation of the specimen. In malignant samples only evaluation of the tumoral foci are acceptable, so appropriate finding of the malignant tissue is mandatory before IHC analysis. Proper focusing in the site of evaluation is generally done by experts after a gross examination which is important since the pathologist often sees tissue features that suggest cancer. It helps also the expert to decide which parts of a large biopsy are the most critical to study under a microscope. For each stained specimen, the ER positive status is determined by experts based on visual inspection of the staining intensity and a direct count of the proportion of positive nuclear staining in the critical regions of the malignant tissue. Then, manual assessments of ER status and immunohistochemical analysis are combined in a panel consensus which is considered as a gold standard for methods' evaluation. Table 1 shows the scoring system [8] used in this work, where the proportion of positive staining is computed by the ratio of the number of positive stained nuclei to the total number of positive and negative nuclei in the whole image.

Our image database is divided into two sets according to the type of the reaction products (AEC or DAB). Dataset 1^(AEC) contains 36 images having an expert's positivity score 1+(10/36), 2+(20/36) and 3+(6/36). Dataset 2^(DAB) comprises 84 images having an expert's positivity score 1+(32/84), 2+(42/84) and 3+(10/84). The studied image database is obtained by a stratified random sampling from the whole patient population containing 480 tissue

Table 1
The scoring system used for ER positive status evaluation.

Score	Intensity staining	Proportion of positive nuclear staining (%)
0	Negative staining	0–10
1+	Weak positive staining	11–33
2+	Moderate positive staining	34–66
3+	Strong positive staining	67–100

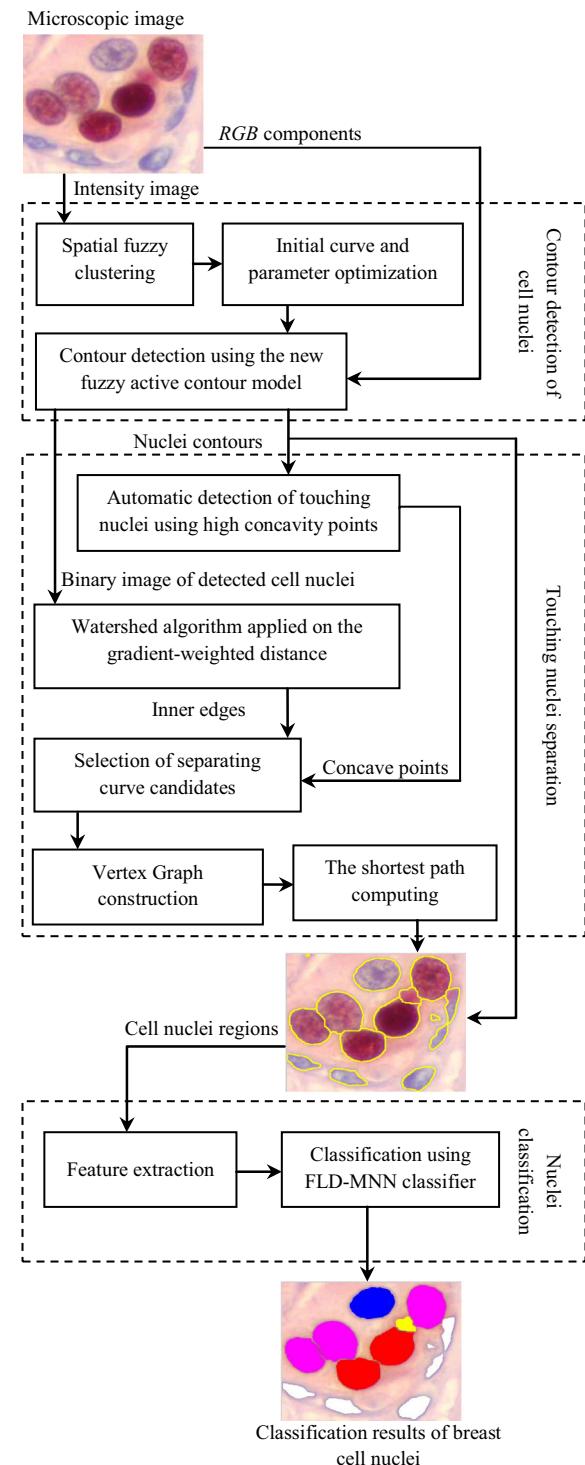


Fig. 2. Flow chart of the proposed image analysis method for automatic assessment of ER status.

images. To reduce the sampling fluctuations and ensure a good representativeness of the image database, each sample is selected independently from each image group in proportion to that group's depiction in the whole population.

2.2. Description of the proposed method

As illustrated in Fig. 2, the flow chart shows that the proposed method is composed mainly of three steps: Contour detection of cell nuclei, touching nuclei separation and nuclei classification. As a first step, a color geometric active contour model incorporating a spatial fuzzy clustering algorithm is proposed to extract the nuclei contours in the breast tissue image. Similarly to the method of Li et al. [13], a spatial fuzzy clustering algorithm is applied to estimate nuclei regions. However, compared to the previous work [13], we propose to include Bayes error based two-region segmentation energy functional [23] to the fuzzy level set method [13] in order to get accurate nuclei detection in the breast tissue images. In fact, Li et al. [13] have used the classical level set method which uses only local information given by curvature and the edge stopping function to detect the object boundaries. The classical level set method neglects the region information, therefore it is unable to provide accurate results in fuzzy and discrete edges especially when applied to noisy images. So, it would be attractive to add the color region information on the fuzzy level set method in order to improve the segmentation accuracy in color images. Besides, the controlling parameters and the initial curve of the proposed model are optimized automatically from the clustering results. In the second step, we extract the touching nuclei regions from the pre-segmented image using high concavity points of nuclei contours. Next, we apply the watershed method to the hybrid distance transformed image. To separate touching nuclei, high concavity points are used to locate separating curve candidates. Then, a concave vertex graph is constructed from concave points and separating edges. The optimal separating curves are selected by computing the shortest path in the graph using Dijkstra algorithm. Finally, a trained multilayer neural network is used to classify FLD features obtained from the segmented nuclei regions. In this work, stained nuclei are classified into five categories according to their textural and morphological

characteristics in order to recognize the cancer affected nuclei in the breast tissue images. The major steps of the proposed method are presented in the following subsections.

2.3. Contour detection of cell nuclei

2.3.1. Spatial fuzzy clustering

It is known that the Fuzzy C-Means (FCM) algorithm is one of the unsupervised techniques that has been successfully applied to clustering, feature analysis, classifier design, target recognition and image segmentation [13,32,33]. The algorithm is an iterative optimization that minimizes the objective function for the partition of dataset, $x = [x_1, x_2, \dots, x_n]^T$, defined as follows:

$$J_m(u, v) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \|x_j - v_i\|^2, \quad (1)$$

where x represents the gray level of an image with n pixels to be partitioned to c clusters (or classes). u_{ij} denotes the membership of pixel x_j in the i th class, v_i is the i th class center and m is a weighting factor that controls the fuzziness of the membership function. The cost function is minimized when pixels close to their class center are assigned high membership values, while those that are far away are given low values. The membership functions and the centroids are updated iteratively by the following:

$$u_{ij} = \left(\sum_{k=1}^c \left(\frac{\|x_j - v_k\|}{\|x_j - v_i\|} \right)^{2/(m-1)} \right)^{-1}, \quad (2)$$

and

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (3)$$

The membership functions are constrained to be in the range $[0,1]$ such that $\sum_{i=1}^c u_{ij} = 1$. Starting with an initial estimation for each centroid, FCM algorithm converges to a solution for v_i representing the local minimum of the cost function.

The major drawback of the standard FCM algorithm in imaging applications is the lack of spatial information. Noise and image artifacts can easily damage the performance of FCM segmentation. So, integrating spatial information in FCM algorithm generates many active researches [32,33]. Chuang et al. [32] proposed a

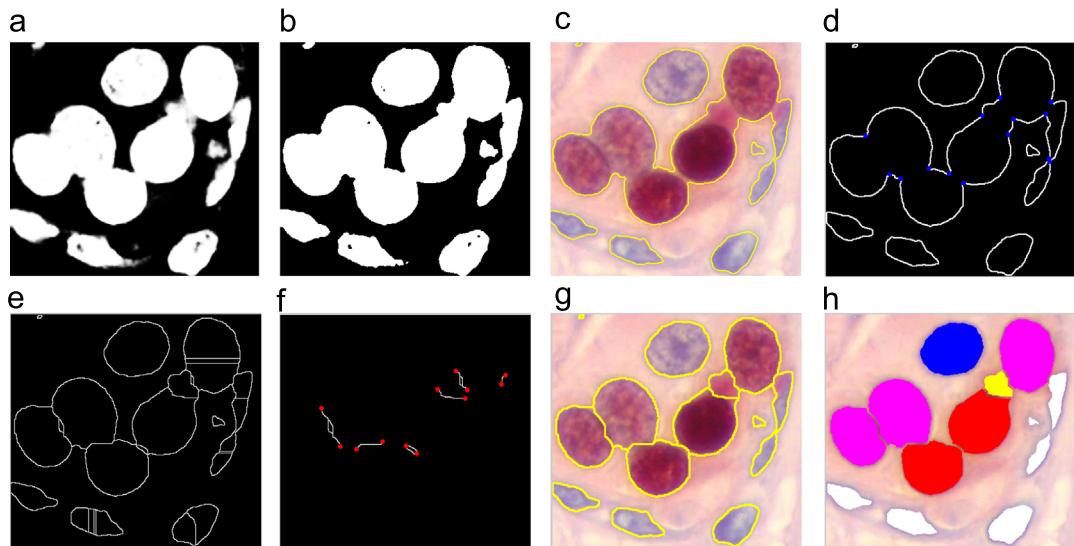


Fig. 3. Different steps of the proposed image analysis method: (a) initial segmentation by spatial FCM, (b) initialization by thresholding, (c) final cell nuclei contours, (d) clustered nuclei detection using of high concavity points, (e) watershed algorithm applied on the distance image, (f) selection of the nearest inner edges to concave points and vertex graph construction, (g) segmentation results of all cell nuclei, (h) classification results of breast cell nuclei: the FLD-MNN classifier marks the segmented nuclei with a color mask (benign nuclei – white; negative nuclei (0) – blue; positive nuclei: (1+) – yellow, (2+) – magenta, (3+) – red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

spatial FCM which incorporates the spatial information into membership functions directly as follows:

$$u'_{ij} = \frac{u_{ij}^p h_{ij}^q}{\sum_{k=1}^c u_{kj}^p h_{kj}^q}, \quad (4)$$

where p and q are parameters that control the importance of the respective contribution. The spatial information is included by the function h_{ij} defined by

$$h_{ij} = \sum_{k \in NB(x_j)} u_{ik}, \quad (5)$$

where $NB(x_j)$ denotes a local window centered around a pixel x_j . The spatial membership u'_{ij} and the cluster centers v_i are updated using Eqs. (2) and (3). To detect the significant objects in the image using spatial FCM algorithm, we took a sliding window size 9×9 , $c=m=2$ and $p=q=1$. Fig. 3(a) shows segmentation results of cell nuclei using spatial FCM algorithm.

