Abstract—Cytologic screening has been widely used for detecting the cervical cancers. In this study, a semiautomatic PC-based cellular image analysis system was developed for segmenting nuclear and cytoplasmic contours and for computing morphometric and textual features to train support vector machine (SVM) classifiers to classify four different types of cells and to discriminate dysplastic from normal cells. A software program incorporating function, including image reviewing and standardized denomination of file names, was also designed to facilitate and standardize the workflow of cell analyses. Two experiments were conducted to verify the classification performance. The cross-validation results of the first experiment showed that average accuracies of 97.16% and 98.83%, respectively, for differentiating four different types of cells and in discriminating dysplastic from normal cells have been achieved using salient features (8 for four-cluster and 7 for two-cluster classifiers) selected with SVM recursive feature addition. In the second experiment, 70% (837) of the cell images were used for training and 30% (361) for testing, achieving an accuracy of 96.12% and 98.61% for four-cluster and two-cluster classifiers, respectively. The proposed system provides a feasible and effective tool in evaluating cytologic specimens.

Index Terms—Cervical cancer, cytology, image analysis, Pap smear, support vector machines (SVMs).

I. INTRODUCTION

CYTOLOGY evaluation is a safe, efficient, and well-established technique for the diagnoses of many diseases. The most famous success in cytology is its ability to reduce the mortality and morbidity of cervical cancer through mass screening. It was reported that the invasive cancer incidence decreased by 47.8% after national screening from 1995 to 2006 in Taiwan [1]. One role of cytology is directed to early detection of dysplasia or preinvasive cancer cells. Once the abnormal cells are detected, the patient can be scheduled for a biopsy examination and subsequent surgical treatment. Consequently, the progression of the cancer can be stopped at an early stage.

Classical cytological diagnosis is based on microscopic observation of specialized cells and qualitative assessment using descriptive criteria, which may be inconsistent because of subjective variability of different observers [2]. To lower the false negative rate in screening, many advanced technologies involving sampling, smear preparation, or screening quality control have been developed and introduced [3]–[6]. Commercial devices that use these technologies can be divided into the following categories based on their approaches: 1) improved slide preparation to reduce sampling error, e.g., thin-layered liquid-based preparation (ThinPrep, SurePath, Tripath) [7], [8]; 2) reduced workload and screening error as in the autoscreening system (ThinPrep Imaging System, Cytex, Boxborough, MA; FocalPoint System, Tripath Imaging, Burlington, NC); 3) improved laboratory quality control like rescreening (Papnet) [9]; and 4) enhanced quality assurance, such as the proficiency test [10]. However, most of these devices do not assist objective diagnosis by providing the calculable variables that would eliminate interpretation errors and interobserver discrepancy [11]. In addition, they are not applicable to the general cytological laboratory because of high cost and technical or linguistic gaps [6], [12].

Thus, without a reproducible and quantitative tool, observer bias is still an unsolved problem in the routine cytological laboratory. Diagnostic divergence caused by visual observation remains.

The technique of computerized image analysis used to assist artificial diagnosis of cell abnormalities or tumors in cytopathology or histopathology also can provide accurate and objective evaluation of nuclear morphology. Quantitative methods for estimating a cytological specimen can be traced back some 30 years and are still continuing to develop [9], [13]–[19]. Reliability, accessibility, cost, efficiency, technical maintenance, and linguistic communication are considerations that need to be taken into account in any new designs. Due to revolution and evolution of new technologies, enhanced power of computation,
decreased cost of hardware and software, and prevalence of the Internet, more and more systems being developed use computational algorithms for cellular image analysis [20]–[22]. Such approaches promise to resolve the limitation of subjective analysis, especially in the fields of bioinformatics, biology, and medicine.

Traditional cytological criteria for differentiating dysplastic cells from normal cells are based on the change in nucleus to cytoplasm (N/C) ratio, nuclear size, irregularity of nuclear shape and nuclear membrane, and density and granularity of nuclear chromatin. Most of these criteria are descriptive and relatively subjective. In contrast, in computed morphometry, the subjective criteria are replaced by quantitative, calculable, and comparative variables [23]–[25]. This study aimed at setting up a reproducible and reliable analytical tool to facilitate interpretation and to create a reliable database of cell images for conducting cytopathological educations and designing clinical diagnostic systems. Generally, there are three types of cervical squamous cells, i.e., superficial, intermediate, and parabasal cells, observed on the Pap smear images. The endocervical columnar epithelia cells are occasionally found, sometimes in small clusters, in the Pap smear samples. We evaluated the cell morphometric variables by using statistical analysis and other feature-selection methods, such as filter and wrapper methods, to select salient variables for the design of classifiers and construction of predictive models to automatically classify different cell types and to discriminate the dysplastic cells from normal cells, respectively. The designed multiple-cluster classifier is valuable in the classification of cell types for constructing cell image database and is useful in applications such as cytological and pathological educations, whereas the constructed predictive model is shown to be able to assist pathologists to increase the diagnostic performance of dysplasia cells.

Support vector machine (SVM) is a supervised learning method widely used for classification [26], [27]. A special property of SVM is that it can simultaneously minimize the empirical classification error and maximize the geometric margin of a classifier. It is a powerful methodology for solving problems in nonlinear classification, function estimation, and density estimation, leading to many applications including image interpretation, data mining, biometric authentication, biotechnological investigation, and other electrical applications. [28]–[32]. It has also been used for automatic detection of cell nuclei on Pap smear images [23].

Recently, automatic [23], [24], [33] and semiautomatic [25], [34] methods have been proposed to detect nuclear contours of cervical cells on the Pap smear images and to discriminate the normal from abnormal dysplasia cells. However, it was argued that cytoplasm contains important features for the discrimination of abnormal cells [35]. More recently, classification of different types of cervical cells has been pinpointed [36], [37]. In the study of Sokouti et al. [36], the cervical cells were classified into normal, low-grade squamous intraepithelial lesion (LSIL), and high-grade squamous intraepithelial lesion (HSIL), but not considering different types of normal cervical cells. Moreover, Genctav et al. [37] classified seven types of cells, including two types of cervical squamous (superficial and intermediate) cells, one type of endocervical (columnar) cell, three types of dysplasia (mild, moderate, and severe) cells, and carcinoma in situ, based on unsupervised classification. Since the columnar cells are occasionally found in the Pap smear images and the carcinoma in situ is in an early stage of cancer, early detection of dysplasia cells has been generally focused in the cytopathological setting of preventive medicine. In this study, in addition to discriminating the abnormal dysplasia from the normal cells, we also purport to classify the cervical cells into dysplasia and three different types of normal cells. The preliminary results were reported in Huang et al. [38].

In our previous work, only 503 cells (27%) selected from 1814 acquired cervical cells with peer-agreement were used for manual segmentation of the nuclear and cytoplasmic contours (NCCs), which were then applied as the masks to calculate morphologic and textural features to design a two-cluster classifier for discriminating abnormal from normal cells and a multiple-cluster classifier for classifying four different types of cervical cells based on the SVM [38]. However, it was argued that 27% images selected were not representative and might render the study highly biased. In this study, in order to eliminate such a bias, three pathologists were again invited to make further discussion regarding those images which did not reach agreement in categorization in the first round to convince each other to include more cases for analyses, resulting in a total of 1198 (66%) cell images selected for cell analysis and classifier design. Furthermore, we improved the previous work by developing a semiautomatic system integrating a NCC detector for automatically detecting NCCs and a manual random walk (RW) detector for manual segmentation to further improve the performance of contour detection for those cells which were not well detected by the NCC detector.

Moreover, two experiments were conducted to verify performance of the SVM classifiers. In the first experiment, twofold cross validation was used to obtain the SVM parameters, followed by tenfold cross validation to verify performance of the classifiers. Filter and wrapper methods were used to select salient features for building the SVM classifiers. On the other hand, in the second experiment, the dataset was randomly divided into training set and testing set, which contain 837 (70%) and 361 (30%) cells, respectively. The SVM classifiers were trained with salient features selected using wrapper methods based on the training set, and tested with the cell images in the testing set. The rest of this paper is organized as follows. Section II describes the materials and methods proposed and adopted in this study. Section III demonstrates the experimental results. Finally, brief discussion and conclusion are made in Sections IV and V, respectively.

II. MATERIALS AND METHODS

All the analysis and experiments are conducted based on the designed programs and the equipment in a routine cytologic laboratory. Here, we describe a prototypic system composed of a microscope, digital camera, personal computer, automated cellular processing and analyzing programs, and SVM classifier.

The flowchart of the overall experimental procedure is illustrated in Fig. 1. It can be divided into the following steps:
acquisition and categorization of cell images, image editing and processing, contour segmentation of cell nucleus and cytoplasm, measurement and analysis of cellular morphology and texture, classification of cell types using SVM, and assessment of diagnostic performance. Customized programs were designed in our laboratory to simplify the analysis procedures, facilitate cellular image review, organize the file names of cell images, and calculate the morphometric features of cells. Additionally, statistical significance of the selected morphometric and textural features, performance between the SVM and the Fisher linear discriminant classifiers, and efficacy among different feature-selection methods were also compared.

**A. Acquisition and Categorization of Cell Images**

Cytological images were captured using a high-resolution digital camera (Olympus C-5060, Japan) mounted on a microscope (Olympus BX 51, Japan) and stored as digital format. Cell images were captured with three different scales of magnification (100x, 200x, and 400x) by excluding the images with minimal magnification, extensive cellular overlapping, and interference by other inflammatory cells or debris. A total of 1814 cell images were captured from 42 Pap smears. Fig. 2 demonstrates the program designed for image interpretation and file name standardization and renaming. An image can be inspected easily in 100x, 200x, and 400x magnification simultaneously by pathological personnel. All the images were renamed using a scheme based on cell type, standardized sequence number, magnification, requesting pathological number, and diagnostic code for easy documentation and retrieval by the analysis programs and SVM classifier.

**B. Image Editing and Processing**

Extraction of individual cellular images and generation of their corresponding nucleus and cytoplasm masks were achieved by using a designed program. Its friendly graphic user interface makes the image processing and editing tasks easy to operate for clinical laboratory personnel. First, the program automatically adjusted the color of the image on the computer to a visual view very similar to the view on the microscope. Second, individual cells were selected with the provided cropping tool, followed by the segmentation of NCCs. Fig. 3 illustrates the procedure in automatically generating nuclear and cytoplasmic masks of a selected cell. An extracted cell image and its masks were saved.
as two different but associated files, which were then used for calculating nuclear and cytoplasmic variables by an automated analysis program. Fig. 3(a) demonstrates a cell cropped from a Pap smear image by the designed program for detecting contours using the integrated NCC detector [39] and RW detector [40]. The NCC detector, as detailed in the following section, is used to automatically detect the cell contours. If the user is not satisfied with the segmented contours, the RW detector is further applied for adjusting the contours by drawing sketchy curves inside nucleus, in cytoplasm, or outside the cell, respectively, to mark textures within different cell regions. For images containing multiple cells, the NCC detector was not able to accurately detect the NCCs. The RW detector could be used to effectively resolve these cases. As illustrated in Fig. 3(b), the segmented regions of the nucleus and cytoplasm are converted into masks represented in blue and green, respectively. The masks were then used to locate the nuclear and cytoplasmic regions on the original cell image to automatically calculate morphometric and textural variables using the analyzing program. Fig. 3(c) illustrates the mask overlaid on a cell image to facilitate the calculation of nucleus and cytoplasm variables. The graphical user interface of the automated analysis program is shown in Fig. 3(d).

C. Automatic NCC Detection

The automatic NCC detection includes two consecutive phases: cytoplasmic contour detection followed by nuclear contour detection. Although Otsu’s method is effective in the determination of the threshold value to separate an object from the background, it cannot provide a proper threshold when the standard deviations or the number of data are significantly different between two groups, i.e., object and background [41]. The adaptable threshold decision (ATD) method, adopting standard deviations of groups, numbers of data, and group intervals, was demonstrated to be able to determine the optimal threshold to effectively separate the cytoplasm from the background.

