COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 113 (2014) 539–556

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Automatic cervical cell segmentation and classification in Pap smears

Thanatip Chankong[a], *Nipon Theera-Umpon (Senior Member, IEEE)*[a,b,*], *Sansanee Auephanwiriyakul (Senior Member, IEEE)*[c,b]

[a] *Department of Electrical Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand*
[b] *Biomedical Engineering Center, Chiang Mai University, Chiang Mai 50200, Thailand*
[c] *Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand*

ARTICLE INFO

Article history:
Received 12 June 2013
Received in revised form
26 October 2013
Accepted 18 December 2013

Keywords:
Cervical cell classification
Cervical cell segmentation
Pap smear
ThinPrep
Cervical cancer screening

ABSTRACT

Cervical cancer is one of the leading causes of cancer death in females worldwide. The disease can be cured if the patient is diagnosed in the pre-cancerous lesion stage or earlier. A common physical examination technique widely used in the screening is Papanicolaou test or Pap test. In this research, a method for automatic cervical cancer cell segmentation and classification is proposed. A single-cell image is segmented into nucleus, cytoplasm, and background, using the fuzzy C-means (FCM) clustering technique. Four cell classes in the ERUDIT and LCH datasets, i.e., normal, low grade squamous intraepithelial lesion (LSIL), high grade squamous intraepithelial lesion (HSIL), and squamous cell carcinoma (SCC), are considered. The 2-class problem can be achieved by grouping the last 3 classes as one abnormal class. Whereas, the Herlev dataset consists of 7 cell classes, i.e., superficial squamous, intermediate squamous, columnar, mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ. These 7 classes can also be grouped to form a 2-class problem. These 3 datasets were tested on 5 classifiers including Bayesian classifier, linear discriminant analysis (LDA), K-nearest neighbor (KNN), artificial neural networks (ANN), and support vector machine (SVM). For the ERUDIT dataset, ANN with 5 nucleus-based features yielded the accuracies of 96.20% and 97.83% on the 4-class and 2-class problems, respectively. For the Herlev dataset, ANN with 9 cell-based features yielded the accuracies of 93.78% and 99.27% for the 7-class and 2-class problems, respectively. For the LCH dataset, ANN with 9 cell-based features yielded the accuracies of 95.00% and 97.00% for the 4-class and 2-class problems, respectively. The segmentation and classification performances of the proposed method were compared with that of the hard C-means clustering and watershed technique. The results show that the proposed automatic approach yields very good performance and is better than its counterparts.

* *Corresponding author at*: Biomedical Engineering Center, Chiang Mai University, Chiang Mai 50200, Thailand. Tel.: +66 5394 2083; fax: +66 5322 1485.
E-mail addresses: tchankong@gmail.com (T. Chankong), nipon@ieee.org, nipon.tor@gmail.com, DrNipon@chiangmai.ac.th (N. Theera-Umpon), sansanee@ieee.org (S. Auephanwiriyakul).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.12.012

1. Introduction

Cervical cancer is the fourth leading cause of cancer death in females worldwide [1]. The prognosis for cervical cancer depends on the stage of the cancer at the time of detection. The disease can be cured if diagnosed in the pre-cancerous lesion stage or earlier. Papanicolaou test or Pap test is a physical examination technique widely used to prevent cervical cancer by finding cells that have the potential to turn cancerous. It was estimated that, in the year 2006, systematic screening can reduce mortality rates from cervical cancer by 70% or more [2].

In Thailand, cervical cancer is the second most common cancer among women [3], with a high mortality rate of nearly 300,000 per year. The screening program administered by the Ministry of Public Health and the National Health Security Office suggested that women aged 35–60 years should undergo Pap smear examination every five years [4]. With an overall population number of 67 million in Thailand [5], the number of samples to be examined per year is large comparing to the number of cytologists who process the screening. The study of automatic cervical cell classification has been done for over 40 years to cope with this labor-intensive task of manual screening and also to reduce the error of screening result. A number of commercial automated screening systems have been approved by the FDA for quality control with examples including PAPNET (Neuromedical Systems Inc.), FocalPoint Slide Profiler^{TM} (formerly AUTOPAP; BD TriPath), ThinPrep Pap Test, ThinPrep Imaging System (Hologic Inc.) and Imager^{TM} (Cytyc). Several research works have shown that these automated systems indeed improve the accuracy of the screening result and reduce the false-negative rate [6–8]. However, cost effectiveness is a major drawback of these systems with the cost of PAPNET test far exceeds that of manual screening.

Uncertainty in diagnostic capability was also reported [9]. It is therefore suggested that the automated system should be used as an aiding tool in conjunction with the expert's opinion rather than relying on the system as a primary screening and diagnosing tool [10,11]. Over the past few years, trend of research in automated cervical cancer screening has shifted from cytology screening to histology image [12,13] and colposcopic image [14,15]. Histology image is not only used in cervical cancer screening, but it is used in the other kinds of cancer screening also [16–18]. However, cytology screening is still a default screening method in most countries due to its relatively low cost and its effectiveness in cervical cancer prevention if the screening is routinely performed.

The screening process normally starts with gathering cervical cell samples from the uterine cervix and mounting it on a glass slide. The collected sample is visually inspected under a microscope to identify the target cell or classify each cell into categories. The basic characteristics used to classify the stage of cells are mainly the characteristics of cell nuclei and cytoplasm such as shape, size, texture, ratio of nucleus and cytoplasm. From image processing point of view, the first step in extracting information from cell components is to correctly identify a region of each component (nucleus, cytoplasm, and non-cell components) by segmentation process. There are several research works on nucleus segmentation [19–22]. However, when one would like to classify each cervical cell into categories with only nucleus information, it might not yield a good performance. Hence, segmenting whole cell is more desirable [23–26]. However, there is no classification result reported in these works. After the segmentation step, each cell is then classified using specific classifiers based on the extracted features from cell components as mentioned earlier or by using filters to discriminate classes without feature extraction process [27]. However, the classification performance in Ref. [27] is not quite high.

In this research, we propose a method for automatic segmentation and classification of cervical cell images. In the segmentation process, we use a patch-based fuzzy C-means (FCM) clustering technique. A cell image is segmented by using the over-segment FCM technique into nucleus, cytoplasm, and background. For comparison, we use the hard C-means clustering technique and the watershed segmentation technique in the segmentation step as well. The segmented image is then used to extract related features to be the input to classifiers. Five classifiers including Bayesian classifier, linear discriminant analysis, K-nearest neighbor, artificial neural networks, and support vector machine, are considered. The usefulness of features based on nucleus and entire cell is also investigated.

This paper is organized as follows: the following section describes basic knowledge of the segmentation technique, mathematical morphologies, feature extraction, classifiers, followed by the details of cell datasets used in this research. Section 3 provides the experimental results and discussion. The conclusion is drawn in Section 4.

2. Materials and methods

2.1. *Cervical cell segmentation using patch-based fuzzy C-means clustering*

We invented the patch-based fuzzy C-means (FCM) clustering method to segment nuclei and cytoplasm of white blood cells [28]. It was later applied to segment nuclei of cervical cells from the conventional Pap test [29]. The FCM is good for clustering data with uncertainty. We, therefore, chose the FCM to cluster uncertain cell image intensity data. In this research, the segmentation method is tested with cervical cells from, not only the conventional Pap test, but also the ThinPrep® Pap test. The segmentation of cytoplasm is also performed. The segmentation algorithm is as follows:

Convert color input image to grayscale
Apply median filter
Apply FCM algorithm
Sort FCM centers in ascending order
For *each patch corresponding to sorted FCM centers (from dark to bright)*
  If *((FCM center of patch)* $\leq T_N$*), then label patch as nucleus*
  Else if *(*$T_N <$ *(FCM center of patch)* $\leq T_C$*), then label patch as cytoplasm*
  Else *label patch as background*
End *(For each patch)*

In segmentation, a single-cell image is converted to a grayscale image before applying a 7×7 median filter to smooth the image and eliminate noise. The processed image is then segmented into 2 or 3 regions, i.e., nucleus and non-nucleus (for the ERUDIT Pap smear dataset), or nucleus, cytoplasm, and background (for the LCH and Herlev Pap smear dataset), using the FCM clustering method. Rather than considering every pixel value and directly cluster them into 2 or 3 clusters, we overly cluster the pixels into patches where each patch is represented by its center value. The median filter size and the number of clusters can be chosen by experiments because of image variation in different datasets. To combine patches into 2 final clusters (nucleus and non-nucleus), we select the threshold T_N for nucleus in accordance with the percentages of all patch centers. The patch with the value of centers less (darker) than the nucleus threshold is labeled as nucleus. In the experiments, the nucleus threshold T_N is varied from 60% to 130% (with 10% incremental step) of the mean of patch centers. The threshold that gives the minimum error between automatic segmentation and manual segmentation is selected as the nucleus threshold. If the segmentation results in multiple objects, the object in which its centroid is the closest to the image center is selected to be the object of interest. On the other hand, to combine patched into 3 final clusters (nucleus, cytoplasm, and background), the cytoplasm threshold T_C has to also be determined. The nucleus threshold can be found using exactly the same method described in the 2-cluster case. The cytoplasm threshold T_C is varied from 90% to 160% (with 10% incremental step) of the mean of patch centers. The cytoplasm threshold is selected by experiment like in the nucleus segmentation. Finally, the patch with the value of center less than the nucleus threshold is labeled as nucleus. The patch with the value of center between the nucleus threshold and cytoplasm threshold is labeled as cytoplasm. The remaining patches are labeled as background.

2.2. Mathematical morphological operations

Morphological operations are non-linear, translation invariant transformations [30]. In this research, morphological opening and closing are applied to clean up the segmented images. The basic morphological operations involving an image S and a structuring element E are:

$$\text{dilation} : S \oplus E = \cup\{E + s : s \in S\}, \tag{1}$$

$$\text{erosion} : S \ominus E = \cap\{S - e : e \in E\}, \tag{2}$$

where $\cap$ and $\cup$ denote the set intersection and union, respectively, and $A + x$ denotes the translation of a set A by a point x, i.e.,

$$A + x = \{a + x : a \in A\}. \tag{3}$$

The closing and opening operations, derived from the erosion and dilation, are defined by:

$$\text{closing} : S \bullet E = (S \oplus (-E)) \ominus (-E), \tag{4}$$

$$\text{opening} : S \circ E = (S \ominus E) \oplus E, \tag{5}$$

where $-E = \{-e : e \in E\}$ denotes the 180° rotation of E about the origin.

2.3. Feature extraction

After applying the automatic segmentation based on FCM clustering, a set of features is extracted from nucleus and cytoplasm in each image. There are 6 features calculated from nucleus alone. Three more features need information from cytoplasm. The features are selected based on the nucleus and entire cell morphology during the cell division process. In the normal condition, chromosomes are not visible under the examination by a light microscope. But during the cell division process, chromatin condenses and becomes visible [31–33]. For the abnormal cells, the cell division occurs with an uncontrolled and high rate. The abnormal cell division activity can be seen from the nucleus characteristics [34]. Chromatin patterns in stained slide are well visible. When examine under the light microscope, one of the features used for grading the level of abnormality in cervical cell is the texture of nucleus, i.e., the coarseness of texture and hyperchromasia [34]. The coarseness of nucleus texture represents the distribution of chromatin. It increases according to the level of severity [35]. In the normal grade, granular of nucleus appears to be smooth and fine. Cells in the LSIL grade show mildly coarse granulars. The HSIL and SCC grades show coarsely distributed chromatin textures [36].