2.3.2. The proposed fuzzy active contour model

To detect object boundaries in a color image with the classical geodesic active contour, a curve C should be found which minimizes the following energy functional:

$$E(C) = \int_0^{L(C)} g_{col}(C(q))ds + v \int_{\Omega} g_{col}(C(q))dA \quad (6)$$

where $\Omega \subset \mathbb{R}^2$ is the 2-D domain of a color image. $C(q) = (x(q), y(q))$ is a differentiable closed curve in Ω with a position parameter $q \in [0, 1]$. The distance ds is the Euclidian metric, dA is the element of area and v is a fixed parameter that control the area of the curve. g_{col} is a stopping function for color images. It should be strictly positive in homogeneous regions and close to zero on the edges [24]. A typical stopping function is proposed [19]:

$$g_{col}(x, y) = \frac{1}{1 + \alpha \Lambda^2} \quad (7)$$

where α is a contract factor and Λ is the largest eigenvalue of the metric tensor g_{ij} given by

$$g_{ij} = \begin{bmatrix} 1 + R_x^2 + G_x^2 + B_x^2 & R_x R_y + G_x G_y + B_x B_y \\ R_x R_y + G_x G_y + B_x B_y & 1 + R_y^2 + G_y^2 + B_y^2 \end{bmatrix} \quad (8)$$

with R_x , G_x and B_x are respectively the gradient according the direction x of the Red, Green and Blue components after Gaussian convolution, i.e. $R_x = \partial(G_\sigma * I_R) / \partial x$, $G_\sigma(x, y) = \sigma^{-1/2} \exp(-|x^2 + y^2|/4\sigma)$ and σ is the smoothing parameter.

The general curve evolution along its normal direction \vec{N} with speed F , is $\partial C / \partial t = F \cdot \vec{N}$. The evolution speed is generally a function of the image gradient and the curve characteristics [12]. The numerical solution of the curve evolution problem is obtained by the fast level set formulation [20]:

$$\begin{cases} \frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi) \left(\lambda \operatorname{div} \left(g_{col} \frac{\nabla \phi}{|\nabla \phi|} \right) + v g_{col} \right) + \mu \operatorname{div} \left(\left(1 - \frac{1}{|\nabla \phi|} \right) \nabla \phi \right) \\ \phi(x, y, t=0) = \phi_0(x, y) \end{cases} \quad (9)$$

where ϕ is a level set function defined on Ω , whose zero-level set $C = \{(x, y) / \phi(x, y) = 0\}$ defines the boundary contour of the objects such that $\phi > 0$ inside the segmented objects and $\phi < 0$ outside. μ is a positive parameter that control the penalization term $\operatorname{div}((1 - 1/|\nabla \phi|) \nabla \phi)$ which is used to keep the level set function close to a signed distance function during the evolution. λ is a constant controlling the curvature term, $\phi_0(x, y)$ is the initial contour and δ_ϵ denotes the regularized Dirac function obtained as follows:

$$\delta_\epsilon(z) = \frac{1}{\pi} \left(\frac{\epsilon}{\epsilon^2 + z^2} \right) \quad (10)$$

where ϵ is a positive constant regulating the Dirac function [20,21]. Finally, the discrete scheme of the level set function is given by

$$\phi^{k+1}(x, y) = \phi^k(x, y) + \tau L(\phi^k(x, y)) \quad (11)$$

where $L(\phi^k(x, y))$ is the approximation of the right-hand side in (9) and τ is the time step. We note that all the spatial derivatives $\partial \phi / \partial x$ and $\partial \phi / \partial y$ can be approximated by the central difference, since the upwind scheme [18] in the traditional level set methods is no longer needed due to the diffusion proprieties of the penalization term [20].

The fuzzy level set method proposed in Li et al. [13] consists on constructing the initial curve from the clustering results. Then, the parameters that control and regularize the level set evolution (i.e. λ , μ , v and τ) are also derived from the results of fuzzy clustering directly. The initial level set equation is given by

$$\phi_0(x, y) = 4\epsilon(B_k - 0.5) \quad (12)$$

where B_k denotes the binary image obtained by the thresholding of the membership degree at each pixel u_k corresponding to the region of interest in the image and noted R_k :

$$B_k = \begin{cases} 1 & \text{if } R_k \geq b_0 \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

where b_0 is an adjustable threshold which is used to estimate interesting regions in the image. In this work b_0 is fixed at 0.5 (Fig. 3(b)) and we took $\epsilon=1$ to compute the regularized Dirac function δ_ϵ and the initial contour ϕ_0 . The controlling parameters associated with the level set method are determined adaptively from the initial level set function ϕ_0 . It is noticeable that level set evolution will be faster if the object of interest is large. So, the time step τ is assigned to the ratio a/ℓ , since it will be big for large objects:

$$\tau = \frac{a}{\ell} = \frac{\int_I H_\epsilon(\phi_0) dx dy}{\int_I \delta_\epsilon(\phi_0) dx dy} \quad (14)$$

where a and ℓ represent respectively the area and the length of the estimated regions. $H_\epsilon(\phi)$ is the regularized Heaviside function such that $\delta_\epsilon(\phi) = H'_\epsilon(\phi)$. The parameter λ is set to be $0.1(a/\ell)$ and μ is equal to $0.2(a/\ell)^{-1}$.

An enhanced balloon force V is proposed in Li et al. [13] to determine the advancing direction of the level set function using the membership degree according to the specific object of interest R_k :

$$V(R_k) = 1 - 2R_k \quad (15)$$

The new balloon force such that $-1 \leq V(R_k) \leq 1$ is a matrix in which the sign of value at each image pixel defines the way of the level set evolution: positive for shrinkage and negative for expansion.

In the new fuzzy active contour model, we propose to include statistical information based two region segmentation model into the fuzzy level set method using the following energy functional [23]:

$$E(\phi) = - \int_{\Omega} (H_\epsilon(\phi) \log(P_1) - (1 - H_\epsilon(\phi)) \log(P_2)) dx dy \quad (16)$$

where P_1 and P_2 are the probability densities of the regions inside and outside the curve C respectively. They are chosen Gaussian functions. These Gaussian functions are computed here with the color components of an image as they convey more statistical information about homogeneous regions than density functions evaluated in gray-level images:

$$P_{ij}(I_j | (\sigma_{ij}, \mu_{ij})) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left(-\frac{(I_j - \mu_{ij})^2}{2\sigma_{ij}^2}\right) \quad (17)$$

with $i=1,2$ corresponding to the probability density inside and outside the curve and $j=1,2,3$ associated respectively to the RGB components of the image I . The mean μ_{ij} and the standard deviation σ_{ij} are computed as

$$\mu_{1j}(\phi) = \frac{\int_{\Omega} I_j H_e(\phi) dx dy}{\int_{\Omega} H_e(\phi) dx dy}, \quad \mu_{2j}(\phi) = \frac{\int_{\Omega} I_j (1 - H_e(\phi)) dx dy}{\int_{\Omega} (1 - H_e(\phi)) dx dy},$$

$$\sigma_{1j}(\phi) = \sqrt{\frac{\int_{\Omega} (I_j - \mu_{1j})^2 H_e(\phi) dx dy}{\int_{\Omega} H_e(\phi) dx dy}}, \quad \sigma_{2j}(\phi) = \sqrt{\frac{\int_{\Omega} (I_j - \mu_{2j})^2 (1 - H_e(\phi)) dx dy}{\int_{\Omega} (1 - H_e(\phi)) dx dy}}$$

Then the probability distribution functions are defined as

$$P_1 = \frac{1}{3} \sum_{j=1}^3 P_{1j}(I_j | (\sigma_{1j}, \mu_{1j})), \quad P_2 = \frac{1}{3} \sum_{j=1}^3 P_{2j}(I_j | (\sigma_{2j}, \mu_{2j})) \quad (18)$$

The minimization of the energy (16) with respect to ϕ according to the gradient descent rule is given by

$$\frac{\partial \phi}{\partial t} = \delta_e(\phi) \log \left(\frac{P_1}{P_2} \right) \quad (19)$$

Finally, the formulation of the new fuzzy active contour model for color images is obtained as follows:

$$\frac{\partial \phi}{\partial t} = \left| \nabla \phi \right| \left(\gamma \log \left(\frac{P_1}{P_2} \right) + \lambda \operatorname{div} \left(g_{col} \frac{\nabla \phi}{|\nabla \phi|} \right) + V(R_k) g_{col} \right) + \mu \operatorname{div} \left(\left(1 - \frac{1}{|\nabla \phi|} \right) \nabla \phi \right) \quad (20)$$

where $\delta_e(\phi)$ is replaced here by $|\nabla \phi|$ in order to intensify the effective range in the whole image [23]. $\lambda, \mu, V(R_k)$ and the initial curve ϕ_0 are computed as defined below. γ is a positive parameter which controls the region term. It should be mentioned here that γ takes a larger value when the image is noisy. In our work, the parameter γ is also estimated adaptively from the clustering results and it is set to be equal to $0.1(a/\tau)$ so that the region term has the same influence as the curvature information in the model. The level set function is updated as usual according to Eq. (11) where $L(\phi^k(x, y))$ is the approximation of the right-hand side in the evolution Eq. (20). The main steps of the proposed fuzzy active contour model are summarized in Table 2. The final cell nuclei contours using the proposed model are represented in Fig. 3(c).