The nucleus contour detection phase consists of three steps, including gradient calculation, nuclear contour detection, and contour connection. First, Sobel operator was used to calculate gradients of the cell image. It was found that Sobel operator is very sensitive to noise; hence, the mean vector difference enhancer was applied to highlight the gradient of the object contour and to suppress the gradient of the noise contour [42]. Second, since the nucleus of a cervical cell on the Pap smear image is generally much smaller than its cytoplasm and the background of the cropped image, especially for a normal cell, the ATD method cannot provide an optimal threshold to precisely separate the nucleus from the cytoplasm. Therefore, the method based on maximum gray-level gradient difference (MGLGD) was adopted for detecting the nuclear contour [39]. The main idea behind the MGLGD is that the difference of mean gray levels between the region inside and the region outside the contour is maximal at the contour of an object. In addition, the gradients of most pixels located on the contour are much larger than the gradients of other pixels. Finally, the cytoplasmic contour obtained in the cytoplasm detection phase was used to determine the initial contour by moving ten pixels toward the nucleus for detecting the contour of cell nucleus. In each iteration progressing to the final nuclear contour, the new contour may not form a closed boundary. In this case, an additional pixel was added to close the contour to step forward the next iteration. Fig. 4 illustrates the results obtained from individual steps. Although the NCC detector is promising in detecting NCCs of most single-cell images, it cannot effectively segment overlapping cells or cells deteriorated by artifacts. The RW detector can be used to cope with the aforementioned problems. As demonstrated in Fig. 5(a), NCC detector detects the cytoplasmic contours covering two cytoplasmic contours. After being adjusted by an RW detector, the NCCs are effectively detected.
points on the nucleus or cytoplasm perimeter. After \( L \) has been determined, the maximum axle width \( W \) can be obtained by searching the line, with one point located at the left side and the other one at the right side of the maximum axle, perpendicular to the maximum axle and has the longest width. The axle center (AC) is the intersection between the maximum axle and the minimum axle. The CG can be obtained by accounting the coordinate of individual pixels \((p_x, p_y)\) on the perimeter, i.e.,

\[
CG = \left( \frac{\sum_{i=1}^{N} p_x^i}{N}, \frac{\sum_{i=1}^{N} p_y^i}{N} \right)
\]

(1)

where \( N \) indicates the number of points located on the perimeter. The MAP, AAP, MGP, and AGP can be calculated from the following equation:

\[
MAP = \text{Max}_{i} \left[ (AC_x - p_x^i)^2 + (AC_y - p_y^i)^2 \right]
\]

(2a)

\[
AAP = \sum_{i=1}^{N} \left[ (AC_x - p_x^i)^2 + (AC_y - p_y^i)^2 \right] / N
\]

(2b)

\[
MGP = \text{Max}_{i} \left[ (CG_x - p_x^i)^2 + (CG_y - p_y^i)^2 \right]
\]

(2c)

\[
AAP = \sum_{i=1}^{N} \left[ (CG_x - p_x^i)^2 + (CG_y - p_y^i)^2 \right] / N.
\]

(2d)

Initially, five Haralick features [47], i.e., entropy, energy, contrast, homogeneity, and linearity, and three Tamura features [48], i.e., coarseness, contrast, and roughness, were considered. In addition, roundness, compactness, and elongation proposed by Yonekawa et al. [49] were also adopted. After statistical analysis, two Haralick features, i.e., entropy of the co-occurrence matrix (ECM) and contrast of the co-occurrence matrix (CCM), and two Tamura features, i.e., coarseness and contrast, were also applied to analyze textures of the cell nucleus. A co-occurrence matrix is a 2-D \( n \times n \) matrix obtained from an image \( I \). The element in location \((i, j)\) of the co-occurrence matrix indicates the number of occurrences for a pixel \( q \) with color value \( i \) in image \( I \) having a pixel \( q^* \) with color value \( j \) in its neighborhood situated at a distance of \( d \) pixels away. The entropy and contrast of an image can be calculated from the co-occurrence matrix according to the following equations:

\[
\text{Entropy} = -\sum_{i} \sum_{j} P(i, j) \log P(i, j)
\]

(3)

\[
\text{Contrast} = \sum_{i} \sum_{j} (i - j)^2 P(i, j)
\]

(4)

where \( P(i, j) \) indicates the probability of a pixel with gray-level \( i \) having a pixel with gray-level \( j \) adjacent to it. In this study, four different distances \((d = 1, 2, 3,\) and \(4\)) and four different directions \((\theta = 0^\circ, 45^\circ, 90^\circ,\) and \(135^\circ)\) were tested for obtaining the combination of \( d \) and \( \theta \) with best classification performance. It was observed that the performance changed in a range of 0.03–0.12%. The parameters with \( d = 1 \) and \( \theta = 45^\circ \) achieving the best performance were adopted.

---

**D. Measurement of Cellular Morphology and Texture**

A software program, namely personal computer-based cellular image analysis (PCCIA) system, was designed to perform the cellular (nucleus and cytoplasm) image analysis. PCCIA can automatically measure the morphometric variables of the cells either individually or in a batch. Fig. 5(d) illustrates the terms, including axle center, center of gravity (CG), perimeter, maximum length, and maximum width, defined for calculating morphometric variables [47]–[50]. With the help of internal calibration using the micrometer image, various variables, including nuclear perimeter, area, maximum axle length \( L \), maximum axle width \( W \), N/C ratio, maximum length from axle center to perimeter (MAP), average length from axle center to perimeter (AAP), maximum length from center of gravity to perimeter (MGP), and average length from center of gravity to perimeter (AGP), were obtained for the evaluation of the nuclear size and shape irregularity of a cell.

As shown in Fig. 5(d), the maximum axle length \( L \) is the length of the longest line among the lines connecting any two

---

**Fig. 5.** (a) Contours automatically detected by NCC and (b) contours adjusted by RW detector. (c) Detected nuclear and cytoplasmic morphometric variables. (d) Illustration of morphometric variables, including perimeter, axle center, CG, maximum axle length, and maximum axle width, defined for quantitative analysis of the cell shape.
Tamura features have been widely adopted for characterizing low-level statistical properties of texture [48]. In this study, coarseness and contrast were selected for effectively describing the texture within the cell nucleus. To evaluate coarseness, we first take the average for each pixel \((x, y)\) over its neighborhoods with a window size of \(2k \times 2k\) pixels:

\[
A_k(x, y) = \frac{1}{(2k+1)^2} \sum_{i=x-2k}^{x+2k-1} \sum_{j=y-2k}^{y+2k-1} f(i, j)
\]

where \(k = 1, 2, \text{ and } 3\). Then, the differences between two pairs of nonoverlapped blocks with the size of \(2k \times 2k\) pixels at both sides of the pixel \((x, y)\) are taken horizontally and vertically:

\[
E_{k,h} = |A_k(x + 2k-1, y) - A_k(x - 2k-1, y)|
\]
\[
E_{k,v} = |A_k(x, y + 2k-1) - A_k(x, y - 2k-1)|.
\]

For each pixel, the optimal value of \(k\) that maximizes \(E\) is selected for the measurement of coarseness by taking the average of \(A_{opt}(x, y)\), as shown in the following equation:

\[
f_{crs} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} A_{opt}(i, j)
\]

where \(m \times n\) indicates the image size. In addition, the contrast can be obtained from the following equation:

\[
f_{con} = \frac{\sigma}{(\mu_4/\sigma^4)^{1/4}}
\]

in which \(\mu_4\) and \(\sigma\) indicate the fourth moment and the standard deviation of the image.

After having analyzed a cell image based on its accompanied masks, the evaluated variables are saved in a file with format compatible to Microsoft Excel and used for further processing. Thus, these organized documents can be used for manual interpretation and for computerized automatic analysis. Table I shows image examples of four different cell types and their corresponding morphometric and textural variables computed by PCCIA.

### E. Designing Classifiers Using SVM

It is widely believed that SVM is superior to traditional statistical and neural network classifiers. The goal of SVM is to separate multiple clusters with a set of unique hyperplanes that have the greatest margins to the edge of each cluster. On the other hand, hyperplanes that separate two clusters is not unique for other linear classifiers. For a two-cluster classification example, the hyperplane separating two classes that leaves the maximum margins from both clusters [51], [52]. Regarding a nonlinear classifier, various kernels including polynomial, radial basics function, and hyperbolic tangent can be used for mapping the original sample space into a new Euclidian space, and then the linear classifier can be designed for classification.

In the first experiment, statistical analyses of morphometric and textural variables were first conducted by using analysis of variance (ANOVA) and \(t\)-test to test significance \((p < 0.05)\) of individual variables. Variables that reach significance were selected as candidate features followed by other feature selection methods, as detailed in the next section, for selecting compatible variables to design a two-cluster classifier to detect dysplastic from normal cells and a four-cluster classifier to discriminate among different types of cells. Cells were randomly divided into two groups (two folds) for cross validation to train the optimal SVM parameters. These parameters were fixed at the validation stage, in which tenfold cross validation was applied to validate the performance of the classifier and the predictive model.

In the second experiment, the dataset containing 1198 cell images, each consisting of 29 nuclear variables and 29 cytoplasmic variables, was randomly divided into training and testing sets, in which the training and testing sets contain 837(70%) and 361(30%) cell images, respectively. In the training phase, tenfold cross validation was adopted to train the model with best performance using the training set, which was followed by the testing phase to verify the performance of the classifiers using the testing set.

The SVM software packages (LIBSVM) developed by Chang and Lin [53] was adopted for classification of different cell types. The selected features of the datasets have been normalized to meet the requirements of LIBSVM. The experimental procedure is summarized in the following.

1) Normalize the variables into the values between 0 and 1 according to the following equation:

\[
X'(n) = \frac{X(n) - \text{Min}}{\text{Max} - \text{Min}}
\]

where Max and Min indicates the maximal and minimum values, respectively, in the dataset; \(X(n)\) is the original value and \(X'(n)\), \(0 \leq X'(n) \leq 1\), is the value after normalization.

2) Label the normal and dysplastic cells with positive (superficial cell: \(y_i = +1\); intermediate cell: \(y_i = +3\); parabasal
cell: \( y_i = +5 \) and negative integers (dysplastic cells: \( y_i = -4 \)), respectively. All the variables corresponding to a cell were then saved as a sequence of data with the following format:

\[
[y] [i] = [v] [1] \ldots
\]

in which label (or class) is an integer indicating the category of the cells, whereas index is an integer representing the morphometric variable. For example, the sequence \( +1 \):0.708 2:0.987 \ldots \) indicates a superficial cell with the values of the first and second normalized variables of 0.708 and 0.987, respectively.

3) **Experiment 1:** In the training phased, cells in the dataset were randomly divided into two subsets, in which images of four different cell types were equally distributed, for cross validation to get the optimal SVM parameters, i.e., \( C \) (cost parameter) and \( \gamma \) (kernel parameter). During the testing phase, cell data were randomly divided into ten subsets (folds) for cross validation by fixing the SVM parameters to the values obtained at the training phase. In each iteration, nine folds were used for training the model, while the rest one was used for validation. The procedure was repeated for ten times, each using an individual fold for validation.

4) **Experiment 2:** The dataset was randomly divided into two independent sets, i.e., training and testing sets. In the training phase, cells in the training set were divided into ten subsets (folds) for cross validation to obtain the model with best performance. In the testing phase, cells in the testing set were used for verifying the performance of the SVM classifiers built in the training phase.