Hence, six nucleus-based features are selected as follows:

Feature 1: area of nucleus,

$$A_{nu} = \text{Total number of pixels in the nucleus region.} \tag{6}$$

Feature 2: compactness of nucleus,

$$C_{nu} = \frac{P_{nu}^2}{A_{nu}}, \tag{7}$$

where P_{nu} is the perimeter of the nucleus.

Feature 3: major axis of nucleus,

$$L_{nu} = \text{the length of the major axis of an ellipse that completely encloses the nucleus region.} \tag{8}$$

Feature 4: minor axis of nucleus,

$$D_{nu} = \text{the length of the minor axis of an ellipse that completely encloses the nucleus region.} \tag{9}$$

Feature 5: aspect ratio of nucleus,

$$R_{nu} = \frac{W_{nu}}{H_{nu}}, \tag{10}$$

where W_{nu} is the width of the nucleus and H_{nu} is the height of the nucleus region.

Feature 6: homogeneity of nucleus.

$$H_{nu} = \sum_{i=1}^{K}\sum_{j=1}^{K}\frac{P(i,j)}{1+|i-j|}, \tag{11}$$

where $P(i,j)$ is the probability of occurrence of a pair of pixel values (i,j) in the nucleus region computed from gray-level co-occurrence matrix. K is the number of gray levels in the image.

The first 5 features consider the shape of nucleus, whereas the last one takes into account the coarseness of nucleus.

The other 3 features based on the entire cell are as follows:

Feature 7: nucleus-to-cytoplasm (N/C) ratio,

$$NC = \frac{A_{nu}}{A_{cy}}, \tag{12}$$

where A_{nu} is the nucleus area and A_{cy} is the cytoplasm area.

Feature 8: compactness of the entire cell,

$$C_{en} = \frac{P_{en}^2}{A_{en}}, \tag{13}$$

where P_{en} is the perimeter of the entire cell and A_{en} is the entire cell area.

Feature 9: area of entire cell,

$$A_{en} = \text{Total number of pixels in the entire cell region.} \tag{14}$$

During the training process, the values of each feature are normalized to be in the range of 0–1 by dividing each of them by the corresponding largest value of that feature in the training set. These largest values are also used to normalize input data in the testing process of the cross validation.

2.4. Classifiers

The performance of 5 classifiers on discriminating different cervical cell classes is evaluated. Because the descriptions of classifiers are widely available in the literature [37–39], only a brief description of each classifier used in this research is provided here as follows:

- *Bayesian classifier*: Assumed that a priori probability distribution is normal, the posterior probabilities for all classes are calculated using the Bayes' rule. The class of each cell image is assigned based on the maximum likelihood method.
- *Linear discriminant analysis* (*LDA*): It is assumed that the cell classification problem is linearly separable. The discriminant function is calculated and used as a rule to assign the cell into separated classes.
- *K-nearest neighbor* (*KNN*): The test cell is classified based on the Euclidean distance between itself and all cells in the training set. The distances of all samples are sorted and the Kth minimum distance is determined. In this research, 7 nearest neighbors ($K = 7$) are used. The test cell is assigned to the majority class of the nearest neighbors.
- *Artificial neural networks* (*ANN*): This classifier is a non-linear statistical classifier which can be used to classify the data with complex relation between input and output. The ANN consists of a number of neurons organized in layers. This research uses a 3-layer backpropagation neural network consisting of 5–9 neurons in the input layer depending on the number of features used, 1 hidden layer with 15 neurons, and 2, 4, or 7 neurons in the output layers depending on the number of classes being considered.
- *Support vector machine* (*SVM*): This classifier finds the optimal hyperplane that separates clusters of data in the feature space. The data is mapped by a kernel function into a different space where the data can be separated by a selected hyperplane. In this research, the radial basis function (RBF) kernel is used with $\sigma = 1.8$.

It is worthwhile noting that the value of each aforementioned chosen parameter was determined from extensive experiments by varying its value. The values of parameters that provided the best results were chosen for our experiments.

2.5. Data descriptions

2.5.1. ERUDIT Pap smear dataset

There are 552 single-cell images, including 138 normal cells and 414 abnormal cells, obtained from Ref. [40]. The resolution of the images in this dataset is 0.201 μm per pixel. They were classified into 4 classes based on the degree of abnormality as follows:

- normal cell images with 63 superficial cells and 75 intermediate cells (138 cells in total),
- low-grade squamous intraepithelial lesions (LSIL) (138 cells),
- high-grade squamous intraepithelial lesions (HSIL) (138 cells),
- squamous cell carcinoma (SCC) (138 cells).

Each cell image in this dataset covers nucleus but may not cover the entire cell. Therefore, the entire cytoplasm may or may not be present in this dataset. Sample images of different classes of cervical cells in the ERUDIT dataset are shown in Fig. 1.

2.5.2. LCH Pap smear dataset

Data collection for the LCH dataset was approved by the Lampang Cancer Hospital Ethics Committee. We obtained ThinPrep® slides from Anatomical Pathology Laboratory, Cytology Department, Lampang Cancer Hospital, Thailand. The cell images were captured by a BX51 microscope (Olympus, Japan) with a DP52 digital camera attachment with optical magnification of 40×. The resolution of the images in the LCH dataset is 0.09 μm per pixel. We manually cropped each single-cell image which was classified by an expert into 4 classes. The plan was to have a balance 2-class dataset, i.e., the number of cells in normal and abnormal classes is equal. A total of 300 cells were classified by the expert into 4 classes consisting of:

- 150 normal cells with 80 normal cells and 70 atypical squamous cells of undetermined significance (ASC-US) cells,
- 64 low grade squamous intraepithelial lesion (LSIL) cells,
- 38 high grade squamous intraepithelial lesion (HSIL) cells,
- 48 squamous cell carcinoma (SCC) cells.

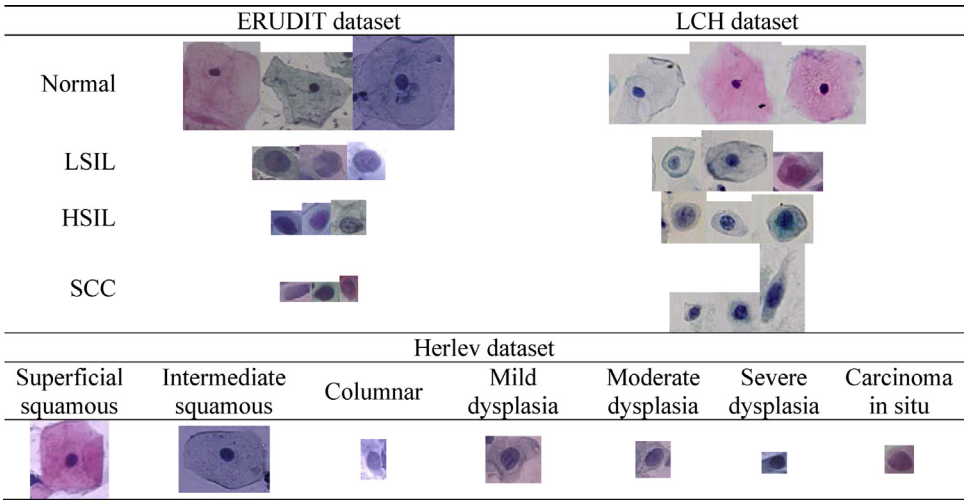

Fig. 1 – Sample cervical cells in ERUDIT, LCH, and Herlev Pap smear datasets.

Therefore, there are 150 normal cells and 150 abnormal cells images in total with various degrees of abnormality. Each cell image in this dataset covers the entire cell. Therefore, both nucleus and cytoplasm information is present in this dataset. Sample images of different classes of cervical cells in the LCH dataset are also shown in Fig. 1.

2.5.3. *Herlev Pap smear dataset*

There are 917 single-cell images obtained from Ref. [41]. The resolution of the images in this dataset is 0.201 μm per pixel. Each cell is in one of 7 classes, i.e.,

- superficial squamous (74 cells),
- intermediate squamous (70 cells),
- columnar (98 cells),
- mild dysplasia (182 cells),
- moderate dysplasia (146 cells),
- severe dysplasia (197 cells),
- carcinoma in situ (150 cells).

Samples cells in the 7 classes are shown in Fig. 1. According to Ref. [27], in 2-class problem, superficial squamous and intermediate squamous classes are considered as a normal class, whereas an abnormal class consists of cells from mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ classes. Columnar cells are classified as neither normal nor abnormal. Hence, there are 144 and 675 cells in normal class and abnormal class, respectively.

2.6. *Evaluation measures*

There are 2 types of evaluations used in this work, i.e., segmentation evaluation and classification evaluation.

2.6.1. *Evaluation measures for segmentation*

For segmentation evaluation, the nuclei and whole cell ground truths are needed. The performance of the detected segmentation boundary is measured using pixel-based precision and recall [27], i.e.,

$$\text{Precision} = \frac{\text{No. of correctly detected pixels}}{\text{No. of all detected pixels}} = \frac{TP_s}{TP_s + FP_s}, \quad (15)$$

$$\text{Recall} = \frac{\text{No. of correctly detected pixels}}{\text{No. of all pixels in the ground truth}} = \frac{TP_s}{TP_s + FN_s}, \quad (16)$$

where TP_s is the number of correctly detected pixels. FP_s is the number of detected pixels that are not in the ground truth and FN_s is the number of pixels in the ground truth that are not detected. We also utilized another measurement, i.e., Zijdenbos similarity index (*ZSI*) [42] calculated as follows:

$$ZSI = \frac{2TP_s}{2TP_s + FP_s + FN_s}, \quad (17)$$

where $ZSI \in [0,1]$. According to Ref. [42], the detected segmentation boundary and the ground truth is matched excellently if *ZSI* is bigger than 0.7.

All three indices measure the correctness of the segmentation. We also utilized the probability of error (*PE*) and the Hausdorff distance to measure the error of the segmentation. The probability of error (*PE*) [16] is defined by:

$$PE = P(O)P(B|O) + P(B)P(O|B), \quad (18)$$

where $P(O)$ and $P(B)$ are a priori probabilities of object (nucleus or whole cell) and background in an image, respectively. $P(B|O)$ is the probability of classifying object as a background, whereas, $P(O|B)$ is the probability of classifying background as an object. We always report *PE* as a percentage of error, hence, we multiply *PE* with 100.

The Hausdroff distance (*H*) [43] is used to measure the effectiveness of the segmentation method. It is defined by:

$$H(A, B) = \max(h(A, B), h(B, A)), \quad (19)$$

where $h(A,B) = \max_{a\in A}\min_{b\in B}||a-b||$ is the directed Hausdroff distance from A to B. $A = \{a_1, a_2, \ldots, a_p\}$ is a set of finite points of expert segmentation and $B = \{b_1, b_2, \ldots, b_p\}$ is a set of finite of algorithm segmentation.