2.4. Touching nuclei separation

2.4.1. Touching nuclei detection using high concavity points

Touching and overlapping cell nuclei can be identified by the existence of concavity points on the nuclei contours. So, detection of high concavity points on the nuclei boundaries is an important step in the separation process [35–37]. In our work, the concave points in the nuclei boundaries are selected using a robust corner detector technique based on the global and local curvature features of the contour [38]. Other algorithms may be used for

Table 2

Algorithm of the new fuzzy active contour model for detection of nuclei contours.

-
- Input: Given the RGB image
1. Classify pixels of the gray level image into 2 classes using the spatial fuzzy clustering algorithm
 2. Initialize $\alpha = 50$, $\sigma = b_0 = 0.5$ and $\epsilon = 1$.
 3. Construct the initial contour $\phi^{k=0} = \phi_0$ by (12)
 4. Compute the active contour parameters $\lambda, \mu, \gamma, V, \tau$ and the edge function g_{col} by (7).
 5. For a given level set function ϕ^k , compute the probability density functions P_1 and P_2 using (17) and (18).
 6. Update ϕ^{k+1} using (11) and (20).
 7. Check whether the solution is stationary. If not, $k = k + 1$ and repeat from 5.
-

detecting concave points [39,40]. We suggest that by using local and global curvature proprieties more accurate results and robust to noise can be reached (Fig. 3(d)).

Suppose that the j th extracted contour is represented by $C^j = \{p_1^j, p_2^j, \dots, p_n^j\}$, where $p_i^j = (x_i^j, y_i^j)$ are points on the contour, x_i^j, y_i^j are the coordinates of i th point in the j th contour. First, the curvature of each point on the contour j is computed using the following equation:

$$K_i^j = \frac{\dot{x}_i^j \ddot{y}_i^j - \dot{y}_i^j \ddot{x}_i^j}{((\dot{x}_i^j)^2 + (\dot{y}_i^j)^2)^{3/2}}, \quad (21)$$

where $(\dot{x}_i^j, \dot{y}_i^j), (\ddot{x}_i^j, \ddot{y}_i^j)$ are respectively the first and second derivatives of the point which are computed by convoluting the boundary with Gaussian derivatives, i.e. $\dot{x}_i^j = x_i^j * G_{\sigma_c}$ and σ_c is the width of the Gaussian function. Second, the initial list of corner candidates is constructed by points corresponding to local maxima of the curvature function. Detected points of maximum curvature correspond to both concave and convex points. To keep only concave points, we suppress points which have positive curvature from the initial corner list (Fig. 4(a)). Third, round corners are removed from the list using an adaptive threshold T of absolute curvature. In fact, a region of support (ROS) is defined for each concave point candidate by the contour portion bounded with the point's two nearest curvature minima in order to distinguish round concave points from the obtuse ones (Fig. 4(b)). The adaptive threshold is given by

$$T(u) = R \times \bar{K} = R \frac{1}{L_1 + L_2 + 1} \sum_{i=u-L_2}^{u+L_1} |K(i)|, \quad (22)$$

where R is a coefficient that define round concave points to be suppressed, \bar{K} is the mean curvature of the ROS, u is the position of the concave point candidate on the contour and $L_1 + L_2$ is the size of the ROS centered at u . Round concave points are eliminated from the list if the curvature of the point candidate is lower than $T(u)$. Finally, false concave points due to quantization noise and trivial details are removed by evaluating the angles of concave point candidates in a dynamic region of support [38]. The angle of the considered point is determined by constructing the tangent direction of the ROS from the two sides of the concave point candidate (Fig. 4(c)). A true concave point should satisfy the following criterion [38]:

$$\text{The concave point is } \begin{cases} \text{true} & \text{if } \theta_c \leq \theta_{obtuse}, \\ \text{false} & \text{if } \theta_c > \theta_{obtuse}. \end{cases} \quad (23)$$

where θ_{obtuse} is a parameter designating the maximum obtuse angle that a concave point candidate can have and still be considered as a true concave point. To detect high concavity points on the nuclei boundaries, we have chosen $\sigma_c = 5$, $R = 1.5$ and $\theta_{obtuse} = 155^\circ$ in all experiments presented in this work (Fig. 4(d)).

2.4.2. Watershed algorithm

In our work, we have used shape information about clustered nuclei in a simple way to refine watershed results [3,41]. In fact, high concavity points on the nuclei contours are used to locate the most significant separating edges. Then, a post processing step is applied to obtain the optimal separating edges inside nuclei. To get initial separating curves of clustered cell nuclei, we have applied the watershed algorithm by immersion [25] on a gradient-weighted distance transform. The distance transform combines two image transformations: the geometric distance transform and the intensity gradients transform [34]. It takes into consideration nuclear configurations and color gradient information within nuclei. This distance can overcome the limitations of both geometrical and gradient transforms. On the one hand, the geometric distance is only good at dealing with regular shapes.

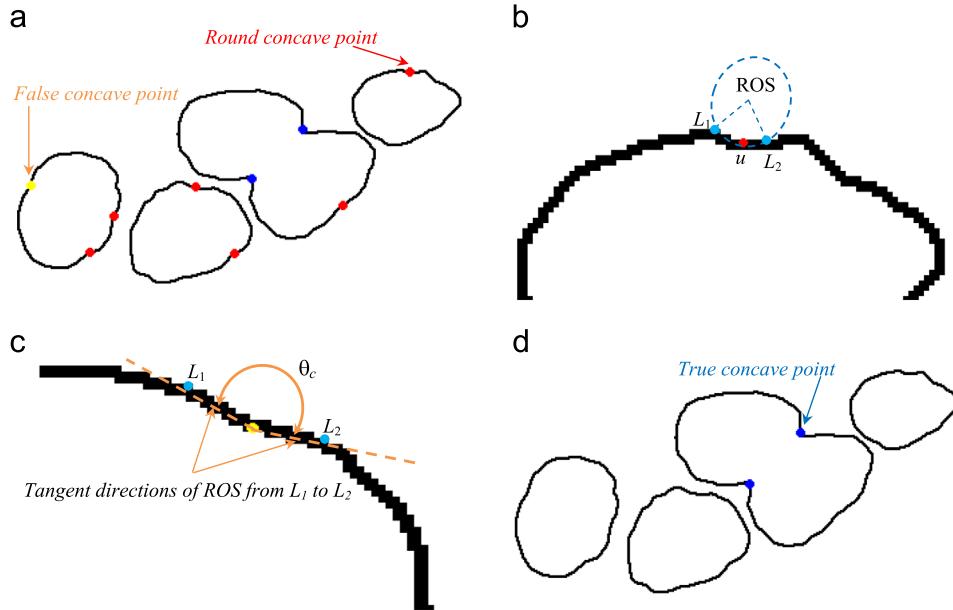


Fig. 4. Concave points detection: (a) initial concave point candidates, (b) region of support (ROS) of a round concave point, (c) evaluation of the angle of a concave point candidate and (d) true concave points.

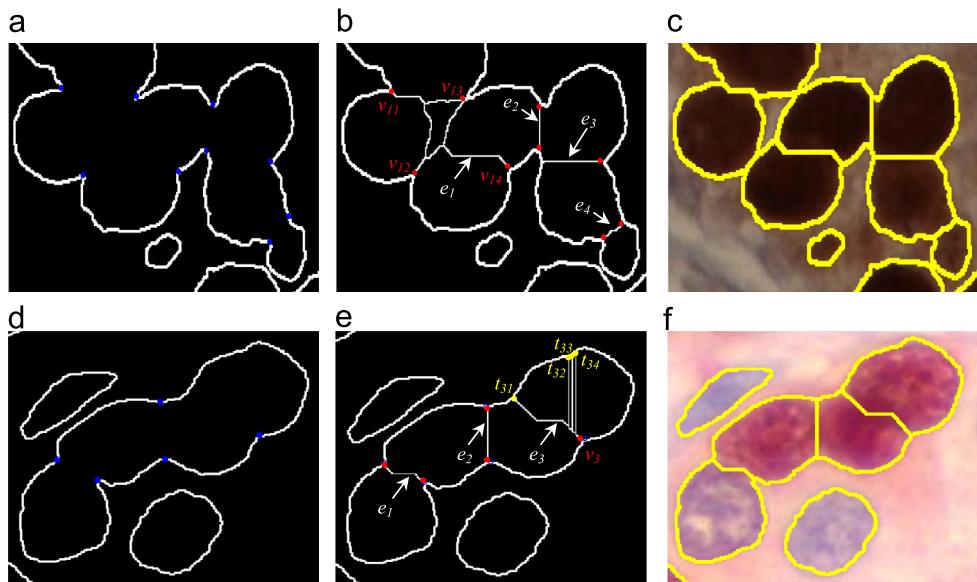


Fig. 5. Segmentation results of some complex cell nuclei configurations: (a), (d) concave point detection. (b), (e) Construction of the concave vertex graph Gp. (c), (f) Optimal path selection to separate touching nuclei.