**F. Feature Selection**

Feature selection takes the advantages of reducing the number of features and the size of storage requirements, decreasing training and computational time, facilitating data visualization and understanding, and improving predictive performance [54], [55]. The algorithms of feature selection can often be classified into three approaches including filter, wrapper, and embedded methods [54]. The filter method is a preprocessing procedure that selects a subset of features based on statistical measures independent of the designed classifiers. In contrast, the wrapper method assesses individual subsets of features in a recursive way by considering their predictive efficacy to a given classifier. It is more computational intensive than the filter method, but is believed to be able to provide more efficient outcome. The subset with a smallest number of features achieving the highest predictive accuracy is used for classifier construction. Recently, a genetic algorithm (GA), an alternative wrapper method, was also proposed as an useful method for feature selection [56], [57]. Sometimes, this strategy is also used for the adjustments of cost value and kernel parameter of SVM together with the selection of features [58] when designing a classifier. For example, it was applied to construct predictive models for the diagnoses of breast cancers [59], [60] and hypertension [61]. On the other hand, embedded method selects features during the process of model construction by considering the cost function of a model [62].

1) **Filter Method Based on Statistical Analysis:** The filter method eliminates unfavorable features but reserves salient features by taking the advantages of low computation complexity. Measures such as information entropy, distance, dependence, and consistency are generally used to select salient features before classification [61]. This method only considers the relevance between an individual independent variable (feature) and the dependent variable by ranking the independent variables according to their relevance to the dependent variable. The interdependence among independent variables is not considered at all. The wrapper method, on the other hand, selects various subsets of features for directly feeding into the classifier to obtain the best combination of features with greatest accuracy. Although it is promising in obtaining a classifier with best performance, it is an NP-hard problem needing tremendous computation time [63], especially for datasets with a great number of features.

In this investigation, filter methods, including entropy [60] and F-score [64], were adopted for feature selection. Entropy is defined as

\[
E^k = \sum \sum s_{ij}^k \log s_{ij}^k + (1 - s_{ij}^k) \log(1 - s_{ij}^k)
\]

where \( s_{ij}^k \) is the similarity measure between any two samples \( i \) and \( j \) of feature \( k \) and is defined as \( s_{ij}^k = e^{-d_{ij}} \), with \( d_{ij} \) indicating the Euclidean distance of feature \( k \) between samples \( x_i \) and \( x_j \). As manifested in the definition, the greater the Euclidean distance, the smaller is the similarity between two samples. On the other hand, F-score is defined as

\[
F^k = \frac{(\mu_1^k - \mu_2^k)^2 + (\mu_2^k - \mu_2^k)^2}{(\sigma_1^2)^2 + (\sigma_2^2)^2}
\]

in which \( \mu_1^k, \mu_2^k \), and \( \mu_2^k \) indicate means of class 1, class 2, and combination of class 1 and class 2 samples, respectively; \( \sigma_1^2 \) and \( \sigma_2^2 \) represent standard deviations of samples in class 1 and class 2, respectively.

2) **Wrapper Method Based on Recursive Feature Elimination (RFE), Recursive Feature Addition (RFA), and GA:** The wrapper method assesses individual subsets of features in a recursive way by considering their predictive efficiency to a given classifier. For a vector space with \( n \) features, the RFE algorithm removes unimportant features based on backward sequential selection by iteratively deleting one feature at a time, resulting in a suboptimal combination of \( r \) (\( r < n \)) features with best predictive performance [54]. For SVM-RFE, it starts with all features by deleting a feature repeatedly until \( r \) features are left, which leads to a largest margin separating two classes. Weight magnitude, which is inversely proportional to the margin, is generally used as the ranking criterion in determining importance of individual features. The eliminated feature \( p \) is the one that minimizes the variation of weight

\[
\|w_p\| = \sum_{i,j=0}^{N} \lambda_i \lambda_j y_i y_j K(x_i, x_j).
\]

In addition to weight (or margin), other measures, such as generalization error [62], gradient of weight [65], and Fischer’s
ratio [66], were also proposed for feature ranking [62]. In this study, mean cross validation accuracy was used as a measure of feature ranking for determining the eliminated feature in each iteration.

In contrast to RFE, RFA is based on forward sequential selection by iteratively adding one feature at a time. It starts with the most important feature followed by repeatedly adding an additional feature at a time until \( r \) features have been selected leading to a largest margin separating two classes.

The GA can find optimal solution within an acceptable time, and is faster than dynamic programming using an exhaustive searching strategy. By taking the advantage of the GA in quickly searching the optimal features, a nonlinear hyperplane with greatest margin can be obtained by using SVM to classify two clusters. Classification of multiple clusters can be easily expanded. The GA was modified to combine with LIBSVM for feature selection and classification [57].

G. Assessment

The cytopathological experts’ classification of cells was assumed to be the gold standard in this investigation. For the four-cluster classifier, confusion matrix accompanied with calculated accuracy and error rates were used to assess its classification performance. To assess the diagnostic performance of the two-cluster model constructed by SVM, the accuracy, sensitivity, and specificity of the model were compared with the model based on the linear discriminant analysis (LDA) built using the SPSS software package. Performance indices are calculated according to the following formula:

\[
\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (13a)
\]

\[
\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (13b)
\]

\[
\text{Specificity} = \frac{TN}{(TN + FP)} \quad (13c)
\]

\[
\text{PPV} = \frac{TP}{(TP + FP)} \quad (13d)
\]

\[
\text{NPV} = \frac{TN}{(TN + FN)} \quad (13e)
\]

where TP, TN, FP, and FN indicate the numbers of true positive, true negative, false positive, and false negative, respectively; PPV and NPV represent positive predictive value and negative predictive value, respectively. In addition, area under ROC curve (AUC) and confusion matrix were also used for sensitivity and error analyses. The significance of individual variables in differentiating the normal (S, I, and P) from the abnormal (HSIL and LSIL) cells was assessed using Student’s \( t \)-test with significance of \( p < 0.05 \).

III. Experimental Results

Fig. 6 demonstrates examples of NCCs detected with the adopted NCC-RW methods and other methods proposed in previous investigations. Table II compares the performance of contour detection among different methods. The detected NCCs were used to generate masks for calculating morphological and textural features to design classifiers.

A. Experiment 1: Cross Validation

After statistical analysis, only the nuclear morphometric variables and N/C ratio that reach significant difference \( (p < 0.05) \) were chosen for further study, while all the other cytoplasmic variables, except N/C ratio, were excluded \( (p > 0.05) \). For descriptive analysis, means and standard deviations of the 12 nuclear morphometric variables and N/C ratio are summarized in Table III. With regard to dysplastic cells, the morphometric variables including perimeter (46.4 ± 14.4 μm), area (187.3 ± 96.1 μm²), maximum axle length (15.7 ± 4.2 μm), maximum axle width (12.3 ± 3.3 μm), N/C ratio (35.3% ± 50.6%), MAP (110.0 ± 31.5 μm), AAP (84.2 ± 21.5 μm), MGP (99.8 ± 26.9 μm), AGP (83.4 ± 21.3 μm), ECM (7.4 ± 0.4), CCM (48.6 ± 19.3), coarseness (81.5 ± 31.5), and contrast (41.6 ± 48.3) were all found to be significantly greater than those of
the normal cells \((p < 0.001)\). All the variables demonstrated the capability to differentiate among four different cell types \((p < 0.001, \text{ANOVA})\) and to discriminate dysplasia from pooled normal cells \((p < 0.001, t\text{-test})\). The statistical results show that the dysplastic cells have larger size \(i.e.,\) larger perimeter, area, maximum length and maximum width), higher nuclear proportion \(i.e.,\) larger N/C ratio), greater irregularity in nuclear shape \(i.e.,\) larger MAP, AAP, MGP, and AGP), and greater contrast.

The classification and diagnostic performance of the SVM classifiers and a comparison with the classifiers constructed with LDA are summarized in Table IV. The classification of morphometric variables using SVM reveals greater classification performance \(accuracy: 96.58\%, total\ error: 41\ cases\) in discriminating four different cell types compared to the methodology established using LDA \(accuracy: 90.48\%, total\ error: 114\ cases\). Regarding the diagnostic performance in detecting the dysplastic cells, SVM-based model presents higher sensitivity \(87.42\%\) with less classification errors while still satisfying satisfactory accuracy \(97.91\%)\) when compared to the LDA model \(sensitivity: 68.21\%, accuracy: 95.83\%). In addition, the area under ROC curve \(\text{AUC}\) for the model constructed using SVM \(\text{AUC} = 0.990\) is greater than the model constructed with LDA \(\text{AUC} = 0.984\).

As shown in Table III, although nuclear textual variable CCM can be used to discriminate dysplastic cell from pooled normal cells, there is no significant difference \((p > 0.05, t\text{-test})\) between the dysplastic cell and the parabasal cell. Furthermore, by concerning the discrimination among three types of normal \(\text{superficial, intermediate, and parabasal}\) cells, it was found that maximum length, maximum width, MAP, and MGP are not able to discriminate between intermediate and parabasal cells \((p > 0.05, t\text{-test})\). Hence, these five variables were excluded to redesign the four-cluster and two-cluster classifiers by using only eight variables. Table V demonstrates that the
TABLE VI
COMPARISON OF CLASSIFICATION AND DIAGNOSTIC PERFORMANCE BETWEEN SVM MODELS CONSTRUCTED USING FEATURES SELECTED WITH FILTER AND WRAPPER METHODS

<table>
<thead>
<tr>
<th>Feature Method</th>
<th>Entropy</th>
<th>F-score</th>
<th>SVM-RFE</th>
<th>SVM-RFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S cell (N=492)</td>
<td>96.98%</td>
<td>96.78%</td>
<td>96.37%</td>
<td>96.91%</td>
</tr>
<tr>
<td>I cell (N=339)</td>
<td>97.94%</td>
<td>97.94%</td>
<td>98.53%</td>
<td>97.71%</td>
</tr>
<tr>
<td>P cell (N=216)</td>
<td>95.37%</td>
<td>95.37%</td>
<td>96.30%</td>
<td>95.37%</td>
</tr>
<tr>
<td>D cell (N=151)</td>
<td>91.39%</td>
<td>91.39%</td>
<td>91.39%</td>
<td>91.39%</td>
</tr>
<tr>
<td>All cells (N=1198)</td>
<td>96.16%</td>
<td>96.34%</td>
<td>96.16%</td>
<td>96.16%</td>
</tr>
</tbody>
</table>

B. Experiment 2: Independent Training and Testing

In the second experiment, 1198 cell images, each consisting of 29 nuclear variables and 29 cytoplasmic variables, were randomly divided into training and testing sets, in which the training and testing sets contain 837(70%) and 361(30%) cells, respectively. In the training phase, tenfold cross validation was adopted to train the model with best performance, which was followed by the testing phase to verify the performance of classification. As indicated in Table VII, the accuracies of cell classification and dysplasia cell diagnosis are 96.12% and 98.61%, respectively, using features selected by SVM-RFE and SVM-RFA methods. The features selected by SVM-RFE are 9 for four-cluster classifier and 8 for two-cluster classifier, whereas the features selected by SVM-RFA include eight features for both four-cluster and two-cluster classifiers. Furthermore, the features, including N/C, ECM, CCM, Perimeter, AQP, and L, were selected by SVM-RFE in both two-cluster and four-cluster classifiers. Additional features, such as area, coarseness, and MGP were also recruited in the four-cluster classifier, and contrast recruited in the two-cluster classifier. With regard to the SVM-RFA selection method, the features, including N/C, area, ECM, CCM, perimeter, and AQP, were selected for both classifiers. Additional features selected include coarseness and MAP in the four-cluster classifier and contrast and AAP in the two-cluster classifier.