2.6.2. *Evaluation measures for classification*

For the classification evaluation, we consider 3 scenarios, i.e., 2-class, 4-class, and 7-class classification problems. The applications of some evaluation measures are straightforward in a 2-class problem (detection problem). However, their applications in a 4- or 7-class problem may not be straightforward. Therefore, the evaluation measures used to evaluate classification performance of each classifier need to be described here. Six measures were considered in this research including accuracy, sensitivity, specificity, Spearman rank-order correlation coefficient (R_s) [27], the Cohen's kappa coefficient (κ) [27], and the weighted kappa coefficient (κ_w) [27].

Accuracy is the percentage of data that are correctly classified to the correct true class, i.e., normal as a normal class or abnormal as an abnormal class in the 2-class problem. In the 4-class problem, it is directly applied to the correctly classified data in normal, LSIL, HSIL, and SCC classes. In 7-class problem, it is the percentage of correctly classified data in superficial squamous, intermediate squamous, columnar, mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ. Accuracy for a C-class problem is defined as:

$$Accuracy = \frac{\sum_{i=1}^{C} \text{No. of correctly classified cells in class } i}{\sum_{i=1}^{C} \text{Total no. of cells in class } i}. \quad (20)$$

Sensitivity is the percentage of the abnormal data that are correctly classified as abnormal (true positive). It is defined as:

$$Sensitivity = \frac{TP}{TP + FN}, \quad (21)$$

where TP is true positive and FN is false negative (the number of data in abnormal class classified as normal). The sensitivity calculation in the 2-class problem is straightforward. However, in the 4-class problem, we need to regroup LSIL, HSIL, and SCC to form one abnormal class. In the 7-class problem, we group superficial squamous, intermediate squamous, columnar as one normal class and abnormal class is from mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ. After that, the sensitivity calculation is the same as in the 2-class problem.

Specificity is the percentage of the normal data that are correctly classified as normal (true negative). It is defined as:

$$Specificity = \frac{TN}{TN + FP}, \quad (22)$$

where TN is true negative and FP is false positive (the number of data in normal class classified as abnormal). Similar to the sensitivity calculation, the specificity calculation in the 2-class problem is straightforward, whereas LSIL, HSIL, and SCC need to be regrouped as one abnormal class in the 4-class problem. Normal class is from superficial squamous, intermediate squamous, columnar and mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ are grouped as abnormal class in 7-class problem.

Spearman rank-order correlation coefficient (R_s) [27] is the correlation between ground truth ranking (U_i) and algorithm ranking (V_i) of cell, $i = 1,2,\ldots,N$, where N is the total number of cells. If both rankings are highly correlated, the magnitude of R_s will be close to 1. The sign of R_s is the indication of the direction of the relation between both rankings. Normally, if cells are in the same class, they will have a same rank. R_s is defined as:

$$R_s = \frac{\sum_{i=1}^{N}(U_i - \bar{U})(V_i - \bar{V})}{\sqrt{\sum_{i=1}^{N}(U_i - \bar{U})^2}\sqrt{\sum_{i=1}^{N}(V_i - \bar{V})^2}}, \quad (23)$$

where $\overline{U}$ and $\overline{V}$ are the rank means of ground truth ranking and algorithm ranking, respectively.

The Cohen's kappa coefficient (κ) [27], is a chance-correlated agreement between desired class and assigned class. This coefficient can be measured from the number of data (N_{ij}) labeled as desired class i and assigned class j from the confusion matrix. The degree of similarity between classes i and j is denoted by weight $\mathbf{W} = [W_{ij}]_{C\times C}$ where C is the number of classes. $W_{ij} \in [0,1]$ and $W_{ij} = 1$ for $i = j$. If $i \neq j$, W_{ij} will be close or equal to 0 if both classes are very different. The weighted relative observed agreement between desired class and assigned class is:

$$P_o = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{C} W_{ij}N_{ij}. \quad (24)$$

The weighted relative agreement expected just by chance is:

$$P_e = \frac{1}{N^2}\sum_{i=1}^{C}\sum_{j=1}^{C} W_{ij}r_i c_j, \quad (25)$$

where $r_i = \sum_{j=1}^{C} N_{ij}$ and $c_j = \sum_{i=1}^{C} N_{ij}$. The chance-correlated weighted relative agreement or the weighted kappa coefficient (κ_w) is defined by:

$$\kappa_w = \frac{P_o - P_e}{1 - P_e}. \quad (26)$$

If $\mathbf{W}$ is an identity matrix, κ_w will become the Cohen's kappa coefficient (κ). Both coefficients will be equal to 1 if both desired class and assigned class are in completely agreement. In the experiment, we set $\mathbf{W} = \mathbf{I}_2$ and $\mathbf{W} = \mathbf{I}_4$ for the 2-class and 4-class problems, respectively, where $\mathbf{I}_C$ is a $C \times C$

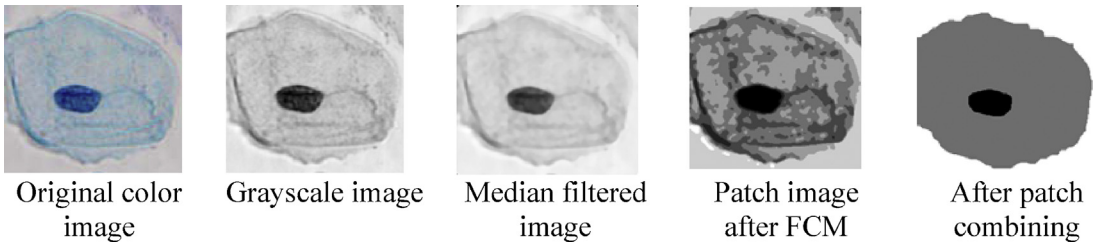

Fig. 2 – Sample segmentation using our method.

identity matrix. Whereas in the 7-class problem, according to Ref. [27], we set

$$\mathbf{W} = \begin{bmatrix} 1 & 0.5 & 0 & 0.25 & 0.25 & 0 & 0 \\ 0.5 & 1 & 0 & 0.25 & 0.25 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0 & 1 & 0.5 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0 & 0.5 & 1 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.25 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0 & 0.25 & 0.5 & 0.5 & 1 \end{bmatrix} \quad (27)$$

3. Results and discussion

For the sake of generalization of the classification results, the leave-one-out cross validation (LOOCV) was applied throughout the experiments. All results shown in this section are for the validation data in the LOOCV.

3.1. *Cervical cell segmentation*

The aim of cervical cell segmentation is to segment a cell into 3 regions, i.e., nucleus, cytoplasm, and background. As mentioned in the previous section, each cell image in the ERUDIT dataset does not cover the entire cell because only nucleus was analyzed. On the other hand, we collected the LCH dataset by considering the entire cell. Also, an entire cell is covered in the Herlev dataset. Therefore, we could segment cells in the Herlev and LCH datasets into 3 aforementioned regions, while only 2 regions (nucleus and non-nucleus) were segmented in the ERUDIT dataset. The parameters for the patch-based FCM clustering were: 7 clusters and the threshold for nucleus

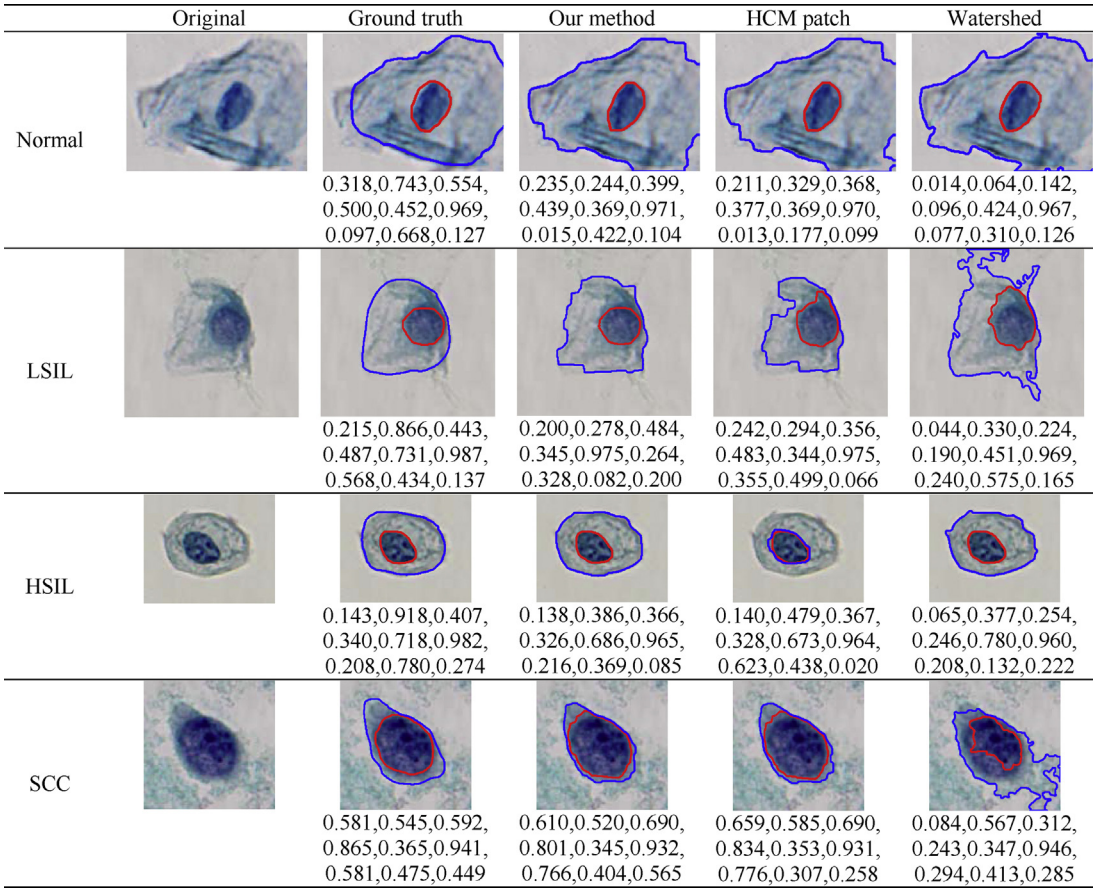

Fig. 3 – Sample of cell segmentation results with the 9 features for LCH dataset.

Table 1 – Nucleus and entire cell segmentation results for LCH dataset (mean ± STD).