On the other hand, the gradient transformation is sensitive to imaging noise, and usually results in over-segmentation [34]. The geometric distance D and the color gradient transform Λ are combined in a single representation given by the following expression [34]:

$$D' = D \times \exp\left(1 - \frac{\Lambda - \Lambda_{\min}}{\Lambda_{\max} - \Lambda_{\min}}\right) \quad (24)$$

Here D is the Euclidean distance computed over the image of the binary image of the pre-segmented nuclei and Λ is the gradient transform of the entire image represented in the RGB space. Λ_{\min} and Λ_{\max} are the minimum and maximum values of the color gradient Λ needed for normalization. Watershed algorithm is then

applied on the inverse of the distance transformation D' . This inverse is denoted S , and is computed as follows:

$$S = G_{\sigma_s} * (\max(D') - D') \quad (25)$$

where G_{σ_s} is a Gaussian smoothing operator used to reduce image noise with width controlled by σ_s . Initial cell nuclei separation using watershed algorithm is shown in Fig. 3(e), where σ_s is set to be 2.5. The choice of the smoothing parameter and its influence on the separation results are discussed in Section 3.

2.4.3. Concave vertex graph construction and shortest path selection

The segmentation results obtained after the watershed step consists of the outer boundary of clustered nuclei and the inner

edges which represent the initial separating edge candidates. To select the most significant inner edges, we proceed by searching the nearest end points of these edges to the concave vertices (see Fig. 3(f)). In fact, a local window (11×11) centered on the concave point is used to detect the minimal Euclidean distance between end points and concave vertices. Only the edges nearest to concave points are kept in the list of separating curve candidates. Then, a concave vertex graph G_p is constructed from the vertex set of selected inner edges E (white lines in Fig. 3(f)) and the set of end points V (red vertices in Fig. 3(f)). The vertices of graph G_p are then equal to $V \cup E$. To select the optimal separating curves of touching nuclei, we have applied Dijkstra algorithm [42] to compute the shortest path between the terminal vertices for each edge in the graph. The final separation results are obtained by combining the outer boundary of cell nuclei and the shortest paths (Fig. 3(g)).

Some segmentation results of complex cell nuclei configurations are illustrated in Fig. 5. In the first case, we show that the detected end points in the graph can be more than two vertices (see the first row of Fig. 5). As we can see in Fig. 5(b), the inner edge e_1 contains four terminal points (v_{1k} , where $k \in [1,4]$). In this case, all the shortest paths between the points v_{1k} are iteratively computed using the Dijkstra algorithm. In the second row of Fig. 5, the graph contains three separated inner edges (e_i ; $i = 1, 2, 3$) with their terminal points. We note that e_3 is selected by only one concave point in the nuclei boundaries and it is clear that this inner edge has other end points in the upper part of the cell nuclei (see Fig. 5(d) and (e)). So, to obtain the optimal separating curve in this case, we extract first all the terminal points from the inner edge (i.e. yellow points t_{3k} , where $k \in [1,4]$ in Fig. 5(b)). Then, the Dijkstra algorithm is used to find the shortest path p_k between the nearest concave point v_3 and terminal points t_{3k} . Finally, the optimal path is selected by minimizing a cost function that ensures the detection of enclosed regions which have a good homogeneous intensity [43]. For each path the minimum of two saliency costs is computed by

$$c_s = \min\left(\frac{|p_k| + \lambda \iint_{R_L} I(x,y) dx dy}{\iint_{R_L} dx dy}, \frac{|p_k| + \lambda \iint_{R_R} I(x,y) dx dy}{\iint_{R_R} dx dy}\right) \quad (26)$$

where $|p_k|$ is length of path that separates the enclosed contour C into two regions R_L and R_R . The term $\iint_R I(x,y) dx dy$ measures the total color variation inside the enclosed region and reflects the intensity homogeneity in region R [43]. The normalization over the area $\iint_R dx dy$, enclosed by C and the shortest path p_k , is used to avoid small regions detection resulting from image noise. λ is a blending parameter which is chosen 0.5 in this work. The optimal path for separating two homogenous cell nuclei is the path for which the cost function c_s is minimal (Fig. 5(f)). The main steps of the proposed separation method for touching cell nuclei are summarized in Table 3.

2.5. Nuclei classification and automatic assessment of ER status

2.5.1. Nuclei feature extraction and FLD preprocessing

In this work, the FLD-MNN classifier [44] is designed to identify cancer nuclei and to distinguish between different types of staining. There are many classifiers that can be used for pattern recognition. However, it has been proven that MNN is a powerful tool for several classification tasks [44,45]. To get accurate classification results of the segmented cell nuclei, we have chosen the most discriminant nuclear features to construct the training dataset of the network. For classification purposes, we used four morphological features and nine textural features of 2882 sample stained nuclei which are extracted from 80 training images labeled by pathologists, the remaining IHC images (40 images) are kept for the test. Indeed, to select the optimal number of the learning images, we performed five experiments using different image sets

Table 3

The improved watershed algorithm for touching cell nuclei separation.

-
- Input: Given the pre-segmented image by the new fuzzy active contour model
1. Extract the nuclei contours C and construct the binary image of detected cell nuclei regions.
 2. Initialize $\sigma_c = 5$, $R = 1.5$, $\theta_{obtuse} = 155^\circ$, $\sigma_s = 2.5$, $cost = \emptyset$ and $paths = \emptyset$.
 3. Detect the concave vertices on the contour C .
 4. Apply the watershed algorithm on the inverse of the hybrid distance transform using (24) and (25).
 5. Extract inner edges and find end points for each one of them.
 6. Select the separating edge candidates (E) by searching the end points nearest to concave vertices (V).
 7. For each edge candidate $e_i \in E$
 - If (the number of end points $v_i \geq 2$)
 - Construct the concave vertex graph $G_p = (v_i \cup e_i)$
 - For each vertex $v_i(k)$
 - For each vertex $v_i(l)$ ($l \neq k$)
 - Find the shortest path between $v_i(k)$ and $v_i(l)$ and save in $paths$.
 - Else if (the number of end points $v_i = 1$)
 - Search end points t_i in the edge such that $t_i \neq v_i$
 - Construct the concave vertex graph $G_p = (v_i \cup t_i \cup e_i)$
 - For each vertex $t_i(k)$
 - Find the shortest path between v_i and $t_i(k)$
 - Apply the path $p(k)$ to separate the enclosed contour C into two regions R_L and R_R .
 - Compute the cost function c_s by (26) and save in $cost$.
 - Pick up the path with the lowest c_s and save in $paths$.
 - Else return $paths$.
 - 8. Add $paths$ to nuclei contours C to obtain final separation results.
-

for training (20, 40, 60 and 80 images). In each experiment, the network is trained with an image set and the average validation error is reported using a five-fold cross-validation method [45]. The optimal number of image dataset used in the learning process, corresponds to the minimal validation error over the training image datasets. The images involved in the training set are different from those in the test set and they belong to different anonymous patients.

The morphological features are the perimeter, the area, the convexity and the eccentricity, which are usually used by pathologists to describe variations in size and shape of cancer nuclei [48]. Textural features represent the nuclear chromatic changes in the image. They are expressed in this study by the mean, the variance and third order moment in each color channel (Red, Green and Blue) of the segmented cell nuclei [14]. Then, FLD analysis is used to reduce the dimensionality of the training dataset and to produce independent components which are generally more suitable for classification purposes than PCA or original features. The new input training data is composed of $c - 1$ FLD features which are obtained by maximizing the ratio of between-class variation to within-class variation [47], where c is the number of cell nuclei classes. In our study, cell nuclei are divided into five categories: benign nuclei, negative stained nuclei (0) and three degrees of positive stained nuclei (1+, 2+ and 3+). The MNN structure is designed during the training experiments. In addition to the input layer which is composed of the four FLD features, we used a single hidden layer and one output layer. The optimal number of hidden nodes is selected using five-fold cross-validation technique [45] and the output layer consists of five neurons which correspond to the number of the desired nuclei classes. For the MNN training, the target values are either 1 (desired nuclei class) or -1 (non-desired nuclei class). The tangent sigmoid transfer function is used for all neurons. To accelerate training, the network is trained with Davidon Fletcher Powel (DFP) [46] algorithm until the maximum number of iterations is reached or the mean square error is less than 10^{-5} .

2.5.2. Nuclei classification using FLD–MNN classifier and automatic assessment of ER status

When the training process is completed and the optimal MNN structure is selected for the input data, the last synaptic weights of the network are saved to be ready for the testing procedure.