IV. DISCUSSION

The merit of cytology in discriminating abnormal from normal cells has been widely recognized and accepted in cervical screening programs. However, the false negative cases leading to postponement of optimal treatment are discouraging. In the false negative group, approximately two-thirds are related to sampling/preparation errors, i.e., the inflammatory, bloody or mucinous background, and cellular crowds. The other one-third are related to screening errors and interpretive errors causing abnormal cells not being correctly classified [2]. In our study, in order to reduce the sampling error, the cell samples for image analysis are taken from liquid-based prepared smears instead of conventional smears. Such liquid-based monolayer collection and preparation can reduce artifacts inherent in the conventional smears, e.g., poor fixation, thick and overlapping groups, obscuring inflammation, blood and mucus, etc. It was tried previously in the development of automated image analysis system for cervical smear [13], [15]. Liquid-based smears can offer...
the stability for reproducible high-quality cellular morphology and can be used to establish the database of cellular images. In addition to better smear preparation, Mackin et al. [67] proposed 3-D imaging and automated segmentation algorithms for analysis of images with thick and overlapping cellular clusters. Again, the liquid-based smears nullify the deployment of high-cost 3-D imaging system by greatly eliminating the occurrence of overlapping cells.

**A. Cell Segmentation**

The segmentation of cervical cells is an important step for cell image analysis and feature extraction. To accurately segment cell nucleus, cytoplasm, and background remains a challenging issue. Recently, although automatic [23], [24], [33] and semiautomatic [25] methods have been proposed for the segmentation of the cervical cell from Pap smear images, some shortcomings can still be found. For example, the methods proposed by Plissiti et al. [23] and Sobrevilla et al. [24] only focused on the detection on nuclei without taking cytoplasm detection into consideration. Furthermore, the performance of automatic detection methods is degraded by cell overlapping, saturation and hue of cell images, and artifacts caused by vaginal secretion and blood staining [24]. Although nuclei are considered as the most informative regions in reflecting structural change when affected by diseases, it is well known that N/C ratio which needs the information of cytoplasm area is deemed as an important feature for discriminating the dysplasia from normal cells.

Tracing the complex multicellular processes, such as cell migration, cell cycle, and cell differentiation, in a living cell population from fluorescence microscopic images has been conducted in previous works [68]–[71]. Chen et al. [68] proposed an automatic system for segmentation, classification, and tracking of nuclei of cancer cells in a dynamic time-lapse setting for studying cell migration, cell cycle, and cell differentiation. Lu et al. [71] developed a system for automatic classification of different cell phases of the cultured Zebrafish cells. The system proposed by Du et al. [69] was demonstrated to be effective in extracting intensity, shape, and texture features of 3-D in vivo images for the classification of different cell phases of live Drosophila embryos. Based on the cell features provided by the CellProfiler [46], Jones et al. [70] presented a supervised machine learning approach using iterative feedback for quickly and easily classifying and scoring cells with diverse cellular morphologies automatically.

Consideration of segmenting cells from fluorescence microscopic images is different from the matured fixed papanicolaou-stained visible-light microscopic images of cervical cells presented in this study. As argued by Carpenter et al. [46], the automated image analysis system developed to customize for a specific cell type, assay, or image set is rarely applicable to different cell types. Hence, the system developed for tracking cells in different phases from fluorescent microscopic images cannot be directly applied in cell analysis of Pap smear images. The semiautomatic NCC-RW method can deal with overlapping cells with better segmentation performance. To verify the segmentation performance of the proposed NCC-RW detector, it is compared with other methods.

Table II compares the contour detection performance of the NCC-RW method adopted in this study with other methods. As indicated in this table, NCC-RW is better than the other methods in the detection of cytoplasmic contours. Regarding the detection of nuclear contours, the performance of NCC-RW is similar to CellProfiler while still outperforming the other two methods. The semiautomatic system integrating the automatic NCC detector and the manual RW detector provides flexible semiautomatic methods for accurately detecting both NCCs of the cervical cells, especially the overlapping cells.

**B. Cell Classification**

In this study, the 13 morphometric variables were first selected as salient features by using ANOVA and were categorized into four groups, including 1) nuclear size (perimeter, area, maximum length, and maximum width), 2) shape (MAP, AAP, MGP and AGP), 3) nuclear texture such as density or distribution of chromatin pattern (ECM, CCM, coarseness, and contrast), and 4) N/C ratio. By taking individual variables into consideration, the statistical results (see Table III) show that the 13 selected morphometric variables are able to differentiate four different cell types ($p < 0.001$, ANOVA) and to discriminate the normal from the dysplastic cervical cells ($p < 0.001$, t-test). It also indicates that the dysplastic cells have larger size (i.e., larger perimeter, area, maximum length, and maximum width), higher nuclear proportion (i.e., larger N/C ratio), greater irregularity in nuclear shape (i.e., larger MAP, AAP, MGP, and AGP), and higher nuclear irregularity (i.e., greater contrast). These findings are all compatible with previous human observation [72] and computed morphometric data [16], i.e., the dysplastic squamous cells have larger nuclear size, more irregular nuclear shape, hyperchromatia, and higher N/C ratio.

As indicated in Tables V and VI, with regard to the differentiation of four different types of cells, the outcome shows that the total accuracy only slightly raises from 96.58% to 96.91% accompanied with an increase of classification accuracy for parabasal (from 93.06% to 95.83%) and dysplastic (from 92.05% to 92.72%) cells and a slight decrease in discriminating superficial (from 98.58% to 98.37%) and intermediate (from 97.94% to 97.35%) cells after excluding the five variables that fail to discriminate two of the four different types of cells. Furthermore, concerning the discrimination between the dysplastic and the normal cells, the diagnostic performance is slightly improved as well, with an increase of accuracy from 97.91% to 98.16%, sensitivity from 87.42% to 88.74%, specificity from 99.43% to 99.52%, PPV from 95.65% to 96.40%, and NPV from 98.21% to 98.39%.

Filter methods, such as F-score and entropy, were also applied for feature selection. The diagnostic performance (accuracy: 98.33%) was only slightly improved compared to the traditional statistical method (accuracy: 98.16%) to discriminate dysplastic from normal cells; moreover, the classification performance for features selected using entropy in discriminating four different types of cells was decreased. Hence, the filter method
TABLE VIII
COMPARISON OF FEATURES SELECTED WITH DIFFERENT METHODS FOR FOUR-CLUSTER AND TWO-CLUSTER CLASSIFIERS IN TWO DIFFERENT EXPERIMENTS (CV: CROSS VALIDATION; ITT: INDEPENDENT TRAINING AND TESTING)

<table>
<thead>
<tr>
<th>Exp</th>
<th>Method</th>
<th>Class</th>
<th>Peri</th>
<th>Area</th>
<th>L</th>
<th>W</th>
<th>N/C</th>
<th>MAP</th>
<th>AAP</th>
<th>MGP</th>
<th>AGP</th>
<th>ECM</th>
<th>CCM</th>
<th>Coar</th>
<th>Cont</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>RFE</td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RFA</td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT</td>
<td>RFE</td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RFA</td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the classifiers designed with independent training and testing sets in the second experiment, eight variables were selected with SVM-RFE, while there are nine and eight features selected with SVM-RFD for four-cluster and two-cluster classifiers, respectively. Table VIII compared the features selected with different feature selection methods in two different experiments. It can be found that four features, including perimeter, AGP, ECM, and CCM were all selected regardless of variation in feature selection methods or experiments. On the other hand, three features, i.e., area, N/C, and coarseness, were selected by the four-cluster classifier, while contrast was selected by the two-cluster classifier no matter the feature selection methods or experimental designs. Among these selected features, nuclear area, nuclear perimeter, and N/C ratio were also selected by Genctav et al. [37] in cell classification.

After running Spearman’s correlation test using statistical software package (SPSS ver. 17), most of the adopted variables demonstrate significant correlation. For example, the Tamura features (coarseness versus contrast) or Haralick features (ECM versus CCM) show strong correlations ($p < 0.0001$). Among the features adopted in this study, only the coarseness of the Tamura features and ECM of the Haralick features do not show significant dependence ($p > 0.05$), and can be used to effectively discriminate abnormal dysplastic from normal cells and to differentiate four different cell types. For designing classifiers with selected features, features that are statistically dependent with each other can be used together to reinforce the performance of classification. As demonstrated in Genctav et al. [37], nuclear features including nuclear area (same as Area in this study), nuclear perimeter (same as Perimeter in this study), nuclear longest diameter (similar to $L$ or MGP), and nucleus shortest diameter (similar to $W$), which are statistically dependent, are used together for classification. In addition, the N/C ratio and nuclear area used in both their and our studies are also statistically dependent.

Although Nunobiki et al. [19] reported the usefulness of RGB color specification in analyzing the variation of color properties for Papnicolaou-stained cervical smears, in this study, only gray-scale information was used to calculate cell textures for automated cytologic diagnosis with very good performance. In addition to appropriate morphometric variables for single cellular analysis, Zahniser et al. [15] proposed that a combination of three analytic methods, contextual analysis, single cell...
analysis, and intermediate cell analysis could achieve a better overall classification of cervical smears, which is consistent to our investigation. Thus, the reliability of microscopic findings can be validated by an automated computed measurement system.

V. CONCLUSION

In this study, an SVM-based classifier was successfully used to train and predict the normality of the cervical squamous cells, which is better than the LDA classifier implemented using SPSS program. The SVM algorithm was implemented using the LIBSVM program (http://www.csie.ntu.edu.tw/~cjlin/libsvm). By using the model, we can quickly, easily, and automatically perform analysis, classification, and regression on the datasets. After having compared several feature-selection methods, SVM-RFA was found to be capable of selecting the optimal combination of salient features to achieve best classification accuracy and diagnostic performance. The classification accuracy of the SVM-based multiple-cluster classifier constructed based on the eight selected variables in discriminating four different types of cells is impressive by achieving a total accuracy as high as 97.16%. On the other hand, the two-cluster predictive model designed using seven selected salient features enabled us to effectively distinguish the dysplastic from normal cells with specificity (99.90%), and AUC (0.995). Its classification outcomes are much better than the results predicted by the model constructed based on the features selected using statistical or filter methods.

In conclusion, an automated image analysis system has been developed for morphologic measurement and evaluation of cervical cells. By using the designed image analysis system, we can establish objective quantitative evaluation from images obtained from different laboratories. More subsequent investigation with statistical measurements is needed to elucidate its practical utility in a laboratory and its ability to improve the diagnostic performance of a laboratory. It is expected to help diminish the number of the gray-zone diagnosis, especially between the atypical squamous cells with undetermined significance (ASCUS) and LSIL, as well as between the ASCUS and HSIL. This approach could also assist cellular classification in other cytological fields. Through a combination of findings from clinical, cytomorphic, and morphometric variables, we can resolve additional problems related to interobserver variation and ambiguous diagnosis.

ACKNOWLEDGMENT

The authors would like to thank Prof. J. L. Semmlow of Rutgers University for his help in editing this paper.

REFERENCES


Yung-Fu Chen received the Diploma degree from the National Taipei Institute of Technology, Taipei, Taiwan, in 1983, and the M.S. degree from the University of Missouri-Rolla, Rolla, MO, USA, in 1990, both in electrical engineering, and the Ph.D. degree in biomedical engineering from National Cheng Kung University, Tainan, Taiwan, in 2002.

Since 1992, he has been with the Central Taiwan University of Science and Technology, Taichung, Taiwan, where he is currently a Professor in the Department of Healthcare Administration and an Adjunct Professor at the Institute of Biomedical Engineering and Material Science. He is also an Adjunct Professor in the Department of Health Services Administration, China Medical University, Taichung. His research interests include medical informatics, innovative applications of radiofrequency identification in business and medicine, and digital preservations.

Po-Chi Huang received the M.D. degree from China Medical University, Taichung, Taiwan, in 1997, and the M.S. degree in biomedical science from National Chung Hsing University, Taichung, in 2003.