		Precision	Recall	ZSI	PE (%)	Hausdroff distance
Nucleus						
Our method	Normal	0.85 ± 0.17	0.87 ± 0.20	0.83 ± 0.17	2.76 ± 4.47	14.97 ± 18.64
	LSI	0.85 ± 0.22	0.71 ± 0.27	0.70 ± 0.22	3.49 ± 3.70	41.49 ± 36.97
	HIS	0.82 ± 0.25	0.84 ± 0.25	0.81 ± 0.23	4.68 ± 5.83	18.36 ± 16.17
	SCC	0.95 ± 0.06	0.83 ± 0.17	0.87 ± 0.11	4.52 ± 6.13	17.82 ± 14.94
	Average	**0.87 ± 0.18**	**0.81 ± 0.22**	**0.81 ± 0.18**	**3.86 ± 5.03**	**23.16 ± 21.68**
HCM patch	Normal	0.87 ± 0.20	0.85 ± 0.22	0.82 ± 0.20	3.18 ± 5.56	15.75 ± 19.99
	LSI	0.88 ± 0.20	0.66 ± 0.27	0.70 ± 0.22	3.89 ± 4.28	40.78 ± 27.64
	HIS	0.85 ± 0.25	0.82 ± 0.26	0.81 ± 0.24	4.70 ± 5.31	19.47 ± 19.33
	SCC	0.96 ± 0.05	0.81 ± 0.17	0.86 ± 0.11	4.65 ± 5.78	18.41 ± 15.12
	Average	**0.89 ± 0.17**	**0.78 ± 0.23**	**0.80 ± 0.19**	**4.11 ± 5.23**	**23.60 ± 20.52**
Watershed	Normal	0.68 ± 0.40	0.73 ± 0.37	0.61 ± 0.40	10.09 ± 24.34	42.54 ± 67.33
	LSI	0.76 ± 0.37	0.65 ± 0.37	0.63 ± 0.37	8.20 ± 20.54	54.07 ± 60.78
	HIS	0.75 ± 0.34	0.73 ± 0.36	0.68 ± 0.32	6.58 ± 26.33	44.14 ± 44.62
	SCC	0.84 ± 0.25	0.70 ± 0.38	0.66 ± 0.33	25.31 ± 37.15	47.76 ± 51.76
	Average	**0.76 ± 0.34**	**0.70 ± 0.37**	**0.65 ± 0.36**	**15.04 ± 27.09**	**47.13 ± 56.12**
Entire cell						
Our method	Normal	0.84 ± 0.25	0.89 ± 0.17	0.81 ± 0.21	20.10 ± 17.29	77.85 ± 71.73
	LSI	0.92 ± 0.14	0.74 ± 0.31	0.76 ± 0.24	24.94 ± 30.34	85.11 ± 79.94
	HIS	0.76 ± 0.24	0.82 ± 0.26	0.71 ± 0.20	23.32 ± 23.19	63.21 ± 47.74
	SCC	0.79 ± 0.16	0.93 ± 0.11	0.83 ± 0.10	9.59 ± 7.00	40.31 ± 32.19
	Average	**0.82 ± 0.20**	**0.84 ± 0.21**	**0.78 ± 0.19**	**19.49 ± 19.45**	**66.62 ± 57.90**
HCM patch	Normal	0.83 ± 0.24	0.88 ± 0.17	0.81 ± 0.19	21.42 ± 17.08	80.17 ± 70.95
	LSI	0.92 ± 0.16	0.67 ± 0.33	0.70 ± 0.25	30.84 ± 31.51	96.98 ± 78.11
	HIS	0.80 ± 0.23	0.80 ± 0.27	0.73 ± 0.20	23.29 ± 24.17	64.17 ± 48.08
	SCC	0.84 ± 0.14	0.88 ± 0.16	0.84 ± 0.12	11.21 ± 14.79	42.62 ± 31.87
	Average	**0.85 ± 0.19**	**0.81 ± 0.23**	**0.77 ± 0.19**	**21.69 ± 21.89**	**70.99 ± 57.25**
Watershed	Normal	0.90 ± 0.22	0.86 ± 0.15	0.85 ± 0.17	18.40 ± 16.71	61.94 ± 50.14
	LSI	0.97 ± 0.08	0.82 ± 0.18	0.87 ± 0.13	10.01 ± 13.91	65.03 ± 56.21
	HIS	0.97 ± 0.05	0.78 ± 0.19	0.85 ± 0.12	14.75 ± 16.43	44.19 ± 29.06
	SCC	0.94 ± 0.17	0.63 ± 0.24	0.70 ± 0.19	27.85 ± 25.44	64.80 ± 31.16
	Average	**0.94 ± 0.13**	**0.77 ± 0.19**	**0.82 ± 0.15**	**17.76 ± 18.12**	**58.99 ± 41.64**

segmentation of 80% of the mean of cluster centers. Sample images at each step of our segmentation method are shown in Fig. 2. For the 3-region case in Herlev dataset, the parameters were the same as in the 2-region case, with an addition of the cytoplasm threshold of 120% of the mean of cluster centers. Whereas, in the LCH dataset, the threshold for nucleus segmentation of 90% of the mean was used and the cytoplasm threshold was the same as in the Herlev dataset. The morphological operations, i.e., opening and closing, were applied successively to smooth the edge of the segment and eliminate the small objects inside the image. The structuring element for these operations was a disk with 7-pixel radius. The disk shape was chosen for a structuring element because it would give smooth and round edge, whereas its size was determined from extensive experiments. For a comparison purpose, we utilized the hard C-means (HCM) clustering technique [44] and the watershed technique [45] in the segmentation process. Every step in the HCM-based segmentation method was exactly the same as in our FCM-based method, except FCM was replaced by HCM. Fig. 3 shows the examples of cell segmentation from our purposed method, the HCM clustering, and the watershed technique for the LCH dataset. The corresponding 9 features of each segmented image are also shown. It can be seen that the extracted features of segmented images from our and HCM methods are close to that of the ground truth. However, the features extracted using the watershed segmentation are much different from the others. This is because the watershed yields some badly segmented images, i.e., too big nuclei or too big entire cells. In fact, only a few badly segmented images can lead to a set of features in totally different scale because we divide each feature by its corresponding maximum value in the training set to normalize each feature to the range of 0–1. Table 1 shows the nucleus and entire cell segmentation evaluation on the LCH dataset. Whereas, Tables 2 and 3 show the nucleus and entire cell segmentation evaluation on Herlev dataset, respectively. It can be seen that the patch-based FCM clustering yields results very similar to the expert's opinions. In Table 2, we also compare our result indirectly with the nucleus segmentation mentioned in Refs. [24–27] for the Herlev dataset. The nucleus segmentation evaluation is comparable with that in Refs. [24–27]. Most of the evaluation measures on our nucleus segmentation are better than that on the HCM and the watershed segmentation. However, for the LCH entire cell segmentation, the watershed has high precision and *ZSI* values, and low *PE* and Hausdroff distance. But from the examples in Fig. 3(b), the segmentation result does not have a shape similar to the expert's opinion.

3.2. *Cervical cell classification for ERUDIT dataset*

Each cervical cell image in the ERUDIT dataset may not cover the entire cell. Therefore, only features extracted from nucleus

Table 2 – Nucleus segmentation results for Herlev dataset.

		Precision	Recall	ZSI	PE (%)	Hausdroff distance
Our method	Sup squa	0.95 ± 0.12	0.75 ± 0.33	0.78 ± 0.29	2.97 ± 10.51	14.71 ± 29.05
	Inter squa	0.98 ± 0.03	0.82 ± 0.25	0.86 ± 0.21	1.25 ± 3.30	13.12 ± 24.39
	Columnar	0.88 ± 0.20	0.78 ± 0.25	0.79 ± 0.19	8.25 ± 7.96	18.27 ± 17.89
	Mild dysp	0.80 ± 0.31	0.86 ± 0.26	0.79 ± 0.28	4.71 ± 5.69	18.69 ± 20.93
	Mod dysp	0.81 ± 0.25	0.88 ± 0.19	0.81 ± 0.21	7.66 ± 7.99	19.49 ± 17.68
	Severe dysp	0.79 ± 0.28	0.88 ± 0.21	0.79 ± 0.25	11.06 ± 12.63	17.26 ± 16.68
	CIS	0.70 ± 0.29	0.88 ± 0.23	0.75 ± 0.25	14.13 ± 12.02	17.64 ± 11.60
	Average	**0.85 ± 0.21**	**0.83 ± 0.25**	**0.80 ± 0.24**	**7.15 ± 8.59**	**17.03 ± 19.75**
HCM patch	Sup squa	0.96 ± 0.12	0.77 ± 0.31	0.81 ± 0.27	1.51 ± 4.83	12.10 ± 24.46
	Inter squa	0.97 ± 0.12	0.78 ± 0.29	0.83 ± 0.24	1.58 ± 4.16	16.13 ± 27.22
	Columnar	0.83 ± 0.26	0.79 ± 0.27	0.76 ± 0.23	9.00 ± 8.61	18.30 ± 18.01
	Mild dysp	0.75 ± 0.33	0.86 ± 0.28	0.76 ± 0.29	5.26 ± 6.56	18.33 ± 18.42
	Mod dysp	0.79 ± 0.26	0.89 ± 0.20	0.80 ± 0.21	8.19 ± 8.26	20.21 ± 18.58
	Severe dysp	0.78 ± 0.29	0.88 ± 0.22	0.78 ± 0.25	10.88 ± 11.88	17.05 ± 14.83
	CIS	0.68 ± 0.32	0.88 ± 0.23	0.72 ± 0.28	14.49 ± 13.78	18.23 ± 12.79
	Average	**0.82 ± 0.24**	**0.83 ± 0.26**	**0.78 ± 0.25**	**7.27 ± 8.30**	**17.19 ± 19.19**
Watershed	Sup squa	0.83 ± 0.37	0.70 ± 0.33	0.76 ± 0.34	0.39 ± 0.65	12.90 ± 23.32
	Inter squa	0.85 ± 0.35	0.72 ± 0.32	0.77 ± 0.34	2.09 ± 11.14	15.57 ± 29.99
	Columnar	0.57 ± 0.47	0.41 ± 0.40	0.43 ± 0.39	21.15 ± 19.55	40.81 ± 29.26
	Mild dysp	0.57 ± 0.43	0.68 ± 0.41	0.56 ± 0.40	11.51 ± 13.84	43.63 ± 38.79
	Mod dysp	0.57 ± 0.44	0.72 ± 0.40	0.55 ± 0.42	15.63 ± 16.12	37.15 ± 33.29
	Severe dysp	0.69 ± 0.39	0.77 ± 0.34	0.65 ± 0.37	17.47 ± 16.76	27.37 ± 26.67
	CIS	0.69 ± 0.40	0.80 ± 0.28	0.65 ± 0.37	20.88 ± 18.14	26.98 ± 22.85
	Average	**0.68 ± 0.41**	**0.69 ± 0.35**	**0.62 ± 0.38**	**12.73 ± 13.74**	**29.20 ± 29.17**
Result in [27]	Sup squa	0.69 ± 0.37	0.63 ± 0.37	0.98 ± 0.12	N/A	N/A
	Inter squa	0.79 ± 0.29	0.73 ± 0.31	0.98 ± 0.12	N/A	N/A
	Columnar	0.85 ± 0.15	0.77 ± 0.18	0.98 ± 0.05	N/A	N/A
	Mild dysp	0.88 ± 0.17	0.86 ± 0.16	0.96 ± 0.16	N/A	N/A
	Mod dysp	0.91 ± 0.10	0.86 ± 0.14	0.97 ± 0.07	N/A	N/A
	Severe dysp	0.90 ± 0.12	0.89 ± 0.11	0.95 ± 0.13	N/A	N/A
	CIS	0.89 ± 0.15	0.90 ± 0.08	0.92 ± 0.17	N/A	N/A
	Average	**0.88 ± 0.15**	**0.93 ± 0.15**	**0.89 ± 0.15**	**N/A**	**N/A**
Result of [24] reproduced in [27]	Sup squa	0.92 ± 0.12	0.88 ± 0.14	0.98 ± 0.02	N/A	N/A
	Inter squa	0.95 ± 0.03	0.92 ± 0.06	0.98 ± 0.02	N/A	N/A
	Columnar	0.83 ± 0.16	0.76 ± 0.20	0.97 ± 0.08	N/A	N/A
	Mild dysp	0.92 ± 0.13	0.90 ± 0.16	0.96 ± 0.08	N/A	N/A
	Mod dysp	0.89 ± 0.15	0.87 ± 0.17	0.94 ± 0.13	N/A	N/A
	Severe dysp	0.88 ± 0.15	0.90 ± 0.13	0.90 ± 0.19	N/A	N/A
	CIS	0.84 ± 0.18	0.88 ± 0.11	0.86 ± 0.24	N/A	N/A
	Average	**0.83 ± 0.20**	**0.96 ± 0.13**	**0.87 ± 0.19**	**N/A**	**N/A**

are valid. We tried to use the first 5 features including the area, compactness, major axis, minor axis, and aspect ratio of nucleus. This set of features represents shape and size of nucleus. We also added one more coarseness-based feature (feature 6: homogeneity of nucleus) to see how it affects the classification.