After the nuclei segmentation phase, FLD features can be easily computed from the segmented nuclei regions. Then, they are presented to the trained network in order to classify them into the five nuclei classes. In Fig. 3(h), we show the classification results of breast cell nuclei, where the FLD–MNN classifier marks

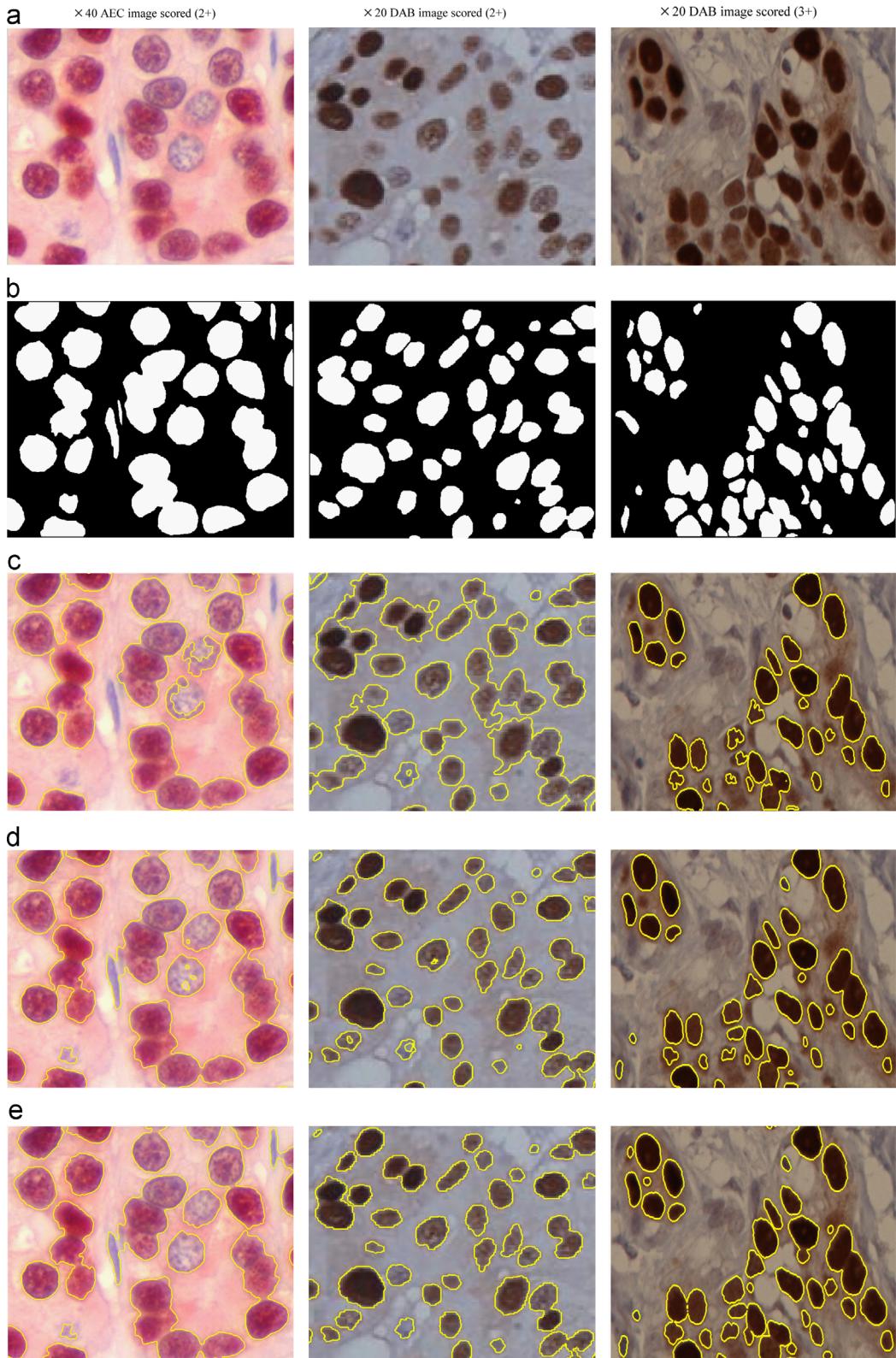


Fig. 6. Contour detection results of all breast cell nuclei: (a) the original image and (b) the ‘ground-truth’ regions of all cell nuclei constructed manually by pathologists. Contour detection results by (c) MGAC model [24], (d) fuzzy level set method [13], and (e) the proposed active contour model.

the segmented nuclei with a color mask (Benign nuclei – white; Negative nuclei (0) – blue; Positive nuclei: (1+) – yellow, (2+) – magenta, (3+) – red). For automatic assessment of ER status, the number of positive stained nuclei is computed automatically from the labeled image. Then, the ER-score of the biopsy image is evaluated by calculating the percentage of positive nuclei to the total number of positive and negative nuclei. That percentage is referred to the scoring scale in Table 1 (Section 2.1).

3. Results and discussion

The proposed method is implemented on a computer with Intel Core 2.1 GHz CPU, 4 GB RAM and it is applied on each image of the image database. To demonstrate the efficiency of the proposed segmentation algorithm, experimental results are presented in each step and compared with different existing segmentation and classification methods. In addition, results in all experiments are verified with those assessed by experienced pathologists.

3.1. Contour detection results of the new fuzzy active contour model

In this section, contour detection results by the new fuzzy active contour model are compared with ones from the state-of-the-art active contour-based segmentation techniques, such as the fuzzy level set method [13] and the modified geodesic active contour model (MGAC) [24]. Both fuzzy active contour models are initialized by the same initial curve and controlling parameters using spatial FCM algorithm. However, the initial curve and the stopping function in the MGAC model are constructed with a discrimination function which provides initial location of the desired objects. This discrimination function is obtained after statistical analysis of a pixel set chosen from the desired objects using Principal Components Analysis (PCA) and interval estimation [49]. In all experiments, we have chosen 150 sample pixels from stained cell nuclei in different breast tissue images. In Fig. 6(a) and (b) we show the original image and the ‘ground-truth’ regions of all cell nuclei constructed manually by experts, respectively. Fig. 6(c)–(e) show respectively the results of the nuclei contour detection obtained with the MGAC model, the fuzzy level set method and the proposed active contour model. From these results, it is clear that new model based on statistical region information is more efficient than the fuzzy level set method and the MGAC model especially for discrete and weak nuclei edges. As illustrated in Fig. 6(c), the MGAC model can fail to detect precisely all types of nuclei in the image since contextual information (the values of the neighboring pixels) is completely ignored during the discrimination function construction of the desired nuclei. It is true that this model includes the region information of the segmented objects introduced by the Chan-Vese energy functional, but it cannot distinguish between two regions which contain Gaussian noise with the same mean. In Fig. 6(d), we can see obviously that the fuzzy level set method is unable to give precise contours for breast cancer cell nuclei with low contrast. This is due to the weak estimation of the nuclei regions using the spatial FCM algorithm. Besides, the absence of the region term in the fuzzy level set method makes the distinction between two regions containing Gaussian noise very difficult. However, statistical information (the standard deviation and the mean of the segmented objects) introduced in the proposed model makes the curve evolve towards the real nuclei boundaries quickly and provides an accurate segmentation of homogeneous regions even in the noisy cases (Fig. 6(e)).

In order to evaluate the performance of the contour detection methods, all active contour models were applied on a testing image dataset composed of 40 images chosen from the three

expression levels (1+, 2+ and 3+) existing in the whole image database. Two automatic evaluation measures are used in this framework: region accuracy measure and contour distance. The region measure is used to provide the difference between the ‘ground-truth’ regions MR (constructed manually by experts) and the contour detection results CR [5,11]. The region segmentation accuracy is defined as

Region accuracy (%)

$$= 100 \times \left[1 - \frac{[A(CR) \cup A(MR)] - [A(CR) \cap A(MR)]}{[A(CR) \cup A(MR)]} \right], \quad (27)$$

where $A(CR)$ and $A(MR)$ are the areas of segmented regions CR and MR . Another useful measure based on the figure of merit of Pratt is applied in this work to provide a complementary information about segmentation quality. This measure uses the distance between all pairs of corresponding points on two contours in order to compute precisely the difference between them [50]. The contour accuracy percentage is given by

Contour accuracy (%)

$$= 100 \times \left[\frac{1}{\max \{NC, NM\}} \sum_{i=1}^{NM} \frac{1}{1 + \beta \times d_i^2} \right], \quad (28)$$

where NC and NM are the points of edges in the segmented image and ground truth image, respectively. d_i is the Euclidean distance between an edge pixel and the nearest edge pixel of the ground truth and β is an empirical adjustment coefficient which is chosen here 1/9.

Table 4 illustrates performances evaluation of the MGAC model [24], fuzzy level set method [13] and the proposed active contour model for cell nuclei segmentation on the testing image dataset. From these results, we deduce that the new fuzzy active contour model is more robust than the other methods in terms of region and contour segmentation accuracies of cell nuclei. It gives an average region accuracy of 96.87%, whereas region accuracy rates are about 92.93% and 87.94% using respectively the fuzzy level set method and MGAC model. Using the contour measure (Eq. (28)), we found that the segmentation accuracy results of the proposed model are very promising and superior to those obtained by the other active contour models. The average rate of the contour accuracy is more than 87% against 81.92% and 75.76% given by fuzzy level set method and MGAC model, respectively. Furthermore, an advantage of applying the new fuzzy active contour model is the important improvement in the speed of convergence. The fuzzy clustering step took 85.6 s to process each microscopic image in the testing dataset. However, the phase of the discrimination function construction in the MGAC method took 257.74 s.

Table 4

Comparison of contour detection performance for all cell nuclei on the testing image dataset.

	MGAC model	Fuzzy level set method	The proposed model
<i>Region accuracy (%)</i>			
Mean	87.94	92.93	96.87
Standard deviation	3.21	2.75	2.14
Median	88.42	93.71	97.38
Min	78.64	88.48	93.74
Max	97.15	98.25	98.96
<i>Contour accuracy (%)</i>			
Mean	75.76	81.92	87.25
Standard deviation	7.54	6.35	4.85
Median	75.81	83.41	88.32
Min	72.64	79.56	82.61
Max	78.82	84.63	89.72
<i>Total time of convergence (s)</i>	187.25 ± 6.15	106.55 ± 3.61	65.32 ± 1.68

Then, the curve evolution took an average processing time 65.32 s on all the images of the testing database, compared to 106.55 s using the fuzzy level set method and 187.25 s applying the MGAC model. From these experiments, we conclude that the proposed active contour model provides accurate segmentation results and faster speed of convergence when compared with the other active contour methods.