From 2008 to 2011, she was with the Department of Pathology, Taichung Hospital. She is currently with the Department of Pathology, Taipei Hospital, Department of Health, Executive Yuan, Taipei, Taiwan. Her research interests include cytopathology and medical education.

Ker-Cheng Lin received the M.D. degree from National Defense Medical College, Taipei, Taiwan, in 1977, and the M.S. degree in public health from The Johns Hopkins University, Baltimore, MD, USA, in 1996.

He was with Chishan Hospital, Department of Health, from 1998 to 2000 and from 2007 to 2010, serving as the Deputy Superintendent and the Superintendent, respectively. Since 2000, he has been with Nantou Hospital, Department of Health, Executive Yuan, Nantou, Taiwan, where he is currently the Superintendent. He is a gynecologist and an expert in breast cancer surgery.

Hsuan-Hung Lin received the Ph.D. degree in applied mathematics from National Chung Hsing University, Taichung, Taiwan, in 2010.

He is currently an Associate Professor in the Department of Management Information Systems and the Director of Computer Center, Central Taiwan University of Science and Technology, Taichung. His research interests include biomedical information and bioinformatics.

Li-En Wang received the B.S. degree in computer science and information engineering from Da-Yeh University, Changhua, Taiwan, in 2010, and the M.S. degree in health services administration, China Medical University, Taichung, Taiwan, in 2012. He is currently working toward the Ph.D. degree in the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan.

His research interests include medical informatics and clinical decision support system.

Chung-Chuan Cheng received the B.S. degree in information management from the Chaoyang University of Technology, Taichung, Taiwan, in 1996.

He is currently working toward the Ph.D. degree in the School of Electrical Engineering, National Chung Hsing University, Taichung.

His research interests include content-based image retrieval, information management, and web-based business system development.

Tsung-Po Chen received the M.D. degree from National Yang Ming University, Taipei, Taiwan, in 1985.

He is currently working toward the Ph.D. degree at the Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan.

From 1989 to 1994, he was with the Department of Surgery, Taichung Veterans General Hospital. He is currently the Superintendent of Tsung-Po Chen Clinical and the Director-General of Taichung City Clinic Association, Taichung. His research interests include cytopathology and clinical decision support system in colorectal practice.

Yung-Kuan Chan received the M.S. degree in computer science from the New Mexico Institute of Mining and Technology, Socorro, NM, USA, in 1991, and the Ph.D. degree in computer science and information engineering from National Chung Cheng University, Chiayi, Taiwan, in 2000.

From 2001 to 2002, he was an Assistant Professor in the Department of Information Management, Chaoyang University of Technology. From August 2002 to July 2003, he was an Assistant Professor in the Department of Computer Science and Information Engineering, National Hweii Institute of Technology. From August 2003 to July 2005, he was an Assistant Professor, from August 2005 to July 2008, an Associate Professor, and currently a Full Professor and the Chairman of the Department of Management Information Systems, National Chung Hsing University, Taichung. His research interests include image processing, image retrieval, image compression, image hiding, and data mining.

John Y. Chiang received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from the Northwestern University, Evanston, IL, USA, in 1987 and 1990, respectively, all in electrical engineering.

Since 1990, he has been with National Sun Yat-sen University, Kaohsiung, Taiwan, where he is currently a Professor in the Department of Computer Science and Engineering. His research interests include pattern recognition, image processing, and content-based image retrieval.
Semiautomatic Segmentation and Classification of Pap Smear Cells

Yung-Fu Chen, Po-Chi Huang, Ker-Cheng Lin, Hsuan-Hung Lin, Li-En Wang, Chung-Chuan Cheng, Tsung-Po Chen, Yung-Kuan Chan, and John Y. Chiang

Abstract—Cytologic screening has been widely used for detecting the cervical cancers. In this study, a semiautomatic PC-based cellular image analysis system was developed for segmenting nuclear and cytoplasmic contours and for computing morphometric and textual features to train support vector machine (SVM) classifiers to classify four different types of cells and to discriminate dysplastic from normal cells. A software program incorporating function, including image reviewing and standardized denomination of file names, was also designed to facilitate and standardize the workflow of cell analyses. Two experiments were conducted to verify the classification performance. The cross-validation results of the first experiment showed that average accuracies of 97.16% and 98.83%, respectively, for differentiating four different types of cells and in discriminating dysplastic from normal cells have been achieved using salient features (8 for four-cluster and 7 for two-cluster classifiers) selected with SVM recursive feature addition. In the second experiment, 70% (837) of the cell images were used for training and 30% (361) for testing, achieving an accuracy of 96.12% and 98.61% for four-cluster and two-cluster classifiers, respectively. The proposed system provides a feasible and effective tool in evaluating cytoligic specimens.

Index Terms—Cervical cancer, cytology, image analysis, Pap smear, support vector machines (SVMs).

I. INTRODUCTION

Cytology evaluation is a safe, efficient, and well-established technique for the diagnoses of many diseases. The most famous success in cytology is its ability to reduce the mortality and morbidity of cervical cancer through mass screening. It was reported that the invasive cancer incidence decreased by 47.8% after national screening from 1995 to 2006 in Taiwan [1]. One role of cytology is directed to early detection of dysplasia or preinvasive cancer cells. Once the abnormal cells are detected, the patient can be scheduled for a biopsy examination and subsequent surgical treatment. Consequently, the progression of the cancer can be stopped at an early stage.

Classical cytological diagnosis is based on microscopic observation of specialized cells and qualitative assessment using descriptive criteria, which may be inconsistent because of subjective variability of different observers [2]. To lower the false negative rate in screening, many advanced technologies involving sampling, smear preparation, or screening quality control have been developed and introduced [3]–[6]. Commercial devices that use these technologies can be divided into the following categories based on their approaches: 1) improved slide preparation to reduce sampling error, e.g., thin-layered liquid-based preparation (ThinPrep, SurePath, Tripath) [7], [8]; 2) reduced workload and screening error as in the autoscreening system (ThinPrep Imaging System, Cytyc, Boxborough, MA; FocalPoint System, Tripath Imaging, Burlington, NC); 3) improved laboratory quality control like rescreening (Papnet) [9]; and 4) enhanced quality assurance, such as the proficiency test [10]. However, most of these devices do not assist objective diagnosis by providing the calculable variables that would eliminate interpretation errors and interobserver discrepancy [11]. In addition, they are not applicable to the general cytological laboratory because of high cost and technical or linguistic gaps [6], [12]. Thus, without a reproducible and quantitative tool, observer bias is still an unsolved problem in the routine cytological laboratory. Diagnostic divergence caused by visual observation remains.

The technique of computerized image analysis used to assist artificial diagnosis of cell abnormalities or tumors in cytology or histopathology also can provide accurate and objective evaluation of nuclear morphology. Quantitative methods for estimating a cytological specimen can be traced back some 30 years and are still continuing to develop [9], [13]–[19]. Reliability, accessibility, cost, efficiency, technical maintenance, and linguistic communication are considerations that need to be taken into account in any new designs. Due to revolution and evolution of new technologies, enhanced power of computation,
increased cost of hardware and software, and prevalence of the Internet, more and more systems being developed use computational algorithms for cellular image analysis [20]–[22]. Such approaches promise to resolve the limitation of subjective analysis, especially in the fields of bioinformatics, biology, and medicine.

Traditional cytological criteria for differentiating dysplastic cells from normal cells are based on the change in nucleus to cytoplasm (N/C) ratio, nuclear size, irregularity of nuclear shape and nuclear membrane, and density and granularity of nuclear chromatin. Most of these criteria are descriptive and relatively subjective. In contrast, in computed morphometry, the subjective criteria are replaced by quantitative, calculable, and comparative variables [23]–[25]. This study aimed at setting up a reproducible and reliable analytical tool to facilitate interpretation and to create a reliable database of cell images for conducting cytopathological educations and designing clinical diagnostic systems. Generally, there are three types of cervical squamous cells, i.e., superficial, intermediate, and parabasal cells, observed on the Pap smear images. The endocervical columnar epithelial cells are occasionally found, sometimes in small clusters, in the Pap smear samples. We evaluated the cell morphometric variables by using statistical analysis and other feature-selection methods, such as filter and wrapper methods, to select salient variables for the design of classifiers and construction of predictive models to automatically classify different cell types and to discriminate the dysplastic cells from normal cells, respectively. The designed multiple-cluster classifier is valuable in the classification of cell types for constructing cell image database and is useful in applications such as cytological and pathological educations, whereas the constructed predictive model is shown to be able to assist pathologists to increase the diagnostic performance of dysplasia cells.

Support vector machine (SVM) is a supervised learning method widely used for classification [26], [27]. A special property of SVM is that it can simultaneously minimize the empirical classification error and maximize the geometric margin of a classifier. It is a powerful methodology for solving problems in nonlinear classification, function estimation, and density estimation, leading to many applications including image interpretation, data mining, biometric authentication, biotechnological investigation, and other electrical applications. [28]–[32]. It has also been used for automatic detection of cell nuclei on Pap smear images [23].

Recently, automatic [23], [24], [33] and semiautomatic [25], [34] methods have been proposed to detect nuclear contours of cervical cells on the Pap smear images and to discriminate the normal from abnormal dysplasia cells. However, it was argued that cytoplasm contains important features for the discrimination of abnormal cells [35]. More recently, classification of different types of cervical cells has been pinpointed [36], [37]. In the study of Sokouti et al. [36], the cervical cells were classified into normal, low-grade squamous intraepithelial lesion (LSIL), and high-grade squamous intraepithelial lesion (HSIL), but not considering different types of normal cervical cells. Moreover, Gencv et al. [37] classified seven types of cells, including two types of cervical squamous (superficial and intermediate) cells, one type of endocervical (columnar) cell, three types of dysplasia (mild, moderate, and severe) cells, and carcinoma in situ, based on unsupervised classification. Since the columnar cells are occasionally found in the Pap smear images and the carcinoma in situ is in an early stage of cancer, early detection of dysplasia cells has been generally focused in the cytopathological setting of preventive medicine. In this study, in addition to discriminating the abnormal dysplasia from the normal cells, we also purport to classify the cervical cells into dysplasia and three different types of normal cells. The preliminary results were reported in Huang et al. [38].

In our previous work, only 503 cells (27%) selected from 1814 acquired cervical cells with peer-agreement were used for manual segmentation of the nuclear and cytoplasmic contours (NCCs), which were then applied as the masks to calculate morphologic and textural features to design a two-cluster classifier for discriminating abnormal from normal cells and a multiple-cluster classifier for classifying four different types of cervical cells based on the SVM [38]. However, it was argued that 27% images selected were not representative and might render the study highly bias. In this study, in order to eliminate such a bias, three pathologists were again invited to make further discussion regarding those images which did not reach agreement in categorization in the first round to convince each other to include more cases for analyses, resulting in a total of 1198 (66%) cell images selected for cell analysis and classifier design. Furthermore, we improved the previous work by developing a semiautomatic system integrating a NCC detector for automatically detecting NCCs and a manual random walk (RW) detector for manual segmentation to further improve the performance of contour detection for those cells which were not well detected by the NCC detector.

Moreover, two experiments were conducted to verify performance of the SVM classifiers. In the first experiment, twofold cross validation was used to obtain the SVM parameters, followed by tenfold cross validation to verify performance of the classifiers. Filter and wrapper methods were used to select salient features for building the SVM classifiers. On the other hand, in the second experiment, the dataset was randomly divided into training set and testing set, which contain 837 (70%) and 361 (30%) cells, respectively. The SVM classifiers were trained with salient features selected using wrapper methods based on the training set, and tested with the cell images in the testing set. The rest of this paper is organized as follows. Section II describes the materials and methods proposed and adopted in this study. Section III demonstrates the experimental results. Finally, brief discussion and conclusion are made in Sections IV and V, respectively.