3.2.1. *Four-class classification for ERUDIT dataset*

The accuracy, specificity, and sensitivity of all 5 classifiers using 5 and 6 features are shown in Table 4. It can be seen that the best accuracy of 96.20% is achieved when using ANN with 5 features. The specificity and sensitivity in this case are 94.93% and 100%, respectively. Whereas, R_s, κ, and κ_w are 0.94, 0.95 and 0.95, respectively. In case of adding homogeneity feature of nucleus, ANN gives the best accuracy of 95.11% with 99.76% sensitivity and 91.30% specificity. R_s, κ, and κ_w are 0.91, 0.93, and 0.93, respectively. This additional homogeneity feature of nucleus can be considered as useful feature. We also show the classification results of the segmented data from the HCM clustering and the watershed technique in Table 5. Both segmentation methods do not yield a comparable result with the one segmented using our method.

3.2.2. *Two-class classification for ERUDIT dataset*

When the 2-class problem is considered, the best accuracy of 97.83% is achieved also using ANN with 5 features as shown in Table 6. The specificity, sensitivity, R_s, κ, and κ_w in this case are 94.20%, 99.03%, 0.94, 0.94, and 0.94, respectively. The accuracies of 4 classifiers (LDA, KNN, ANN and SVM) are more than 93%. However, if we add the homogeneity feature of nucleus, the accuracy of the same 4 classifiers are more than 94% and the accuracy from Bayes is more than 90%. This means the proposed features are very efficient regardless of the type of classifier. The homogeneity feature also plays an important role in enhancing the accuracy of Bayes, LDA, and ANN. Again, we compare the classification result with the data segmented using the HCM clustering technique and watershed technique

Table 3 – Entire cell segmentation results for Herlev dataset.

		Precision	Recall	ZSI	PE (%)	Hausdroff distance
Our method	Sup squa	0.99 ± 0.04	0.84 ± 0.13	0.90 ± 0.08	14.23 ± 11.73	95.99 ± 63.10
	Inter squa	0.97 ± 0.12	0.87 ± 0.12	0.91 ± 0.13	11.83 ± 12.63	68.94 ± 47.42
	Columnar	0.95 ± 0.07	0.72 ± 0.15	0.81 ± 0.10	24.15 ± 11.69	40.13 ± 17.53
	Mild dysp	0.96 ± 0.07	0.78 ± 0.11	0.85 ± 0.07	21.32 ± 9.77	69.32 ± 31.26
	Mod dysp	0.95 ± 0.07	0.77 ± 0.12	0.85 ± 0.08	21.92 ± 10.43	50.57 ± 17.39
	Severe dysp	0.93 ± 0.08	0.80 ± 0.11	0.85 ± 0.07	20.64 ± 9.37	37.87 ± 15.28
	CIS	0.93 ± 0.08	0.81 ± 0.11	0.86 ± 0.06	20.58 ± 8.63	34.14 ± 11.70
	Average	**0.95 ± 0.08**	**0.80 ± 0.12**	**0.86 ± 0.08**	**19.24 ± 10.61**	**56.71 ± 29.10**
HCM patch	Sup squa	0.99 ± 0.04	0.82 ± 0.14	0.89 ± 0.09	15.79 ± 13.33	99.32 ± 61.97
	Inter squa	0.98 ± 0.09	0.83 ± 0.16	0.88 ± 0.13	16.19 ± 16.57	84.64 ± 63.41
	Columnar	0.93 ± 0.09	0.73 ± 0.16	0.80 ± 0.10	25.18 ± 12.42	38.87 ± 17.69
	Mild dysp	0.95 ± 0.08	0.78 ± 0.12	0.85 ± 0.07	22.16 ± 10.22	68.34 ± 32.02
	Mod dysp	0.93 ± 0.09	0.77 ± 0.12	0.84 ± 0.08	22.99 ± 10.75	50.35 ± 19.66
	Severe dysp	0.91 ± 0.10	0.80 ± 0.11	0.85 ± 0.08	21.52 ± 9.93	37.62 ± 15.78
	CIS	0.92 ± 0.10	0.81 ± 0.11	0.85 ± 0.07	20.69 ± 9.08	33.10 ± 11.19
	Average	**0.95 ± 0.08**	**0.79 ± 0.13**	**0.85 ± 0.09**	**20.64 ± 11.76**	**58.89 ± 31.67**
Watershed	Sup squa	0.96 ± 0.05	0.88 ± 0.10	0.91 ± 0.06	11.42 ± 8.09	65.04 ± 41.17
	Inter squa	0.96 ± 0.12	0.87 ± 0.14	0.91 ± 0.13	12.03 ± 13.41	54.31 ± 35.78
	Columnar	0.77 ± 0.22	0.76 ± 0.17	0.73 ± 0.13	29.58 ± 12.01	31.78 ± 19.49
	Mild dysp	0.74 ± 0.29	0.85 ± 0.14	0.74 ± 0.20	27.83 ± 16.50	52.75 ± 30.03
	Mod dysp	0.66 ± 0.27	0.86 ± 0.14	0.70 ± 0.17	31.90 ± 14.48	41.87 ± 22.11
	Severe dysp	0.58 ± 0.23	0.89 ± 0.16	0.66 ± 0.16	34.86 ± 12.13	32.76 ± 17.56
	CIS	0.63 ± 0.17	0.90 ± 0.13	0.72 ± 0.11	30.95 ± 10.60	24.19 ± 10.22
	Average	**0.76 ± 0.19**	**0.86 ± 0.14**	**0.77 ± 0.14**	**25.51 ± 12.46**	**43.24 ± 25.20**

as shown in Table 7. The best result from our segmentation method is again better.

The result from the 2-class problem is better than 4-class problem most of the time, no matter what segmentation method or what classifier is used. One of the reasons is that when we combine LSIL, HSIL and SCC into one abnormal group, the features computed from all cells in this group is very different from those in normal group. For example, the mean and the standard deviation of the area from our segmentation method in normal and abnormal groups are 0.056 ± 0.110 and 0.309 ± 0.172, respectively. But the mean and the standard deviation of the area from our segmentation method in LSIL, HSIL, and SCC are 0.312 ± 0.176, 0.284 ± 0.164, and 0.330 ± 0.173, respectively. These values are not very different from each other. Hence, it will be difficult to separate cells in these three classes. Other features behave the same way. Using HCM or watershed segmentation methods also yield features with similar pattern.

3.3. *Cervical cell classification for LCH dataset*

We collected each cervical cell image in the LCH dataset so that it covers the entire cell. Therefore, we can use the features extracted from the entire cell. We tried to use the first 6 features representing shape, size, and coarseness of nucleus like in the ERUDIT dataset. Because we also segmented cytoplasm, 3 more features (feature 7: N/C ratio, feature 8: compactness

Table 4 – Four-class classification performance on validation set of LOOCV for ERUDIT dataset segmented with our method.

# of features	Performance	Classifier				
		Bayes	LDA	KNN	ANN	SVM
5	Accuracy	86.78	86.41	79.17	**96.20**	87.86
	Sensitivity	91.79	95.17	100.00	**100.00**	100.00
	Specificity	91.30	87.68	87.68	**94.93**	86.23
	R_s	0.79	0.82	0.85	**0.94**	0.87
	κ	0.82	0.82	0.72	**0.95**	0.84
	κ_w	0.82	0.82	0.72	**0.95**	0.84
6	Accuracy	86.05	84.60	79.53	95.11	88.41
	Sensitivity	92.03	93.72	99.52	99.76	100.00
	Specificity	90.58	89.13	88.41	91.30	89.13
	R_s	0.78	0.80	0.85	0.91	0.88
	κ	0.81	0.79	0.73	0.93	0.85
	κ_w	0.81	0.79	0.73	0.93	0.85

The bold values indicate the best results achieved in each scenario of experiment.

Table 5 – Four-class classification performance on validation set of LOOCV for ERUDIT dataset segmented with HCM and watershed.

# of features	Performance	Classifier									
		Bayes		LDA		KNN		ANN		SVM	
		HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed
5	Acc.	71.01	64.31	74.64	62.86	76.63	74.46	89.31	**84.96**	78.99	72.83
	Sens.	93.48	88.89	94.44	77.54	99.76	96.14	98.31	**96.38**	99.52	92.75
	Spec.	85.51	93.48	86.96	92.75	84.78	89.13	84.78	**92.03**	84.78	89.86
	R_s	0.75	0.72	0.76	0.61	0.81	0.82	0.86	**0.90**	0.82	0.80
	κ	0.61	0.52	0.66	0.50	0.69	0.66	0.86	**0.80**	0.72	0.64
	κ_w	0.61	0.52	0.66	0.50	0.69	0.66	0.86	**0.80**	0.72	0.64
6	Acc.	74.28	61.96	76.63	64.49	76.99	73.91	**90.40**	83.51	78.44	74.64
	Sens.	93.96	83.57	93.00	77.05	99.76	96.14	**99.03**	95.17	99.03	93.48
	Spec.	85.51	94.20	89.13	94.20	84.78	89.13	**84.78**	90.58	84.06	91.30
	R_s	0.75	0.71	0.77	0.64	0.81	0.82	**0.88**	0.88	0.81	0.83
	κ	0.66	0.49	0.69	0.53	0.69	0.65	**0.87**	0.78	0.71	0.66
	κ_w	0.66	0.49	0.69	0.53	0.69	0.65	**0.87**	0.78	0.71	0.66

The bold values indicate the best results achieved in each scenario of experiment.

of entire cell, and feature 9: area of entire cell) were added to evaluate their usefulness.