3.2. Evaluation of the touching nuclei separation method

In order to evaluate the performance of the proposed separation method, the segmentation results of touching nuclei by the improved watershed algorithm are compared with those obtained with the classical watershed method [25] and shape marker-controlled watershed [31]. Fig. 7(a) and (b) show respectively the original image and the 'ground-truth' contours of all cell nuclei constructed manually by experts. The results are illustrated in Fig. 7(c)–(e) which represent the separation results of all nuclei by classical watershed, shape marker-controlled watershed and the improved watershed algorithm. These experiments show clearly that the shape marker-controlled watershed algorithm overcomes the limitations of the classical watershed method by reducing the over-segmentation. However, shape marker scheme can fail in detecting the number of markers when the size and shape change a lot inside the same clustered nuclei (Fig. 7(d)). Nevertheless, the improved watershed algorithm gives more accurate separation results than classical watershed and shape marker methods. In fact, the marking function used in [31] is based on a purely geometrical distance transform. Consequently, the shape and the size of segmented nuclei lack precision. In contrast, the proposed separation method gives a good partition of touching nuclei without losing the shape nuclear characteristics. But it could sometimes produce a little amount of under-segmented nuclei (Fig. 7(e)). This is probably due to the undetected concavity points on the dividing nuclei boundaries that are ambiguous to separate.

Table 5 shows the comparative separation results for 1372 touching nuclei on the complete image database. The segmentation results are branded into three groups: correctly segmented, over-segmented, and under-segmented touching nuclei. To study the sensitivity of the proposed segmentation method to the smoothing parameter of the inversed distance transform, the touching nuclei segmentation is achieved by varying the smoothing parameter σ_s . As shown in Table 5, the proposed watershed algorithm provides the best separation performance when the smoothing parameter is set to 2.5. The experiments show also that our algorithm is not very sensitive to the choice of this parameter. The accuracy variation for touching nuclei segmentation is less than 2% in the whole image database. It should be noted that, as illustrated in Table 5, the proposed method performs improvements by 43.09%, and 23.13% with respect to the classical watershed, and shape marker methods in terms of separation accuracy, respectively, when the smoothing parameter is 2.5. Moreover, the error rate of segmented cell nuclei by our method is reduced to lower than 4%.

3.3. Classification results using FLD–MNN classifier

In this section, we present a comparison study of the classification results using FLD–MNN method with those obtained by PCA–MNN and MNN classifiers. PCA–MNN is a combination of a principal components analysis (PCA) and a multilayer neural network (MNN). PCA is used to create new uncorrelated features from the input data and then MNN is used to classify these new components. MNN is a simple feed-forward neural network trained with the original nuclear features extracted from a set of training images. In our study, two sets of training images are

chosen from the dataset 1^(AEC) and dataset 2^(DAB) and containing respectively 3362 and 6256 segmented nuclei from the different classes. We used 24 training images extracted from dataset 1^(AEC) including 2352 P nuclei, 900 N nuclei and 110 B cells. In dataset 2^(DAB), we selected 56 learning images which comprise 4380 P nuclei, 1600 N nuclei and 276 B cells. For fair comparison of the three neural network-based methods, the MNN structures were designed during the training process using the appropriate features for each method. In Table 6, a comparative study between MNN classifiers in term of validation error rate is reported using five-fold cross-validation technique [45]. To compute the average validation error on the entire training image dataset, the data of sample nuclei is first divided into five equally sized folds. Then, five experiments of training and validation are performed such that within each experiment a different fold of the nuclei data is held-out for validation, and the remaining four-folds for training. From the cross-validation results detailed in Table 6, it is obvious that the FLD–MNN classifier is more efficient and robust than the other neural network-based methods since it provides the lowest validation error rate (less than 6%) over the five folds of the training dataset.

In order to provide a better evaluation of the studied classifiers' performances, all the trained classifiers are tested on the complete image database. Then, we performed a statistical analysis in terms of sensitivity (SE), specificity (SP) and accuracy (AC) of detected positive and negative cell nuclei:

$$SE = \frac{TP}{TP+FN} \quad (29)$$

$$SP = \frac{TN}{FP+TN} \quad (30)$$

$$AC = \frac{TP+TN}{TP+FN+TN+FP} \quad (31)$$

where TP is the number of true positives, TN is the number of true negatives, FN is the number of false negatives and FP is the number of false positives.

Table 7 shows the resulted percentages of SE, SP and AC for the complete image database of the studied networks. From these results, it is clear that the FLD–MNN method is more accurate than the other classifiers for the two image datasets. It provides the highest accuracy average (about 97.8% for P nuclei and 97.6% for N nuclei) with the smallest accuracy standard deviation (less than 1.7% for P and N nuclei) for cancer nuclei in overall image database. Note that the FLD preprocessing improves significantly the performance of the MNN classifier trained with PCA or the original features. The improvement in term of accuracy mean of all cancer nuclei can reach 3.9% and 3.7% using PCA–MNN and MNN, respectively. Fig. 8(a)–(c) show respectively the original image, the 'ground-truth' regions of all cell nuclei classified by pathologists and the classification results of the segmented nuclei using FLD–MNN classifier. The classified nuclei are marked by a color mask (benign nuclei – white; negative nuclei (0) – blue; positive nuclei: (1+) – yellow, (2+) – magenta, (3+) – red). As shown in Fig. 8(b) and (c), there is high agreement between the truth and the obtained classification results.

In Table 8, we report the results of cancerous nuclei counting and we demonstrate high precision of our proposed segmentation approach on a large database of real-life breast tissue images. Compared to the nuclei quantification provided by experts, the proposed method was able to detect cancer nuclei with an accuracy of 97.8% in the complete image database. Furthermore, the performance of the proposed method, in assigning the ER-score to each case of the image database, is compared against a corresponding evaluation by pathologists in Table 8. The classification agreement is 97.2% and 95.2% for dataset 1^(AEC) and dataset 2^(DAB), respectively.

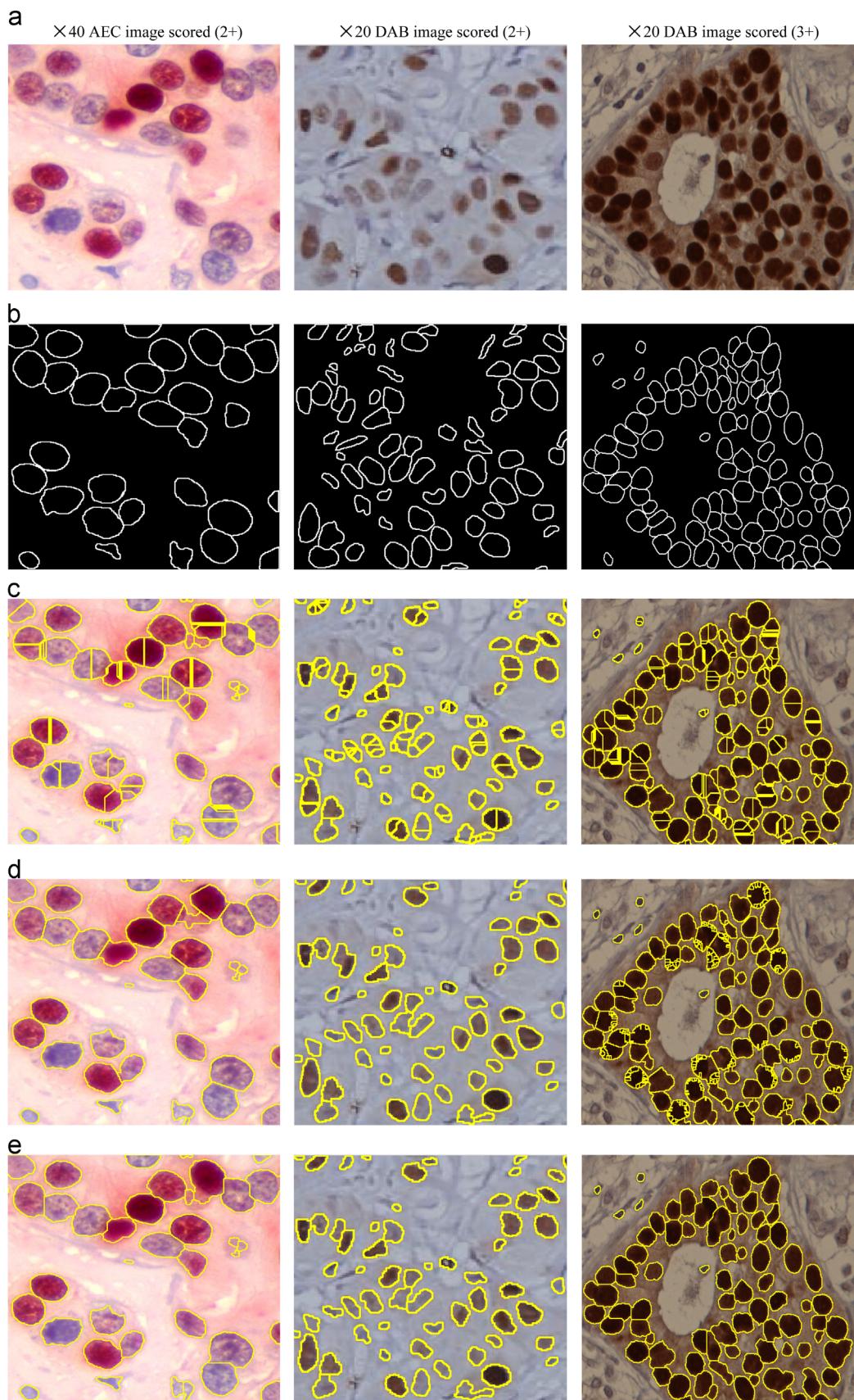


Fig. 7. Separation results of breast cell nuclei: (a) the original image and (b) the ‘ground-truth’ contours of all cell nuclei constructed manually by pathologists. Separation results by (c) classical watershed method [25], (d) shape marker-controlled watershed method [31], and (e) the improved watershed algorithm.

Table 5

Comparative separation results of touching nuclei on the complete image database.

	Classical Watershed method	Shape Marker-controlled Watershed	The improved Watershed algorithm		
			$\sigma_s = 1.5$	$\sigma_s = 2.5$	$\sigma_s = 3.5$
Correctly segmented nuclei (%)	54.76	74.72	96.33	97.85	95.93
Over-segmented nuclei (%)	44.72	15.05	2.42	0.34	0.41
Under-segmented nuclei (%)	0.52	10.23	1.25	1.81	3.66

Table 6Validation error rates (%) of the MNN classifiers using the five-fold cross-validation technique applied on the entire training image dataset (mean values \pm standard deviation).