II. MATERIALS AND METHODS

All the analysis and experiments are conducted based on the designed programs and the equipment in a routine cytologic laboratory. Here, we describe a prototypic system composed of a microscope, digital camera, personal computer, automated cellular processing and analyzing programs, and SVM classifier.

The flowchart of the overall experimental procedure is illustrated in Fig. 1. It can be divided into the following steps:
acquisition and categorization of cell images, image editing and processing, contour segmentation of cell nucleus and cytoplasm, measurement and analysis of cellular morphology and texture, classification of cell types using SVM, and assessment of diagnostic performance. Customized programs were designed in our laboratory to simplify the analysis procedures, facilitate cellular image review, organize the file names of cell images, and calculate the morphometric features of cells. Additionally, statistical significance of the selected morphometric and textural features, performance between the SVM and the Fisher linear discriminant classifiers, and efficacy among different feature-selection methods were also compared.

A. Acquisition and Categorization of Cell Images

Cytological images were captured using a high-resolution digital camera (Olympus C-5060, Japan) mounted on a microscope (Olympus BX 51, Japan) and stored as digital format. Cell images were captured with three different scales of magnification (100x, 200x, and 400x) by excluding the images with minimal magnification, extensive cellular overlapping, and interference by other inflammatory cells or debris. A total of 1814 cell images were captured from 42 Pap smears. Fig. 2 demonstrates the program designed for image interpretation and file name standardization and renaming. An image can be inspected easily in 100x, 200x, and 400x magnification with 24 bits color depth. After excluding the images with peer disagreement in classification in the first round, only 503 (27%) images were obtained, which is not representative and may render the study highly bias. In order to eliminate such a bias, at the second round, three pathologists were again invited to make further discussion regarding the images that did not reach agreement in categorization at the first round to convince each other to include more cases for analyses. Finally, 1198 (66%) images were selected for further analysis and training, in which 492 images were categorized as superficial (S) cells, 339 as intermediate (I) cells, 216 as parabasal (P) cells, and 151 as low-grade and HSIL. In addition to these cellular images, micrometer images at 400x magnification were also captured for calibrating the prototypic system.

B. Image Editing and Processing

Extraction of individual cellular images and generation of their corresponding nucleus and cytoplasm masks were achieved by using a designed program. Its friendly graphic user interface makes the image processing and editing tasks easy to operate for clinical laboratory personnel. First, the program automatically adjusted the color of the image on the computer to a visual view very similar to the view on the microscope. Second, individual cells were selected with the provided cropping tool, followed by the segmentation of NCCs. Fig. 3 illustrates the procedure in automatically generating nuclear and cytoplasmic masks of a selected cell. An extracted cell image and its masks were saved.

Classification of cell types was based on peer agreement which is treated as the gold standard for evaluating the efficacy of the classifier. The digitized cellular images were reviewed by three certificated cytopathologists and six certificated cytotechnologists. Images were displayed at a resolution of 1024 × 768 pixels with 24 bits color depth. After excluding the images with peer disagreement in classification in the first round, only 503 (27%) images were obtained, which is not representative and may render the study highly bias. In order to eliminate such a bias, at the second round, three pathologists were again invited to make further discussion regarding the images that did not reach agreement in categorization at the first round to convince each other to include more cases for analyses. Finally, 1198 (66%) images were selected for further analysis and training, in which 492 images were categorized as superficial (S) cells, 339 as intermediate (I) cells, 216 as parabasal (P) cells, and 151 as low-grade and HSIL. In addition to these cellular images, micrometer images at 400x magnification were also captured for calibrating the prototypic system.
as two different but associated files, which were then used for calculating nuclear and cytoplasmic variables by an automated analysis program. Fig. 3(a) demonstrates a cell cropped from a Pap smear image by the designed program for detecting contours using the integrated NCC detector [39] and RW detector [40]. The NCC detector, as detailed in the following section, is used to automatically detect the cell contours. If the user is not satisfied with the segmented contours, the RW detector is further applied for adjusting the contours by drawing sketchy curves inside nucleus, in cytoplasm, or outside the cell, respectively, to mark textures within different cell regions. For images containing multiple cells, the NCC detector was not able to accurately detect the NCCs. The RW detector could be used to effectively resolve these cases. As illustrated in Fig. 3(b), the segmented regions of the nucleus and cytoplasm are converted into masks represented in blue and green, respectively. The masks were then used to locate the nuclear and cytoplasmic regions on the original cell image to automatically calculate morphometric and textural variables using the analyzing program. Fig. 3(c) illustrates the mask overlaid on a cell image to facilitate the calculation of nucleus and cytoplasm variables. The graphical user interface of the automated analysis program is shown in Fig. 3(d).

C. Automatic NCC Detection

The automatic NCC detection includes two consecutive phases: cytoplasmic contour detection followed by nuclear contour detection. Although Otsu’s method is effective in the determination of the threshold value to separate an object from the background, it cannot provide a proper threshold when the standard deviations or the number of data are significantly different between two groups, i.e., object and background [41]. The adaptable threshold decision (ATD) method, adopting standard deviations of groups, numbers of data, and group intervals, was demonstrated to be able to determine the optimal threshold to effectively separate the cytoplasm from the background.

The nucleus contour detection phase consists of three steps, including gradient calculation, nuclear contour detection, and contour connection. First, Sobel operator was used to calculate gradients of the cell image. It was found that Sobel operator is very sensitive to noise; hence, the mean vector difference enhancer was applied to highlight the gradient of the object contour and to suppress the gradient of the noise contour [42]. Second, since the nucleus of a cervical cell on the Pap smear image is generally much smaller than its cytoplasm and the background of the cropped image, especially for a normal cell, the ATD method cannot provide an optimal threshold to precisely separate the nucleus from the cytoplasm. Therefore, the method based on maximum gray-level gradient difference (MGLGD) was adopted for detecting the nuclear contour [39]. The main idea behind the MGLGD is that the difference of mean gray levels between the region inside and the region outside the contour is maximal at the contour of an object. In addition, the gradients of most pixels located on the contour are much larger than the gradients of other pixels. Finally, the cytoplasmic contour obtained in the cytoplasm detection phase was used to determine the initial contour by moving ten pixels toward the nucleus for detecting the contour of cell nucleus. In each iteration progressing to the final nuclear contour, the new contour may not form a closed boundary. In this case, an additional pixel was added to close the contour to step forward the next iteration. Fig. 4 illustrates the results obtained from individual steps.

Although the NCC detector is promising in detecting NCCs of most single-cell images, it cannot effectively segment overlapping cells or cells deteriorated by artifacts. The RW detector can be used to cope with the aforementioned problems. As demonstrated in Fig. 5(a), NCC detector detects the cytoplasmic contours covering two cytoplasmic contours. After being adjusted by an RW detector, the NCCs are effectively detected in the cell image.
Fig. 5. (a) Contours automatically detected by NCC and (b) contours adjusted by RW detector. (c) Detected nuclear and cytoplasmic morphometric variables. (d) Illustration of morphometric variables, including perimeter, axle center, CG, maximum axle length, and maximum axle width, defined for quantitative analysis of the cell shape.

[see Fig. 5(b)]. Fig. 5(c) illustrates the detected nuclear and cytoplasmic morphometric variables defined in Fig. 5(d).

To compare the performance of contour detection, misclassification error (ME) [43], relative foreground error [43], and modified Hausdorff distance (MHD) [44] were used as the measurement indexes. The proposed NCC-RW method was compared with the edge enhancement nuclear and cytoplasmic contours (EENCC) detector [42], gradient vector flow (GVF) ATD method [45], and CellProfiler [46].

D. Measurement of Cellular Morphology and Texture

A software program, namely personal computer-based cellular image analysis (PCCIA) system, was designed to perform the cellular (nucleus and cytoplasm) image analysis. PCCIA can automatically measure the morphometric variables of the cells either individually or in a batch. Fig. 5(d) illustrates the terms, including axle center, center of gravity (CG), perimeter, maximum length, and maximum width, defined for calculating morphometric variables [47]–[50]. With the help of internal calibration using the micrometer image, various variables, including nuclear perimeter, area, maximum axle length \( L \), maximum axle width \( W \), N/C ratio, maximum length from axle center to perimeter (MAP), average length from axle center to perimeter (AAP), maximum length from center of gravity to perimeter (MGP), and average length from center of gravity to perimeter (AGP), were obtained for the evaluation of the nuclear size and shape irregularity of a cell.

As shown in Fig. 5(d), the maximum axle length \( L \) is the length of the longest line among the lines connecting any two points on the nucleus or cytoplasm perimeter. After \( L \) has been determined, the maximum axle width \( W \) can be obtained by searching the line, with one point located at the left side and the other one at the right side of the maximum axle, perpendicular to the maximum axle and has the longest width. The axle center (AC) is the intersection between the maximum axle and the minimum axle. The CG can be obtained by accounting the coordinate of individual pixels \((p_x, p_y)\) on the perimeter, i.e.,

\[
CG = \left( \sum_{i} \frac{p_x^i}{N}, \sum_{i} \frac{p_y^i}{N} \right)
\]  

where \( N \) indicates the number of points located on the perimeter. The MAP, AAP, MGP, and AGP can be calculated from the following equation:

\[
MAP = \max_i \left( \sqrt{(AC_x - p_x^i)^2 + (AC_y - p_y^i)^2} \right) \quad (2a)
\]

\[
AAP = \sum_{i=1}^{N} \sqrt{(AC_x - p_x^i)^2 + (AC_y - p_y^i)^2} / N \quad (2b)
\]

\[
MGP = \max_i \left( \sqrt{(CG_x - p_x^i)^2 + (CG_y - p_y^i)^2} \right) \quad (2c)
\]

\[
AAP = \sum_{i=1}^{N} \sqrt{(CG_x - p_x^i)^2 + (CG_y - p_y^i)^2} / N. \quad (2d)
\]

Initially, five Haralick features [47], i.e., entropy, energy, contrast, homogeneity, and linearity, and three Tamura features [48], i.e., coarseness, contrast, and roughness, were considered. In addition, roundness, compactness, and elongation proposed by Yonekawa et al. [49] were also adopted. After statistical analysis, two Haralick features, i.e., entropy of the co-occurrence matrix (ECM) and contrast of the co-occurrence matrix (CCM), and two Tamura features, i.e., coarseness and contrast, were also applied to analyze textures of the cell nucleus. A co-occurrence matrix is a 2-D \( n \times n \) matrix obtained from an image \( I \). The element in location \((i, j)\) of the co-occurrence matrix indicates the number of occurrences for a pixel \( q \) with color value \( i \) in image \( I \) having a pixel \( q' \) with color value \( j \) in its neighborhood situated at a distance of \( d \) pixels away. The entropy and contrast of an image can be calculated from the co-occurrence matrix according to the following equations:

\[
\text{Entropy} = - \sum_i \sum_j P(i,j) \log P(i,j) \quad (3)
\]

\[
\text{Contrast} = \sum_i \sum_j (i-j)^2 P(i,j) \quad (4)
\]

where \( P(i,j) \) indicates the probability of a pixel with gray-level \( i \) having a pixel with gray-level \( j \) adjacent to it. In this study, four different distances \((d = 1, 2, 3, \text{and } 4)\) and four different directions \((\theta = 0^\circ, 45^\circ, 90^\circ, \text{and } 135^\circ)\) were tested for obtaining the combination of \( d \) and \( \theta \) with best classification performance. It was observed that the performance changed in a range of 0.03–0.12%. The parameters with \( d = 1 \) and \( \theta = 45 \) achieving the best performance were adopted.
Tamura features have been widely adopted for characterizing low-level statistical properties of texture [48]. In this study, coarseness and contrast were selected for effectively describing the texture within the cell nucleus. To evaluate coarseness, we first take the average for each pixel \((x, y)\) over its neighborhoods with a window size of \(2k \times 2k\) pixels

\[
A_k(x, y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}} \sum_{j=y-2^{k-1}}^{y+2^{k-1}} f(i,j) / 2^{2k}
\]

(5)

where \(k = 1, 2, \text{ and } 3\). Then, the differences between two pairs of nonoverlapped blocks with the size of \(2k \times 2k\) pixels at both sides of the pixel \((x, y)\) are taken horizontally and vertically

\[
E_{k,h} = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)|
\]

\[
E_{k,v} = |A_k(x, y + 2^{k-1}) - A_k(x, y - 2^{k-1})|.
\]

(6)

For each pixel, the optimal value of \(k\) that maximizes \(E\) is selected for the measurement of coarseness by taking the average of \(A_{\text{opt}}(x, y)\), as shown in the following equation:

\[
f_{\text{crs}} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} A_{\text{opt}}(i, j)
\]

(7)

where \(m \times n\) indicates the image size. In addition, the contrast can be obtained from the following equation:

\[
f_{\text{con}} = \frac{\sigma}{(\mu_4/\sigma^4)^{1/4}}
\]

(8)

in which \(\mu_4\) and \(\sigma\) indicate the fourth moment and the standard deviation of the image.