3.3.1. *Four-class classification for LCH dataset*

The accuracy, specificity, and sensitivity of all 5 classifiers using 5–9 features are shown in Table 8. When it is stated that a classifier uses *n* features, it means the first *n* features based on the aforementioned order are used. The best accuracy of 95.00% is achieved when using ANN with 9 features. The specificity and sensitivity in this case are 95.33% and 98.67%, respectively. R_s, κ, and κ_w are 0.94, 0.93 and 0.93, respectively. These 3 additional features increase the accuracies of Bayes, LDA, KNN, and SVM. However, they are still not as good as that of ANN. The results from HCM and watershed segmentation methods are also shown in Table 9. It should be noted that we performed the experiments with 5–9 features. However, the best results from both segmentation methods were achieved when the number of features was 9. To save space, we show only the results when 9 features are utilized. Again, the best result from both methods is not comparable with our segmentation method.

3.3.2. *Two-class classification for LCH dataset*

The 2-class classification results of the 5 classifiers using LOOCV for the LCH dataset is shown in Table 10. ANN with 9 features yields the best accuracy of 97.00%. The specificity and sensitivity are 98.00% and 96.00%, respectively. R_s, κ, and κ_w are all equal to 0.94. The features based on the entire cell help increasing accuracies in all classifiers. The classification result for the data segmented using HCM clustering technique and watershed technique is also shown in Table 11. To save space, only the numbers of features yielding the best results for HCM clustering (5 features) and watershed technique (8 features) are shown. The classification performance from HCM segmentation is lower than that using our segmentation method. But

Table 6 – Two-class classification performance on validation set of LOOCV for ERUDIT dataset segmented with our method.

# of features	Performance	Classifier				
		Bayes	LDA	KNN	ANN	SVM
5	Accuracy	87.50	93.84	96.38	**97.83**	96.74
	Sensitivity	86.23	96.86	100.00	**99.03**	99.28
	Specificity	91.30	84.78	85.51	**94.20**	89.13
	R_s	0.71	0.83	0.90	**0.94**	0.91
	κ	0.70	0.83	0.90	**0.94**	0.91
	κ_w	0.70	0.83	0.90	**0.94**	0.91
6	Accuracy	90.58	94.02	96.38	97.10	97.10
	Sensitivity	90.34	96.14	100.00	98.79	99.28
	Specificity	91.30	87.68	85.51	92.03	90.58
	R_s	0.77	0.84	0.90	0.92	0.92
	κ	0.76	0.84	0.90	0.92	0.92
	κ_w	0.76	0.84	0.90	0.92	0.92

The bold values indicate the best results achieved in each scenario of experiment.

Table 7 – Two-class classification performance on validation set of LOOCV for ERUDIT dataset segmented with HCM and watershed.

# of features	Performance	Classifier									
		Bayes		LDA		KNN		ANN		SVM	
		HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed
5	Acc.	87.14	85.51	93.12	82.25	96.01	95.11	95.83	**95.83**	94.20	91.30
	Sens.	86.96	82.37	95.17	89.86	100.0	97.58	99.03	**98.31**	97.10	90.58
	Spec.	87.68	94.93	86.96	59.42	84.06	87.68	86.23	**88.41**	85.51	93.48
	R_s	0.69	0.69	0.82	0.51	0.89	0.87	0.89	**0.89**	0.84	0.79
	κ	0.69	0.67	0.82	0.51	0.89	0.87	0.88	**0.89**	0.84	0.78
	κ_w	0.69	0.67	0.82	0.51	0.89	0.87	0.88	**0.89**	0.84	0.78
6	Acc.	90.94	84.78	93.12	82.97	96.01	95.11	**96.56**	95.47	94.57	92.75
	Sens.	91.79	81.40	94.69	89.37	100.0	97.58	**99.52**	97.83	97.58	92.75
	Spec.	88.41	94.93	88.41	63.77	84.06	87.68	**87.68**	88.41	85.51	92.75
	R_s	0.77	0.68	0.82	0.54	0.89	0.87	**0.91**	0.88	0.85	0.82
	κ	0.77	0.65	0.82	0.54	0.89	0.87	**0.90**	0.88	0.85	0.82
	κ_w	0.77	0.65	0.82	0.54	0.89	0.87	**0.90**	0.88	0.85	0.82

The bold values indicate the best results achieved in each scenario of experiment.

Table 8 – Four-class classification performance on validation set of LOOCV for LCH dataset segmented with our method.

# of features	Performance	Classifier				
		Bayes	LDA	KNN	ANN	SVM
5	Accuracy	82.33	79.67	75.33	88.00	83.00
	Sensitivity	93.33	82.67	69.33	94.67	81.33
	Specificity	83.33	87.33	90.67	90.67	92.67
	R_s	0.77	0.73	0.74	0.83	0.80
	κ	0.74	0.69	0.61	0.82	0.74
	κ_w	0.74	0.69	0.61	0.82	0.74
6	Accuracy	77.33	82.67	75.33	89.67	83.67
	Sensitivity	94.67	86.00	69.33	94.00	81.33
	Specificity	70.00	90.00	90.00	91.33	95.33
	R_s	0.72	0.77	0.73	0.83	0.82
	κ	0.68	0.74	0.61	0.85	0.75
	κ_w	0.68	0.74	0.61	0.85	0.75
7	Accuracy	85.33	87.67	80.00	91.00	85.33
	Sensitivity	96.67	92.00	78.00	96.00	88.67
	Specificity	84.00	95.33	94.67	96.00	95.33
	R_s	0.85	0.88	0.84	0.91	0.88
	κ	0.79	0.81	0.69	0.86	0.78
	κ_w	0.79	0.81	0.69	0.86	0.78
8	Accuracy	85.67	88.33	81.00	90.67	85.33
	Sensitivity	95.33	92.00	78.67	93.33	87.33
	Specificity	86.00	95.33	95.33	94.67	94.67
	R_s	0.85	0.88	0.84	0.89	0.87
	κ	0.79	0.82	0.70	0.86	0.78
	κ_w	0.79	0.82	0.70	0.86	0.78
9	Accuracy	87.33	91.67	83.67	**95.00**	87.33
	Sensitivity	96.67	95.33	76.67	**98.67**	84.67
	Specificity	87.33	94.67	96.67	**95.33**	96.67
	R_s	0.88	0.91	0.85	**0.94**	0.88
	κ	0.81	0.87	0.74	**0.93**	0.80
	κ_w	0.81	0.87	0.74	**0.93**	0.80

The bold values indicate the best results achieved in each scenario of experiment.

Table 9 – Four-class classification performance on validation set of LOOCV for LCH dataset segmented with HCM and watershed.

# of features	Performance	Classifier									
		Bayes		LDA		KNN		ANN		SVM	
		HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed
9	Acc.	86.33	86.33	89.67	87.33	82.00	87.67	**93.33**	89.67	88.33	**91.00**
	Sens.	94.67	98.00	92.00	92.67	83.33	94.67	**94.67**	97.33	90.67	**98.00**
	Spec.	88.67	93.33	98.00	98.67	96.00	99.33	**96.67**	99.33	96.67	**100.0**
	R_s	0.89	0.88	0.93	0.91	0.88	0.93	**0.94**	0.95	0.92	**0.96**
	K	0.80	0.80	0.84	0.81	0.72	0.81	**0.90**	0.84	0.82	**0.86**
	κ_w	0.80	0.80	0.84	0.81	0.72	0.81	**0.90**	0.84	0.82	**0.86**

The bold values indicate the best results achieved in each scenario of experiment.

the performance from watershed segmentation is higher than that using our segmentation method. However, the classification performance from watershed in 4-class problem is much lower than that using our segmentation method. The reason for this is that the feature values in 2-class problem are very different especially in case of watershed segmentation, but not in 4-class problem. Please be noted that from Fig. 3 and Table 1, the watershed segmentation result is not good comparing to our segmentation method.

3.4. *Cervical cell classification for Herlev dataset*

In the Herlev dataset, the entire cell is presented as in the LCH dataset. Therefore, we can use the features extracted from the entire cell similar to that used in the LCH dataset.

3.4.1. *Seven-class classification for Herlev dataset*

The 7-class classification results of the 5 classifiers using LOOCV for the Herlev dataset is shown in Table 12. ANN with 9

Table 10 – Two-class classification performance on validation set of LOOCV for LCH dataset segmented with our method.

# of features	Performance	Classifier				
		Bayes	LDA	KNN	ANN	SVM
5	Accuracy	79.00	81.33	78.33	92.33	89.00
	Sensitivity	71.33	73.33	70.67	92.67	88.00
	Specificity	86.67	89.33	86.00	92.00	90.00
	R_s	0.59	0.63	0.57	0.85	0.78
	κ	0.58	0.63	0.57	0.85	0.78
	κ_w	0.58	0.63	0.57	0.85	0.78
6	Accuracy	88.00	81.00	79.33	92.67	89.33
	Sensitivity	92.00	72.67	71.33	93.33	88.67
	Specificity	84.00	89.33	87.33	92.00	90.00
	R_s	0.76	0.63	0.59	0.85	0.79
	κ	0.76	0.62	0.59	0.85	0.79
	κ_w	0.76	0.62	0.59	0.85	0.79
7	Accuracy	92.00	89.33	87.33	96.00	92.67
	Sensitivity	92.00	84.00	80.00	97.33	91.33
	Specificity	92.00	94.67	94.67	94.67	94.00
	R_s	0.84	0.79	0.75	0.92	0.85
	κ	0.84	0.79	0.75	0.92	0.85
	κ_w	0.84	0.79	0.75	0.92	0.85
8	Accuracy	92.67	89.00	87.67	96.00	92.00
	Sensitivity	95.33	83.33	80.67	96.67	90.00
	Specificity	90.00	94.67	94.67	95.33	94.00
	R_s	0.85	0.79	0.76	0.92	0.84
	κ	0.85	0.78	0.75	0.92	0.84
	κ_w	0.85	0.78	0.75	0.92	0.84
9	Accuracy	94.67	89.67	88.00	**97.00**	93.67
	Sensitivity	96.00	85.33	79.33	**96.00**	93.33
	Specificity	93.33	94.00	96.67	**98.00**	94.00
	R_s	0.89	0.80	0.77	**0.94**	0.87
	κ	0.89	0.79	0.76	**0.94**	0.87
	κ_w	0.89	0.79	0.76	**0.94**	0.87

The bold values indicate the best results achieved in each scenario of experiment.

Table 11 – Two-class classification performance on validation set of LOOCV for LCH dataset segmented with HCM and watershed.

# of features	Performance	Classifier									
		Bayes		LDA		KNN		ANN		SVM	
		HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed
5	Acc.	89.00	97.00	89.67	96.00	82.00	97.33	**96.33**	97.33	87.33	97.67
	Sens.	88.00	98.67	86.67	93.33	78.67	99.33	**94.67**	96.67	86.00	98.67
	Spec.	90.00	95.33	92.67	98.67	85.33	95.33	**98.00**	98.00	88.67	96.67
	R_s	0.78	0.94	0.79	0.92	0.64	0.95	**0.93**	0.95	0.75	0.95
	K	0.78	0.94	0.79	0.92	0.64	0.95	**0.93**	0.95	0.75	0.95
	κ_w	0.78	0.94	0.79	0.92	0.64	0.95	**0.93**	0.95	0.75	0.95
8	Acc.	89.00	97.00	91.67	96.67	89.67	96.33	95.33	**99.33**	91.67	98.33
	Sens.	86.67	97.33	86.00	94.67	81.33	96.67	95.33	**100.0**	89.33	98.00
	Spec.	91.33	96.67	97.33	98.67	98.00	96.00	95.33	**98.67**	94.00	98.67
	R_s	0.78	0.94	0.84	0.93	0.80	0.93	0.91	**0.99**	0.83	0.97
	K	0.78	0.94	0.83	0.93	0.79	0.93	0.91	**0.99**	0.83	0.97
	κ_w	0.78	0.94	0.83	0.93	0.79	0.93	0.91	**0.99**	0.83	0.97

The bold values indicate the best results achieved in each scenario of experiment.