Classifier	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
MNN	12.8 ± 2.7	12.3 ± 2.2	11.5 ± 1.7	10.1 ± 1.5	9.7 ± 1.2	11.3 ± 1.8
PCA-MNN	10.9 ± 2.8	9.7 ± 2.5	8.9 ± 1.9	8.1 ± 1.3	7.5 ± 0.9	9.1 ± 1.8
FLD-MNN	7.8 ± 1.4	6.8 ± 1.1	5.4 ± 0.9	4.2 ± 0.7	3.9 ± 0.5	5.6 ± 0.9

Table 7Performance comparison of the MNN classifiers in terms of sensitivity (SE), specificity (SP) and accuracy (AC) on the complete image database (mean values \pm standard deviation).

Image datasets	MNN			PCA-MNN			FLD-MNN		
	SE (%)	SP (%)	AC (%)	SE (%)	SP (%)	AC (%)	SE (%)	SP (%)	AC (%)
<i>Dataset 1</i> ^(AEC)									
P nuclei	91.1 ± 2.3	97.8 ± 1.2	96.4 ± 1.7	78.5 ± 2.1	97.9 ± 1	94.2 ± 1.6	96.9 ± 1.2	98.1 ± 0.9	97.4 ± 1.1
N nuclei	93.5 ± 2	92.3 ± 1.7	94.2 ± 1.9	93.7 ± 1.9	97.5 ± 1.5	96.5 ± 1.8	96 ± 1.7	97.6 ± 1.3	97.1 ± 1.5
<i>Dataset 2</i> ^(DAB)									
P nuclei	78.3 ± 2.3	94.7 ± 2.2	93.7 ± 2.1	83.2 ± 2.4	95.9 ± 2.1	94.5 ± 2.1	97.9 ± 2.3	98.6 ± 1.8	98.2 ± 2.1
N nuclei	95.8 ± 2.3	87.4 ± 2.1	91.6 ± 2.2	95.7 ± 2.2	87.8 ± 2	90.1 ± 2	97.6 ± 2.2	98.3 ± 1.6	98.1 ± 2
<i>Overall database</i>									
P nuclei	84.7 ± 2.3	96.2 ± 1.7	95 ± 1.9	80.8 ± 2.2	96.9 ± 1.5	94.3 ± 1.8	97.4 ± 1.7	98.3 ± 1.3	97.8 ± 1.6
N nuclei	94.6 ± 2.1	89.8 ± 1.9	92.9 ± 2	94.7 ± 2	92.6 ± 1.7	93.3 ± 1.9	96.8 ± 1.9	97.9 ± 1.4	97.6 ± 1.7

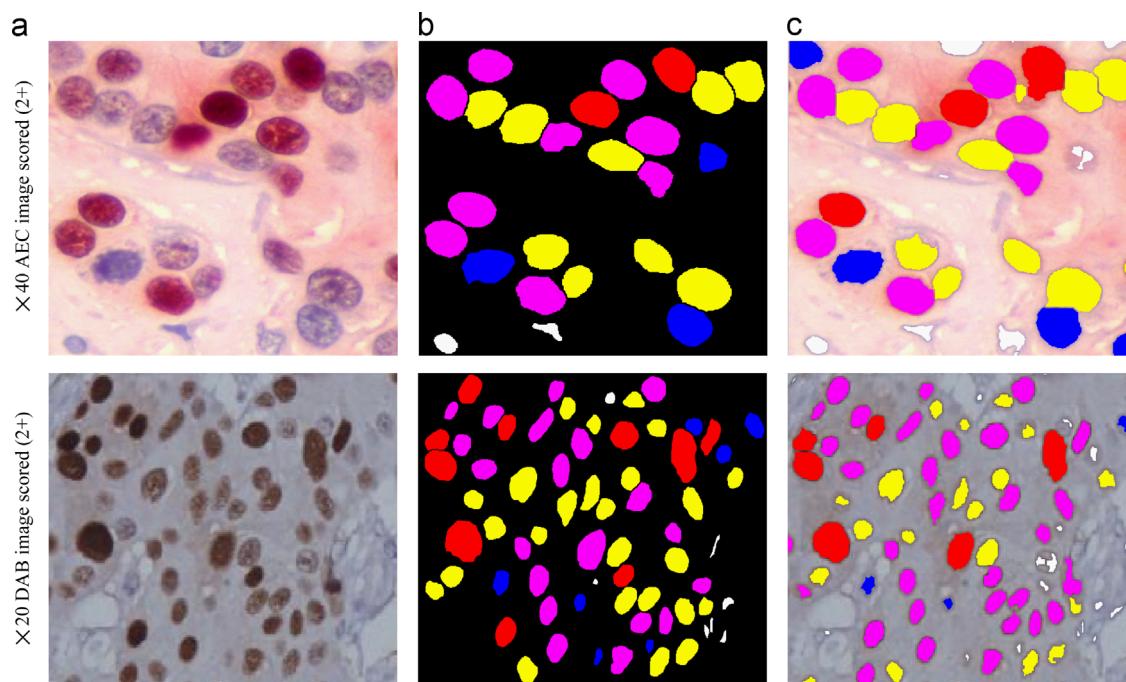
**Fig. 8.** Classification results of breast cell nuclei: (a) the original image, (b) the 'ground-truth' cell nuclei classification identified by pathologists and (c) classification results of the segmented nuclei using FLD-MNN classifier, where the classified nuclei are marked by a color mask (benign nuclei – white; negative nuclei (0) – blue; positive nuclei: (1+) – yellow, (2+) – magenta, and (3+) – red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 8

Results of the computer-assisted ER evaluation system versus expert's assessment on the complete image database.

	Computer-assisted ER evaluation			Expert's assessment		
	Dataset 1 ^(AEC)	Dataset 2 ^(DAB)	Overall Database	Dataset 1 ^(AEC)	Dataset 2 ^(DAB)	Overall Database
P nuclei						
TP	3015	5690	8705	3045	5716	8761
FP	32	46	78	0	0	0
TN	1706	3363	5069	1775	3446	5221
FN	96	122	218	0	0	0
N nuclei						
TP	1548	2995	4543	1617	3078	4695
FP	79	102	181	0	0	0
TN	3173	6058	9231	3203	6084	9287
FN	64	73	137	0	0	0
Nucleicounting AC (%)	97.2	98.1	97.8	100	100	100
TNI (1+)	10	30	40	10	32	42
NFI (1+)	0	0	0	0	0	0
TNI (2+)	19	42	61	20	42	62
NFI (2+)	0	2	2	0	0	0
TNI (3+)	7	12	19	6	10	16
NFI (3+)	1	2	3	0	0	0
Cancer scoring AC (%)	97.2	95.2	95.8	100	100	100

TNI: total number of classified images; NFI: number of False classified images; AC: accuracy.

In other words, computer-assisted image analysis system gives 95.8% in overall accuracy, ranking correctly 115/120 cases. With these results, we conclude that the proposed image analysis method can reliably be used to assist pathologists by providing a second opinion for ambiguous cases that require further attention.

4. Conclusions

In this paper, we proposed a fully automated image analysis method for assessing ER status in breast cancer specimens. The proposed method combines a new fuzzy active contour model and an improved watershed algorithm for cell nuclei segmentation. Then, an FLD-MNN classifier is designed using textural and morphological nuclei features in order to identify positive and negative stained nuclei, for subsequent ER-status evaluation.

A comparative study versus existing techniques is performed in order to demonstrate the effectiveness of the proposed method in terms of the segmentation and the classification accuracies. Furthermore, all obtained results are verified with the truths given by pathologists in order to provide a better evaluation of the proposed methods' performances. First, we have shown that the new fuzzy active contour model gives the fastest speed of convergence among earlier active contour methods while providing precise segmentation results. In addition, the proposed active contour model estimates the initial curve and the controlling parameters from fuzzy clustering automatically which reduce importantly manual intervention and accelerate the contour evolution speed. Second, results of touching nuclei segmentation show clearly that the improved watershed algorithm gives precise results in term of nuclei number and can reliably segment touching nuclei even in complex morphological cases. Moreover, we have shown that the performance of the new separation method is insensitive to the type of nuclei staining and to the irregular configuration of cell nuclei in shape and size. The experiments prove the superior performance of the proposed separation method. Finally, we found encouraging classification results of the segmented cell nuclei using the FLD-MNN classifier. The classification accuracies are respectively about 97.8% and 97.6% for positive and negative stained nuclei. Besides, the positivity score provided by our proposed approach in each case is very close

to the score announced by the experts, achieving an accuracy of 95.8% on a large real-life image database. In conclusion, the proposed image analysis method for automatic assessment of ER status in IHC stained breast specimens can reliably be used as an additional tool to assist pathologists in the cancer diagnosis process.

5. Summary

The aim of this work is to create an automatic image analysis method for breast cancer nuclei scoring in order to assist pathologists to make objective and fast diagnosis. In this paper, we propose a new nuclear segmentation approach for quantitative assessment of estrogen receptors (ER) positive status in breast cancer sections. The originality of this paper results firstly from the proposal of a new fuzzy active contour model combined with an improved watershed algorithm based on concave vertex graph to detect and segment all stained nuclei in the microscopic image. Secondly, we have studied the ability of the proposed nuclei segmentation scheme to cooperate with a trained multilayer neural network classifier in evaluating ER status of several breast cancer specimens. The performance of the method is verified when compared to state-of-art techniques as well as to the manually truths given by experienced pathologists.

Conflict of interest statement

None declared.