After having analyzed a cell image based on its accompanied masks, the evaluated variables are saved in a file with format compatible to Microsoft Excel and used for further processing. Thus, these organized documents can be used for manual interpretation and for computerized automatic analysis. Table I shows image examples of four different cell types and their corresponding morphometric and textural variables computed by PCCIA.

### E. Designing Classifiers Using SVM

It is widely believed that SVM is superior to traditional statistical and neural network classifiers. The goal of SVM is to separate multiple clusters with a set of unique hyperplanes that have the greatest margins to the edge of each cluster. On the other hand, hyperplanes that separate two clusters is not unique for other linear classifiers. For a two-cluster classification example, the hyperplane separating two classes that leaves the maximum margins from both clusters [51], [52]. Regarding a nonlinear classifier, various kernels including polynomial, radial basics function, and hyperbolic tangent can be used for mapping the original sample space into a new Euclidian space, and then the linear classifier can be designed for classification.

In the first experiment, statistical analyses of morphometric and textural variables were first conducted by using analysis of variance (ANOVA) and \(t\)-test to test significance \((p < 0.05)\) of individual variables. Variables that reach significance were selected as candidate features followed by other feature selection methods, as detailed in the next section, for selecting compatible variables to design a two-cluster classifier to detect dysplastic from normal cells and a four-cluster classifier to discriminate among four different types of cells. Cells were randomly divided into two groups (two folds) for cross validation to train the optimal SVM parameters. These parameters were fixed at the validation stage, in which tenfold cross validation was applied to validate the performance of the classifier and the predictive model.

In the second experiment, the dataset containing 1198 cell images, each consisting of 29 nuclear variables and 29 cytoplasmic variables, was randomly divided into training and testing sets, in which the training and testing sets contain 837(70%) and 361(30%) cell images, respectively. In the training phase, tenfold cross validation was adopted to train the model with best performance using the training set, which was followed by the testing phase to verify the performance of the classifiers using the testing set.

The SVM software packages (LIBSVM) developed by Chang and Lin [53] was adopted for classification of different cell types. The selected features of the datasets have been normalized to meet the requirements of LIBSVM. The experimental procedure is summarized in the following.

1) Normalize the variables into the values between 0 and 1 according to the following equation:

\[
X'(n) = \frac{X(n) - \text{Min}}{\text{Max} - \text{Min}}
\]

(9)

where Max and Min indicates the maximal and minimum values, respectively, in the dataset; \(X(n)\) is the original value and \(X'(n)\), \(0 < X'(n) < 1\), is the value after normalization.

2) Label the normal and dysplastic cells with positive (superficial cell: \(y_i = +1\); intermediate cell: \(y_i = +3\); parabasal

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>S cell</th>
<th>I cell</th>
<th>P cell</th>
<th>Dysplasia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell Image</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perimeter (µm)</td>
<td>14.03</td>
<td>38.15</td>
<td>21.13</td>
<td>44.92</td>
</tr>
<tr>
<td>Area (µm²)</td>
<td>20.93</td>
<td>66.55</td>
<td>53.15</td>
<td>228.38</td>
</tr>
<tr>
<td>L (µm)</td>
<td>5.00</td>
<td>13.60</td>
<td>8.04</td>
<td>17.33</td>
</tr>
<tr>
<td>W (µm)</td>
<td>4.60</td>
<td>13.03</td>
<td>6.90</td>
<td>13.65</td>
</tr>
<tr>
<td>N/C Ratio (%)</td>
<td>1.0</td>
<td>8.1</td>
<td>23.2</td>
<td>46.4</td>
</tr>
<tr>
<td>MAP (µm)</td>
<td>34.21</td>
<td>94.37</td>
<td>50.33</td>
<td>117.09</td>
</tr>
<tr>
<td>AAP (µm)</td>
<td>28.39</td>
<td>81.06</td>
<td>45.44</td>
<td>94.92</td>
</tr>
<tr>
<td>MGP (µm)</td>
<td>32.30</td>
<td>85.80</td>
<td>50.22</td>
<td>108.16</td>
</tr>
<tr>
<td>AGP (µm)</td>
<td>28.35</td>
<td>80.65</td>
<td>45.41</td>
<td>94.65</td>
</tr>
<tr>
<td>ECM</td>
<td>6.29</td>
<td>7.39</td>
<td>6.85</td>
<td>7.79</td>
</tr>
<tr>
<td>CCM</td>
<td>38.50</td>
<td>71.34</td>
<td>44.17</td>
<td>50.35</td>
</tr>
<tr>
<td>Coarseness</td>
<td>288.22</td>
<td>113.14</td>
<td>88.07</td>
<td>106.66</td>
</tr>
<tr>
<td>Contrast</td>
<td>1.28</td>
<td>7.51</td>
<td>12.73</td>
<td>60.43</td>
</tr>
</tbody>
</table>
cell: \( y_i = +5 \) and negative integers (dysplastic cells: \( y_i = -4 \)), respectively. All the variables corresponding to a cell were then saved as a sequence of data with the following format:

\[
[label] \{index1\}:[value1] \{index2\}:[value2] \ldots \ldots
\]

in which label (or class) is an integer indicating the category of the cells, whereas index is an integer representing the morphometric variable. For example, the sequence \( +1 1\):0.708 \( 2\):0.987 \( 3\) indicates a superficial cell with the values of the first and second normalized variables of 0.708 and 0.987, respectively.

3) **Experiment 1:** In the training phased, cells in the dataset were randomly divided into two subsets, in which images of four different cell types were equally distributed, for cross validation to get the optimal SVM parameters, i.e., \( C \) (cost parameter) and \( \gamma \) (kernel parameter). During the testing phase, cell data were randomly divided into ten subsets (folds) for cross validation by fixing the SVM parameters to the values obtained at the training phase. In each iteration, nine folds were used for training the model, while the rest one was used for validation. The procedure was repeated for ten times, each using an individual fold for validation.

4) **Experiment 2:** The dataset was randomly divided into two independent sets, i.e., training and testing sets. In the training phase, cells in the training set were divided into ten subsets (folds) for cross validation to obtain the model with best performance. In the testing phase, cells in the testing set were used for verifying the performance of the SVM classifiers built in the training phase.

### F. Feature Selection

Feature selection takes the advantages of reducing the number of features and the size of storage requirements, decreasing training and computational time, facilitating data visualization and understanding, and improving predictive performance [54], [55]. The algorithms of feature selection can often be classified into three approaches including filter, wrapper, and embedded methods [54]. The filter method is a preprocessing procedure that selects a subset of features based on statistical measures independent of the designed classifiers. In contrast, the wrapper method assesses individual subsets of features in a recursive way by considering their predictive efficacy to a given classifier. It is more computational intensive than the filter method, but is believed to be able to provide more efficient outcome. The subset with a smallest number of features achieving the highest predictive accuracy is used for classifier construction. Recently, a genetic algorithm (GA), an alternative wrapper method, was also proposed as an useful method for feature selection [56], [57]. Sometimes, this strategy is also used for the adjustments of cost value and kernel parameter of SVM together with the selection of features [58] when designing a classifier. For example, it was applied to construct predictive models for the diagnoses of breast cancers [59], [60] and hypertension [61]. On the other hand, embedded method selects features during the process of model construction by considering the cost function of a model [62].

1) **Filter Method Based on Statistical Analysis:** The filter method eliminates unfavorable features but reserves salient features by taking the advantages of low computation complexity. Measures such as information entropy, distance, dependence, and consistency are generally used to select salient features before classification [61]. This method only considers the relevance between an individual independent variable (feature) and the dependent variable by ranking the independent variables according to their relevance to the dependent variable. The interdependence among independent variables is not considered at all. The wrapper method, on the other hand, selects various subsets of features for directly feeding into the classifier to obtain the best combination of features with greatest accuracy. Although it is promising in obtaining a classifier with best performance, it is an NP-hard problem needing tremendous computation time [63], especially for datasets with a great number of features.

In this investigation, filter methods, including entropy [60] and F-score [64], were adopted for feature selection. Entropy is defined as

\[
E^k = \sum \sum s_{ij}^k \log s_{ij}^k + (1 - s_{ij}^k) \log(1 - s_{ij}^k)
\]

where \( s_{ij}^k \) is the similarity measure between any two samples \( i \) and \( j \) of feature \( k \) and is defined as \( s_{ij}^k = e^{-\alpha d_{ij}} \), with \( d_{ij} \) indicating the Euclidean distance of feature \( k \) between samples \( x_i \) and \( x_j \). As manifested in the definition, the greater the Euclidean distance, the smaller is the similarity between two samples. On the other hand, F-score is defined as

\[
F^k = \frac{(\mu_1^k - \mu_2^k)^2 + (\mu_2^k - \mu_3^k)^2}{(\sigma_1^2)^2 + (\sigma_2^2)^2}
\]

in which \( \mu_1^k, \mu_2^k \), and \( \mu_3^k \) indicate means of class 1, class 2, and combination of class 1 and class 2 samples, respectively; \( \sigma_1^2 \) and \( \sigma_2^2 \) represent the standard deviations of samples in class 1 and class 2, respectively.

2) **Wrapper Method Based on Recursive Feature Elimination (RFE), Recursive Feature Addition (RFA), and GA:** The wrapper method assesses individual subsets of features in a recursive way by considering their predictive efficiency to a given classifier. For a vector space with \( n \) features, the RFE algorithm removes unimportant features based on backward sequential selection by iteratively deleting one feature at a time, resulting in a suboptimal combination of \( r \) (\( r < n \)) features with best predictive performance [54]. For SVM-RFE, it starts with all features by deleting a feature repeatedly until \( r \) features are left, which leads to a largest margin separating two classes. Weight magnitude, which is inversely proportional to the margin, is generally used as the ranking criterion in determining importance of individual features. The eliminated feature \( p \) is the one that minimizes the variation of weight

\[
\|w_p\|^2 = \sum_{i,j=0}^{N} \lambda_i \lambda_j y_i y_j K(x_i^T x_j).
\]
ratio [66], were also proposed for feature ranking [62]. In this study, mean cross validation accuracy was used as a measure of feature ranking for determining the eliminated feature in each iteration.

In contrast to RFE, RFA is based on forward sequential selection by iteratively adding one feature at a time. It starts with the most important feature followed by repeatedly adding an additional feature at a time until \( r \) features have been selected leading to a largest margin separating two classes.

The GA can find optimal solution within an acceptable time, and is faster than dynamic programming using an exhaustive searching strategy. By taking the advantage of the GA in quickly searching the optimal features, a nonlinear hyperplane with greatest margin can be obtained by using SVM to classify two clusters. Classification of multiple clusters can be easily expanded. The GA was modified to combine with LIBSVM for feature selection and classification [57].