Table 12 – Seven-class classification performance on validation set of LOOCV for Herlev dataset segmented with our method.

# of features	Performance	Classifier				
		Bayes	LDA	KNN	ANN	SVM
5	Accuracy	68.48	71.43	70.56	83.32	79.50
	Sensitivity	84.15	95.11	96.15	96.30	92.89
	Specificity	84.71	69.83	71.07	81.82	90.50
	R_s	0.56	0.68	0.69	0.78	0.77
	κ	0.63	0.66	0.65	0.80	0.76
	κ_w	0.62	0.67	0.67	0.80	0.77
6	Accuracy	69.90	73.72	71.76	84.62	81.03
	Sensitivity	90.37	96.44	96.15	97.19	93.04
	Specificity	88.84	74.38	73.14	88.43	93.39
	R_s	0.68	0.74	0.71	0.83	0.79
	κ	0.65	0.68	0.66	0.82	0.77
	κ_w	0.68	0.71	0.68	0.83	0.79
7	Accuracy	70.01	74.92	68.92	85.71	79.17
	Sensitivity	90.22	96.89	97.33	97.78	92.74
	Specificity	88.43	80.99	70.66	89.67	91.32
	R_s	0.71	0.77	0.71	0.84	0.77
	κ	0.65	0.70	0.63	0.83	0.75
	κ_w	0.68	0.73	0.65	0.84	0.76
8	Accuracy	75.25	78.84	73.50	86.80	80.70
	Sensitivity	91.26	96.00	96.89	98.07	91.56
	Specificity	89.26	78.93	66.53	86.78	88.43
	R_s	0.75	0.76	0.72	0.85	0.77
	κ	0.71	0.75	0.68	0.84	0.77
	κ_w	0.73	0.75	0.67	0.85	0.77
9	Accuracy	82.33	80.26	73.28	**93.78**	85.39
	Sensitivity	95.56	96.89	94.96	**98.96**	94.22
	Specificity	95.04	82.64	66.53	**96.69**	92.56
	R_s	0.86	0.81	0.76	**0.94**	0.88
	κ	0.79	0.76	0.68	**0.93**	0.83
	κ_w	0.82	0.78	0.67	**0.93**	0.83

The bold values indicate the best results achieved in each scenario of experiment.

Table 13 – Confusion matrix of validation set of LOOCV in 7-class classification for Herlev dataset segmented with our method using neural networks with 9 features.

		Predicted						
		Sup squamous	Inter squamous	Columnar	Mild dysplasia	Moderate dysplasia	Severe dysplasia	Carcinoma in situ
Actual	Sup squamous	68	3	2	0	0	0	1
	Inter squamous	1	62	4	0	1	1	1
	Columnar	0	3	91	1	0	1	2
	Mild dysplasia	0	0	1	173	1	7	0
	Moderate dysplasia	1	1	0	1	139	3	1
	Severe dysplasia	0	0	1	2	3	189	2
	Carcinoma in situ	0	0	3	2	4	3	138

features yields the best accuracy of 93.78%. The specificity and sensitivity are 96.69% and 98.96%, respectively. R_s, κ, and κ_w are 0.94, 0.93, and 0.93, respectively. Again, the 6th feature and these 3 additional features help increasing the performance in all classifiers. The confusion matrix of the best classifier is also shown in Table 13. We choose to show this confusion matrix because it represents the experiment on the largest number of classes in this work. The predicted outputs tend to be the last 4 classes due to the unbalancing of the dataset. The numbers of cells in the last 4 classes are added up to be about 74% of that of the entire dataset. To remedy this problem, we may have to modify the classifier or the training process [46]. However, the overall classification performance is still very good here. Again, we compare the result with the best results from HCM and watershed segmentation methods as shown in Table 14. In this case, both HCM clustering technique and watershed technique achieved the best results when the number of features is 9. The results show again that our segmentation method yields features that help in the classification performance. We also compare our results indirectly with Ref. [27]. In Ref. [27], R_s, κ, and κ_w are 0.675, 0.265, and 0.328, respectively. Although, the segmentation performance of our method and [27] are comparable, our classification performance is better.

3.4.2. *Two-class classification for Herlev dataset*

The 2-class classification results of the 5 classifiers using LOOCV for the Herlev dataset are shown in Table 15. ANN with 9 features yields the best accuracy of 99.27%. The specificity and sensitivity are 96.53% and 99.85%, respectively. R_s, κ, and κ_w are all equal to 0.97. The comparison with HCM and watershed segmentation methods is also shown in Table 16. In the table, only the best results from HCM clustering (using 9 features) and watershed technique (using 6 features) are shown. Again, the best result is from the one with our segmentation method. We also compare our results indirectly with Ref. [27], In Ref. [27], R_s, κ, and κ_w are 0.848, 0.848, and 0.848, respectively. Once again, our classification performance is better.

Although the first 5 features based on shape and size of nucleus already provide good classification results, all results indicate that the homogeneity feature of nucleus can enhance the accuracies of all classifiers. It is impossible to evaluate the usefulness of cytoplasm (or entire cell) information in the ERUDIT dataset because each image in the dataset covers nucleus only. Fortunately, we collected the LCH dataset so that each image covers the entire cervical cell. In the Herlev dataset, each cell image covers the entire cell as well. The results show that the 3 additional cytoplasm-based features, including the N/C ratio, compactness, and area of entire cell, once again help enhancing the accuracies of the classifiers. However, using only nucleus information, the classification rate is close to 90% in all results. Hence, only the nucleus information is adequate to classify cervical cells. This is similar to what we found in white blood cell classification problem in that only nucleus information is adequate to classify white blood cells in bone marrow [46]. This finding is very useful because nucleus segmentation is an easier problem than cytoplasm (or entire cell) segmentation. This issue becomes even more serious when adjacent cells are touching or overlapping.

Table 14 – Seven-class classification performance on validation set of LOOCV for Herlev dataset segmented with HCM and watershed.

# of features	Performance	Classifier									
		Bayes		LDA		KNN		ANN		SVM	
		HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed
9	Acc.	75.25	65.54	71.54	63.47	70.99	68.92	**88.88**	**85.39**	83.53	76.12
	Sens.	92.74	91.11	93.48	89.33	96.74	91.26	**98.07**	**96.59**	95.70	88.74
	Spec.	92.56	69.42	75.21	65.70	74.38	67.77	**87.60**	**83.47**	91.74	85.54
	R_s	0.84	0.69	0.79	0.60	0.82	0.65	**0.92**	**0.84**	0.90	0.74
	K	0.71	0.59	0.66	0.56	0.65	0.63	**0.87**	**0.83**	0.80	0.72
	κ_w	0.75	0.60	0.69	0.57	0.70	0.63	**0.88**	**0.83**	0.82	0.72

The bold values indicate the best results achieved in each scenario of experiment.

Table 15 – Two-class classification performance on validation set of LOOCV for Herlev dataset segmented with our method.

# of features	Performance	Classifier				
		Bayes	LDA	KNN	ANN	SVM
5	Accuracy	79.98	83.27	94.63	94.75	93.41
	Sensitivity	78.22	100.00	98.96	97.63	94.67
	Specificity	88.19	4.86	74.31	81.25	87.50
	R_s	0.54	0.20	0.81	0.81	0.79
	κ	0.49	0.08	0.80	0.81	0.78
	κ_w	0.49	0.08	0.80	0.81	0.78
6	Accuracy	90.11	84.00	95.60	97.44	95.12
	Sensitivity	89.63	100.00	98.96	98.96	95.11
	Specificity	92.36	9.03	79.86	90.28	95.14
	R_s	0.72	0.27	0.84	0.91	0.85
	κ	0.71	0.14	0.84	0.91	0.84
	κ_w	0.71	0.14	0.84	0.91	0.84
7	Accuracy	85.71	94.87	96.58	97.31	94.38
	Sensitivity	84.15	97.93	98.96	98.96	94.81
	Specificity	93.06	80.56	85.42	89.58	92.36
	R_s	0.64	0.82	0.88	0.91	0.82
	κ	0.61	0.82	0.88	0.91	0.82
	κ_w	0.61	0.82	0.88	0.91	0.82
8	Accuracy	85.96	94.51	95.97	97.07	94.14
	Sensitivity	84.44	97.63	98.37	98.96	94.37
	Specificity	93.06	79.86	84.72	88.19	93.06
	R_s	0.65	0.80	0.86	0.90	0.82
	κ	0.62	0.80	0.86	0.90	0.81
	κ_w	0.62	0.80	0.86	0.90	0.81
9	Accuracy	89.38	94.14	94.99	**99.27**	95.36
	Sensitivity	88.15	98.37	97.48	**99.85**	95.11
	Specificity	95.14	74.31	83.33	**96.53**	96.53
	R_s	0.72	0.79	0.82	**0.97**	0.86
	κ	0.69	0.78	0.82	**0.97**	0.85
	κ_w	0.69	0.78	0.82	**0.97**	0.85

The bold values indicate the best results achieved in each scenario of experiment.

Table 16 – Two-class classification performance on validation set of LOOCV for Herlev dataset segmented with HCM and watershed.

# of features	Performance	Classifier									
		Bayes		LDA		KNN		ANN		SVM	
		HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed	HCM	Watershed
6	Acc.	89.50	88.40	91.94	87.91	94.87	95.36	96.46	**97.92**	95.36	95.60
	Sens.	89.33	87.26	95.85	97.63	98.37	98.37	98.52	**99.70**	96.44	96.44
	Spec.	90.28	93.75	73.61	42.36	78.47	81.25	86.81	**89.58**	90.28	91.67
	R_s	0.70	0.69	0.71	0.52	0.82	0.83	0.88	**0.93**	0.84	0.85
	K	0.69	0.67	0.71	0.49	0.81	0.83	0.87	**0.93**	0.84	0.85
	κ_w	0.69	0.67	0.71	0.49	0.81	0.83	0.87	**0.93**	0.84	0.85
9	Acc.	92.31	91.94	90.96	90.35	94.38	93.89	**98.05**	97.19	96.95	94.51
	Sens.	92.15	91.41	96.00	95.56	98.07	96.74	**99.26**	98.96	97.33	95.70
	Spec.	93.06	94.44	67.36	65.97	77.08	80.56	**92.36**	88.89	95.14	88.89
	R_s	0.77	0.77	0.67	0.65	0.80	0.79	**0.93**	0.90	0.90	0.82
	K	0.76	0.76	0.67	0.65	0.80	0.79	**0.93**	0.90	0.90	0.82
	κ_w	0.76	0.76	0.67	0.65	0.80	0.79	**0.93**	0.90	0.90	0.82

The bold values indicate the best results achieved in each scenario of experiment.