References

- [1] S. Di Cataldo, E. Ficarra, E. Macii, Computer-aided techniques for chromogenic immunohistochemistry: status and directions, *Comput. Biol. Med.* 42 (2012) 1012–1025.
- [2] L He, L.R. Long, S. Antani, G.R. Thoma, Histology image analysis for carcinoma detection and grading, *Comput. Methods Programs Biomed.* 107 (2012) 538–556.
- [3] A. Mouelhi, M. Sayadi, F. Fnaiech, K. Mrad, K.B. Romdhane, Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method, *Biomed. Signal Process. Control* 8 (2013) 421–436.
- [4] M. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, B. Yener, Histopathological image analysis: a review, *IEEE Rev. Biomed. Eng.* 2 (2009) 147–171.

- [5] S. Di Cataldo, E. Ficarra, A. Acquaviva, E. Macii, Achieving the way for automated segmentation of nuclei in cancer tissue images through morphology-based approach: a quantitative evaluation, *Comp. Med. Imag. Graph.* 34 (2010) 453–461.
- [6] B. Jasani, A. Douglas-Jones, A. Rhodes, S. Wozniak, P.J. Barrett-Lee, J. Gee, R. Nicholson, Measurement of estrogen receptor status by immunocytochemistry in paraffin wax sections, *Methods Mol. Med.* 120 (2006) 127–146.
- [7] S. Kostopoulos, D. Cavouras, A. Daskalakis, I. Kalatzis, P. Bougioukos, G. Kagadis, P. Ravazoula, G. Nikiforidis, Assessing Estrogen Receptors' status by texture analysis of breast tissue specimens and pattern recognition methods, in: Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns: CAIP'07, 2007, pp. 221–228.
- [8] R. Leake, D. Barnes, S. Pinder, I. Ellis, L. Anderson, T. Anderson, R. Adamson, T. Rhodes, K. Miller, R. Walker, Immunohistochemical detection of steroid receptors in breast cancer: a working protocol. UK Receptor Group, UK NEQAS, The Scottish Breast Cancer Pathology Group, and The Receptor and Biomarker Study Group of the EORTC, *J. Clin. Pathol.* 53 (2000) 634–635.
- [9] P. Phukpattaranont, P. Boonyaphiphat, Color based segmentation of nuclear stained breast cancer cell images, *ECTI Trans. Electr. Eng., Electron. Commun.* 5 (2007) 158–164.
- [10] O. Lezoray, H. Cardot, Cooperation of color pixel classification schemes and color watershed: a study for microscopic images, *IEEE Trans. Image Process.* 11 (2002) 783–789.
- [11] L. Yang, P. Meer, D. Foran, Unsupervised segmentation based on robust estimation and color active contour models, *IEEE Trans. Inform. Technol. Biomed.* 9 (2005) 475–486.
- [12] S. Schüpp, A. Elmoataz, J. Fadihi, P. Herlin, D. Bloyet, Image segmentation via multiple active contour models and fuzzy clustering with biomedical applications, in: Proceedings of the 15th International Conference on 'Pattern Recognition': ICPR'00, vol. 1, 2000, pp. 622–625.
- [13] B.N. Li, C.K. Chui, S. Chang, S.H. Ong, Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation, *Comput. Biol. Med.* 41 (2011) 1–10.
- [14] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, *Int. J. Comput. Vision* 4 (1988) 321–331.
- [15] S. Osher, R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, New York, 2003.
- [16] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours, *Int. J. Comput. Vision* 22 (1997) 61–79.
- [17] C. Xu, A. Yezzi, J. Prince, On the relationship between parametric and geometric active contours, in: Proceedings of Asilomar Conference on Signal, Systems and Computers, vol. 1, 2000, pp. 483–489.
- [18] J.A. Sethian, *Level set Methods and Fast Marching Methods*, second ed., Cambridge University Press, New York, 1999.
- [19] R. Goldenberg, R. Kimmel, E. Rivlin, M. Rudzsky, Fast geodesic active contours, *IEEE Trans. Image Process.* 10 (2001) 1467–1475.
- [20] C.M. Li, C.Y. Xu, C.F. Gui, M.D. Fox, Level set evolution without re-initialization: a new variational formulation, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, 2005, pp. 430–436.
- [21] T.F. Chan, L.A. Vese, Active contours without edges, *IEEE Trans. Image Process.* 2 (2001) 266–277.
- [22] L. Chen, Y. Zhou, Y.G. Wang, J. Yang, GACV: geodesic-aided C-V method, *Pattern Recognition* 39 (2006) 1391–1395.
- [23] Z. Ying, L. Guangyao, S. Xiehua, Z. Xinmin, A geometric active contour model without re-initialization for color images, *Image Vision Comput.* 27 (2009) 1411–1417.
- [24] L. Pi, J.S. Fan, C.M. Shen, Color image segmentation for objects of interest with modified geodesic active contour method, *J. Math. Imag. Vision* 1 (2007) 51–57.
- [25] L. Vincent, P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (1991) 583–598.
- [26] A.N. Moga, M. Gabiouj, Parallel marker-based image segmentation with watershed transformation, *J. Parallel Distrib. Comput.* 1 (1998) 27–45.
- [27] K. Haris, S.N. Efstratiadis, N. Maglaveras, A.K. Katsaggelos, Hybrid image segmentation using watersheds and fast region merging, *IEEE Trans. Image Process.* 12 (1998) 1684–1699.
- [28] V. Grau, A.U. Mewes, M. Alcaniz, R. Kikinis, S.K. Warfield, Improved watershed transform for medical image segmentation using prior information, *IEEE Trans. Med. Imaging* 23 (2004) 447–458.
- [29] I. Dagher, K.E. Tom, WaterBalloons: a hybrid watershed Balloon Snake segmentation, *Image Vision Comput.* 26 (2008) 905–912.
- [30] L. Yang, O. Tuzel, P. Meer, D.J. Foran, Automatic image analysis of histopathology specimens using concave vertex graph, *MICCAI, Part I, Lect. Notes Comput. Sci.* 5241 (2008) 833–841.
- [31] J. Cheng, J.C. Rajapakse, Segmentation of clustered nuclei with shape markers and marking function, *IEEE Trans. Biomed. Eng.* 56 (2009) 741–748.
- [32] K.-S. Chuang, H.-L. Tseng, S. Chen, J. Wu, T.-J. Chen, Fuzzy c-means clustering with spatial information for image segmentation, *Computerized Med. Imaging Graphics* 30 (2006) 9–15.
- [33] M. Sayadi, L. Tlig, F. Fnaiech, A new texture segmentation method based on the fuzzy C-means algorithm and statistical features, *Appl. Math. Sci.* 1 (2007) 2999–3007.
- [34] G. Lin, U. Adiga, K. Olson, J.F. Guzowski, C.A. Barnes, B. Roysam, A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks, *Cytometry Part A* 56A (2003) 23–36.
- [35] F. Clopet, A. Boucher, Segmentation of overlapping/aggregating nuclei cells in biological images, in: Proceedings of 19th International Conference on Pattern Recognition: ICPR'08, 2008, pp. 1–4.
- [36] Fatakdawala, H. Basavanhally, J. Xu, G. Bhanot, S. Ganesan, M. Feldman, J. Tomaszewski, A. Madabhushi, Expectation-Maximization-Driven Geodesic Active Contour With Overlap Resolution (EMaGACOR): application to lymphocyte segmentation on breast cancer histopathology, *IEEE Trans. Biomed. Eng.* 57 (2010) 1676–1689.
- [37] S. Kothari, Q. Chaudri, M.D. Wang, Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques, in: ISBI'09 Proceedings of Sixth IEEE International Conference on Symposium on Biomedical Imaging: from Nano to Macro, 2009, pp. 795–798.
- [38] X.C. He, N.H.C. Yung, Corner detector based on global and local curvature properties, *Optical Eng.* 47 (5) (2008) 1–12. (057008).
- [39] W. Wang, Binary image segmentation of aggregates based on polygonal approximation and classification of concavities, *Pattern Recognition* 10 (1998) 1503–1524.
- [40] D. Chetverikov, Z. Szabó, A simple and efficient algorithm for detection of high curvature points in planar curves, in: Proceedings of the 23rd Workshop of the Austrian Pattern Recognition Group, 1999, pp. 175–184.
- [41] A. Mouelhi, M. Sayadi, F. Fnaiech, Automatic segmentation of clustered breast cancer cells using watershed and concave vertex graph, in: IEEE International Conference on 'Communications, Computing and Control Applications': CCCA'11, Hammamet, Tunisia, March 2011.
- [42] E.W. Dijkstra, A note on two problems in connexion with graphs, *Numerische Math.* 1 (1959) 269–271.
- [43] J.S. Stahl, S. Wang, Convex grouping combining boundary and region information, *ICCV* 2 (2005) 946–953.
- [44] X. Long, W.L. Cleveland, Y.L. Yao, A new preprocessing approach for cell recognition, *IEEE Trans. Inform. Technol. Biomed.* 9 (3) (2005) 407–412.
- [45] T. Andersen, T. Martinez, Cross validation and MLP architecture selection, *Proceedings of International Joint Conference on Neural Networks: IJCNN'99* 3 (1999) 1614–1619.
- [46] S. Abid, A. Mouelhi, F. Fnaiech, Accelerating the multilayer perceptron learning with the Davidon Fletcher Powell Algorithm, in: IEEE International Joint Conference on Neural Network: IJCNN'06, Vancouver, Canada, 16–21 July 2006, pp. 3389–3394.
- [47] A.M. Martinez, A.C. Kak, PCA versus LDA, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2001) 228–233.
- [48] L. Jelen, T. Fevens, A. Krzyzak, Classification of breast cancer malignancy using cytological images of fine needle aspiration biopsies, *Int. J. Appl. Math. Comput. Sci.* 18 (2008) 75–83.
- [49] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, third ed., Wiley, New York, 2003.
- [50] I.E. Abdou, W.K. Pratt, Quantitative design and evaluation of enhancement/thresholding edge detectors, *Proc. IEEE* 67 (1979) 753–763.