4. Conclusions

This research proposes a method of automatic cervical cell image segmentation and classification. We used the over-segment fuzzy C-means clustering technique to segment each cell into 2 or 3 regions. Three Pap smear datasets, i.e., ERUDIT, LCH, and Herlev, were tested. The 4-class problem consists of 4 cell classes, i.e., normal, low grade squamous intraepithelial lesion (LSIL), high grade squamous intraepithelial lesion (HSIL), and squamous cell carcinoma (SCC). When the last 3 classes are grouped as one abnormal class, the problem is changed to the 2-class problem in the ERUDIT and LCH datasets. Whereas, the 7-class problem in the Herlev dataset consists of 7 cell classes, i.e., superficial squamous, intermediate squamous, columnar, mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma in situ. The first 2 classes are grouped as one normal class, whereas the last 4 classes are grouped as one abnormal class [27]. For the ERUDIT dataset, ANN with 5 nucleus-based features yielded the accuracies of 96.20% and 97.83% on the 4-class and 2-class problems, respectively. For the Herlev dataset, ANN with 9 cell-based features yielded the accuracies of 93.78% and 99.27% for the 7-class and 2-class problems, respectively. For the LCH dataset, ANN with 9 cell-based features yielded the accuracies of 95.00% and 97.00% for the 4-class and 2-class problems, respectively. We also compare the classification results with the data segmented by the hard C-means clustering technique and watershed technique. All the results show that our segmentation method provides a better set of features for the classifiers. Not only accuracy, in this particular problem, sensitivity is extremely important because it is the indication of the fatal false negatives. From the results, the features extracted from our segmentation method also provide better sensitivity than that from the hard C-means and watershed segmentation methods.

In this work, we implement the segmentation and classification system for a single cervical cell. Our ongoing work is to extend the proposed method to cope with an image with multiple cells. Up to now, the new method can segment all nuclei and perform well in the classification process. We are certain that our method will be able to perform adequately on an image with multiple cells in general. With satisfactory feedbacks from doctors at Lampang Cancer Hospital who provided us a dataset, we plan to deploy our method to aid them in cervical cancer screening at the hospital in this near future. We will also test our method on Pap smear images stored in a system similar to MITIS [47].

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

Financial support from the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0238/2550) to Thanatip Chankong and Nipon Theera-Umpon is acknowledged. We thank Dr. Taweethong Koanantakool, Department of Medical Services, Ministry of Public Health, Thailand, for introducing us the cervical cancer classification problem. We would like to thank Dr. Jan Jantzen for supporting the ERUDIT and Herlev Pap smear datasets. We are thankful to Lampang Cancer Hospital, Thailand, for its supported data collection and technical knowledge. We also thank anonymous reviewers for their valuable comments.

REFERENCES

[1] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, D. Forman, Global cancer statistics, CA: A Cancer Journal for Clinicians 61 (2) (2011) 69–90.
[2] H.C. Kitchenera, P.E. Castleb, J.T. Coxc, Achievements and limitations of cervical cytology screening, Vaccine 24 (S3) (2006) 63–70.
[3] P. Attasara, R. Buasom (Eds.), Hospital-Based Cancer Registry, vol. 27, National Cancer Institute, Department of Medical Services, Ministry of Public Health, 2011.
[4] International Health Policy Program Thailand (IHPP) and Health Intervention and Technology Assessment Program (HITAP), Research for Development of an Optimal Policy Strategy for Prevention and Control of Cervical Cancer in Thailand, Ministry of Public Health, Thailand, 2008.
[5] Central Intelligence Agency, Thailand Demographics Profile 2012, CIA World Factbook, Directorate of Intelligence, 2012.
[6] R. Ashfaq, B. Solares, M.H. Saboorian, Detection of endocervical component by PAPNET™ system on negative cervical smears, Diagnostic Cytopathology 15 (2) (1996) 121–123.
[7] J.S. Lee, L. Kuan, S. Oh, F.W. Patten, D.C. Wilbur, A feasibility study of the AutoPap system location-guided screening, Acta Cytologica 42 (1) (1998) 221–226.
[8] H. Doornewaard, Y.T. van der Schouw, Y. van der Graaf, Reproducibility in double scanning of cervical smears with the PAPNET system, Acta Cytologica 44 (4) (2000) 604–610.
[9] F.A. Kreuger, M. van Ballegooijen, H. Doornewaard, Is PAPNET suitable for primary screening? Lancet 353 (9162) (1999) 1374–1375.
[10] D.V. Coleman, Evaluation of automated systems for the primary screening of cervical smears, Current Diagnostic Pathology 5 (2) (1998) 57–64.
[11] C.V. Biscotti, A.E. Dawson, B. Dziura, L. Galup, T. Darragh, A. Rahemtulla, L. Wills-Frank, Assisted primary screening using the automated ThinPrep imaging system, American Journal of Clinical Pathology 123 (2) (2005) 281–287.
[12] S.M. Freire, R.T. de Almeida, M.D.B. Cabral, E.A. Bastos, R.C. Souza, M.G.P. da Silva, A record linkage process of a cervical cancer screening database, Computer Methods and Programs in Biomedicine 108 (2013) 90–101.
[13] G. Rahmadwati, M. Naghdy, C. Ros, E. Todd, Norachmawati classification cervical cancer using histology images, in: International Conference on Computer Engineering and Applications, 2010, pp. 515–519.
[14] H.G. Acosta-Mesa, B. Zitova, H.V. Rios-Figueroa, N. Cruz-Ramirez, A. Marin-Hernandez, R. Hernandez-Jimenez, B.E. Cocotle-Ronzon, E. Hernandez-Galicia, Cervical cancer detection using colposcopic images: a temporal approach, in: Mexican International Conference on Computer Science, 2005, pp. 158–164.
[15] O.E. Meslouhi, M. Kardouchi, H. Allali, T. Gadi, Semi-automatic cervical cancer segmentation using active contours without edges, in: International Conference on Signal-Image Technology and Internet-Based Systems, 2009, pp. 54–58.

[16] C.G. Loukas, A. Linney, A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia, and vasculature, Computer Methods and Programs in Biomedicine 74 (2004) 183–199.
[17] L. He, L.R. Long, S. Antani, G.R. Thoma, Histology image analysis for carcinoma detection and grading, Computer Methods and Programs in Biomedicines 107 (2012) 538–556.
[18] J. Kong, O. Sertel, H. Shimada, K.L. Boyer, J.H. Saltz, M.N. Gurcan, Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation, Pattern Recognition 42 (2009) 1080–1092.
[19] C. Bergmeir, M.G. Silvente, J.M. Benitez, Segmentation of cervical cell nuclei in high-resolution microscopic images: a new algorithm and a web-based software framework, Computer Methods and Programs in Biomedicine 107 (2012) 497–512.
[20] P. Malm, B.N. Balakrishnan, V.K. Sujathan, R. Kumar, E. Bengtsson, Debris removal in Pap-smear images, Computer Methods and Programs in Biomedicine 111 (2013) 128–138.
[21] M.E. Plissiti, C. Nikou, A. Charchanti, Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images, Pattern Recognition Letters 32 (2011) 838–853.
[22] M.E. Plissiti, C. Nikou, A. Charchanti, Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering, IEEE Transactions on Information Technology in Biomedicine 15 (2) (2011) 233–241.
[23] M.-H. Tsai, Y.-K. Chan, Z.-Z Lin, S.-F. Yang-Mao, P.-C. Huang, Nucleus and cytoplast contour detector of cervical smear image, Pattern Recognition Letters 29 (2008) 1441–1453.
[24] K. Li, Z. Lu, W. Liu, J. Yin, Cytoplasm and nucleus segmentation in cervical smear images using radiating GVF snake, Pattern Recognition 45 (2012) 1255–1264.
[25] P.-Y Pai, C.-C Chang, Y.-K Chan, Nucleus and cytoplast contour detector from a cervical smear image, Expert Systems with Applications 39 (2012) 154–161.
[26] J. Fan, R. Wang, S. Li, C. Zhang, Automated cervical cell image segmentation using level set based active contour model, in: 12th International Conference on Control, Automation, Robotics and Vision, 2012.
[27] A. Genctav, S. Aksoy, S. Onder, Unsupervised segmentation and classification of cervical cell images, Pattern Recognition 45 (2012) 4151–4168.
[28] N. Theera-Umpon, White blood cell segmentation and classification in microscopic bone marrow images, Lecture Notes in Computer Science 3614 (2005) 787–796.
[29] T. Chankong, N. Theera-Umpon, S. Auephanwiriyakul, Cervical cell classification using segmented-nucleus features, in: International Symposium on Biomedical Engineering, 2009.
[30] E.R. Dougherty, An Introduction to Morphological Image Processing, SPIE Press, Bellingham, WA, 1992.
[31] T. Masayoshi, Color Atlas of Cancer Cytology, 3rd ed., Igaku-shoin, Tokyo, 2000.
[32] B. Alberts, A. Johnson, J. Lewis, M. Raff, Molecular Biology of the Cell, Garland Science, New York, 2007.
[33] R.A. Weinberg, The Biology of Cancer, Garland Science, Taylor & Francis, New York, 2007.
[34] B.S. Ducatman, H.H. Wang (Eds.), The Pap Smear: Controversies in Practice, Arnold, London, 2002.
[35] Tulane University, Neoplasia–Hyperchromasia, available at: http://tulane.edu/som/departments/pathology/training/neoplasia_image_12.cfm
[36] L.G. Koss, C. Gompel, C. Bergeron, Introduction to Gynecologic Cytopathology: With Histologic and Clinical Correlations, Lippincott Williams & Wilkins, Baltimore, 1999.
[37] R. Schalkoff, Pattern Recognition: Statistical, Structural and Neural Approaches, John Wiley & Sons, Singapore, 1992.
[38] S. Haykin, Neural Networks and Learning Machines, 3rd ed., Pearson Education, New York, 2009.
[39] V.N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed., Springer-Verlag, New York, 2000.
[40] J. Jantzen, The ERUDIT Papsmear Tutorial, available at: http://fuzzy.iau.dtu.dk/smear/
[41] E. Martin, Pap-Smear Classification (Master's Thesis), Technical University of Denmark, 2003.
[42] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin, A.C. Palmer, Morphometric analysis of white matter lesion in MR images: method and validation, IEEE Transactions on Medical Imaging 13 (4) (1994) 716–724.
[43] D.P. Huttenlocher, G.A. Klanderman, G.A. Kl, W.J. Rucklidge, Compare images using the Hausdroff distance, IEEE Transactions on Pattern Analysis and Machine Intelligence 5 (1993) 850–863.
[44] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[45] P. Soille, Morphological Image Analysis Principles and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
[46] N. Theera-Umpon, S. Dhompongsa, Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification, IEEE Transactions on Information Technology in Biomedicine 11 (3) (2007) 353–359.
[47] G.K. Matsopoulos, V. Kouloulias, P. Asvestas, N. Mouravliansky, K. Delibasis, D. Demetriades, MITIS: a WWW-based medical system for managing and processing gynaecological–obstetrical–radiological data, Computer Methods and Programs in Biomedicine 76 (2004) 53–71.