G. Assessment

The cytopathological experts’ classification of cells was assumed to be the gold standard in this investigation. For the four-cluster classifier, confusion matrix accompanied with calculated accuracy and error rates were used to assess its classification performance. To assess the diagnostic performance of the two-cluster model constructed by SVM, the accuracy, sensitivity, and specificity of the model were compared with the model based on the linear discriminant analysis (LDA) built using the SPSS software package. Performance indices are calculated according to the following formula:

\[
\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (13a) \\
\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (13b) \\
\text{Specificity} = \frac{TN}{(TN + FP)} \quad (13c) \\
\text{PPV} = \frac{TP}{(TP + FP)} \quad (13d) \\
\text{NPV} = \frac{TN}{(TN + FN)} \quad (13e)
\]

where \( TP, TN, FP, \) and \( FN \) indicate the numbers of true positive, true negative, false positive, and false negative, respectively; \( PPV \) and \( NPV \) represent positive predictive value and negative predictive value, respectively. In addition, area under ROC curve (AUC) and confusion matrix were also used for sensitivity and error analyses. The significance of individual variables in differentiating the normal (S, I, and P) from the abnormal (HSIL and LSIL) cells was assessed using Student’s \( t \)-test with significance of \( p < 0.05 \).

III. EXPERIMENTAL RESULTS

Fig. 6 demonstrates examples of NCCs detected with the adopted NCC-RW methods and other methods proposed in previous investigations. Table II compares the performance of contour detection among different methods. The detected NCCs were used to generate masks for calculating morphological and textural features to design classifiers.

```markdown
<table>
<thead>
<tr>
<th>SEGMENTATION METHOD</th>
<th>ME</th>
<th>RAE</th>
<th>MHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVF-ATD</td>
<td>0.035</td>
<td>0.052</td>
<td>0.037</td>
</tr>
<tr>
<td>EENCC</td>
<td>0.085</td>
<td>0.172</td>
<td>0.144</td>
</tr>
<tr>
<td>CellProfiler</td>
<td>0.031</td>
<td>0.047</td>
<td>0.069</td>
</tr>
<tr>
<td>NCC-RW</td>
<td>0.023</td>
<td>0.042</td>
<td>0.025</td>
</tr>
</tbody>
</table>
```

A. Experiment 1: Cross Validation

After statistical analysis, only the nuclear morphometric variables and N/C ratio that reach significant difference \( (p < 0.05) \) were chosen for further study, while all the other cytoplasmic variables, except N/C ratio, were excluded \( (p > 0.05) \). For descriptive analysis, means and standard deviations of the 12 nuclear morphometric variables and N/C ratio are summarized in Table III. With regard to dysplastic cells, the morphometric variables including perimeter \((46.4 \pm 14.4 \, \mu m)\), area \((187.3 \pm 96.1 \, \mu m^2)\), maximum axle length \((15.7 \pm 4.2 \, \mu m)\), maximum axle width \((12.3 \pm 3.3 \, \mu m)\), N/C ratio \((35.3\% \pm 50.6\%)\), MAP \((110.0 \pm 31.5 \, \mu m)\), AAP \((84.2 \pm 21.5 \, \mu m)\), MGP \((99.8 \pm 26.9 \, \mu m)\), AGP \((83.4 \pm 21.3 \, \mu m)\), ECM \((7.4 \pm 0.4)\), CCM \((48.6 \pm 19.3)\), coarseness \((81.5 \pm 31.5)\), and contrast \((41.6 \pm 48.3)\) were all found to be significantly greater than those of
the normal cells ($p < 0.001$). All the variables demonstrated the capability to differentiate among four different cell types ($p < 0.001$, ANOVA) and to discriminate dysplasia from pooled normal cells ($p < 0.001$, t-test). The statistical results show that the dysplastic cells have larger size (i.e., larger perimeter, area, maximum length and maximum width), higher nuclear proportion (i.e., larger N/C ratio), greater irregularity in nuclear shape (i.e., larger MAP, AAP, MGP, and AGP), and greater contrast.

The classification and diagnostic performance of the SVM classifiers and a comparison with the classifiers constructed with LDA are summarized in Table IV. The classification of morphometric variables using SVM reveals greater classification performance (accuracy: 96.58%, total error: 41 cases) in discriminating four different cell types compared to the methodology established using LDA (accuracy: 90.48%, total error: 114 cases). Regarding the diagnostic performance in detecting the dysplastic cells, SVM-based model presents higher sensitivity (87.42%) with less classification errors while still attaining satisfactory accuracy (97.91%) when compared to the LDA model (sensitivity: 68.21%, accuracy: 95.83%). In addition, the area under ROC curve (AUC) for the model constructed using SVM ($AUC = 0.990$) is greater than the model constructed with LDA ($AUC = 0.984$).

As shown in Table III, although nuclear textual variable CCM can be used to discriminate dysplastic cell from pooled normal cells, there is no significant difference ($p > 0.05$, t-test) between the dysplastic cell and the parabasal cell. Furthermore, by concernsing the discrimination among three types of normal (superficial, intermediate, and parabasal) cells, it was found that maximum length, maximum width, MAP, and MGP are not able to discriminate between intermediate and parabasal cells ($p > 0.05$, t-test). Hence, these five variables were excluded to redesign the four-cluster and two-cluster classifiers by using only eight variables. Table V demonstrates that the
classification performance and diagnostic performance have increased slightly from 96.58% to 96.91% and from 97.91% to 98.16%, respectively.

As shown in Table VI, after feature selection using the wrapper methods, i.e., SVM-RFE and SVM-RFA, the performance has been improved by achieving an accuracy of 97.16% and 98.83% for four-cluster classifier and two-cluster predictive model, respectively. On the other hand, the predictive performance for classifiers designed with features selected using filter methods, i.e., entropy (four-cluster: 96.16%, two-cluster: 98.33%) and F-score (two-cluster: 97.58%), is not enhanced.

### B. Experiment 2: Independent Training and Testing

In the second experiment, 1198 cell images, each consisting of 29 nuclear variables and 29 cytoplasmic variables, were randomly divided into training and testing sets, in which the training and testing sets contain 837(70%) and 361(30%) cells, respectively. In the training phase, tenfold cross validation was adopted to train the model with best performance, which was followed by the testing phase to verify the performance of classification. As indicated in Table VII, the accuracies of cell classification and dysplasia cell diagnosis are 96.12% and 98.61%, respectively, using features selected by SVM-RFE and SVM-RFA methods. The features selected by SVM-RFE are 9 for four-cluster classifier and 8 for two-cluster classifier, whereas the features selected by SVM-RFA include eight features for both four-cluster and two-cluster classifiers. Furthermore, the features, including \( N/C \), ECM, CCM, Perimeter, AGP, and \( L \), were selected by SVM-RFE in both two-cluster and four-cluster classifiers. Additional features, such as area, coarseness, and MGP were also recruited in the four-cluster classifier, and contrast recruited in the two-cluster classifier. With regard to the SVM-RFA selection method, the features, including \( N/C \), area, ECM, CCM, perimeter, and AGP, were selected for both classifiers. Additional features selected include coarseness and MAP in the four-cluster classifier and contrast and AAP in the two-cluster classifier.

### IV. DISCUSSION

The merit of cytology in discriminating abnormal from normal cells has been widely recognized and accepted in cervical screening programs. However, the false negative cases leading to postponement of optimal treatment are discouraging. In the false negative group, approximately two-thirds are related to sampling/preparation errors, i.e., the inflammatory, bloody or mucinous background, and cellular crowds. The other one-third are related to screening errors and interpretive errors causing abnormal cells not being correctly classified [2]. In our study, in order to reduce the sampling error, the cell samples for image analysis are taken from liquid-based prepared smears instead of conventional smears. Such liquid-based monolayer collection and preparation can reduce artifacts inherent in the conventional smears, e.g., poor fixation, thick and overlapping groups, obscuring inflammation, blood and mucus, etc. It was tried previously in the development of automated image analysis system for cervical smear [13], [15]. Liquid-based smears can offer...
the stability for reproducible high-quality cellular morphology and can be used to establish the database of cellular images. In addition to better smear preparation, Mackin et al. [67] proposed 3-D imaging and automated segmentation algorithms for analysis of images with thick and overlapping cellular clusters. Again, the liquid-based smears nullify the deployment of high-cost 3-D imaging system by greatly eliminating the occurrence of overlapping cells.

A. Cell Segmentation

The segmentation of cervical cells is an important step for cell image analysis and feature extraction. To accurately segment cell nucleus, cytoplasm, and background remains a challenging issue. Recently, although automatic [23], [24], [33] and semiautomatic [25] methods have been proposed for the segmentation of the cervical cell from Pap smear images, some shortcomings can still be found. For example, the methods proposed by Plissiti et al. [23] and Sobrevilla et al. [24] only focused on the detection on nuclei without taking cytoplasm detection into consideration. Furthermore, the performance of automatic detection methods is degraded by cell overlapping, saturation and hue of cell images, and artifacts caused by vaginal secretion and blood staining [24]. Although nuclei are considered as the most informative regions in reflecting structural change when affected by diseases, it is well known that N/C ratio which needs the information of cytoplasm area is deemed as an important feature for discriminating the dysplasia from normal cells.

Tracing the complex multicellular processes, such as cell migration, cell cycle, and cell differentiation, in a living cell population from fluorescence microscopic images has been conducted in previous works [68]–[71]. Chen et al. [68] proposed an automatic system for segmentation, classification, and tracking of nuclei of cancer cells in a dynamic time-lapse setting for studying cell migration, cell cycle, and cell differentiation. Lu et al. [71] developed a system for automatic classification of different cell phases of the cultured Zebrasfish cells. The system proposed by Du et al. [69] was demonstrated to be effective in extracting intensity, shape, and texture features of 3-D in vivo images for the classification of different cell phases of live Drosophila embryos. Based on the cell features provided by the CellProfiler [46], Jones et al. [70] presented a supervised machine learning approach using iterative feedback for quickly and easily classifying and scoring cells with diverse cellular morphologies automatically.

Consideration of segmenting cells from fluorescence microscopic images is different from the matured fixed papanicolaou-stained visible-light microscopic images of cervical cells presented in this study. As argued by Carpenter et al. [46], the automated image analysis system developed to customize for a specific cell type, assay, or image set is rarely applicable to different cell types. Hence, the system developed for tracking cells in different phases from fluorescent microscopic images cannot be directly applied in cell analysis of Pap smear images. The semiautomatic NCC-RW method can deal with overlapping cells with better segmentation performance. To verify the segmentation performance of the proposed NCC-RW detector, it is compared with other methods.

Table II compares the contour detection performance of the NCC-RW method adopted in this study with other methods. As indicated in this table, NCC-RW is better than the other methods in the detection of cytoplasmic contours. Regarding the detection of nuclear contours, the performance of NCC-RW is similar to CellProfiler while still outperforming the other two methods. The semiautomatic system integrating the automatic NCC detector and the manual RW detector provides flexible semiautomatic methods for accurately detecting both NCCs of the cervical cells, especially the overlapping cells.

B. Cell Classification

In this study, the 13 morphometric variables were first selected as salient features by using ANOVA and were categorized into four groups, including 1) nuclear size (perimeter, area, maximum length, and maximum width), 2) shape (MAP, AAP, MGP and AGP), 3) nuclear texture such as density or distribution of chromatin pattern (ECM, CCM, coarseness, and contrast), and 4) N/C ratio. By taking individual variables into consideration, the statistical results (see Table III) show that the 13 selected morphometric variables are able to differentiate four different cell types (p < 0.001, ANOVA) and to discriminate the normal from the dysplastic cervical cells (p < 0.001, t-test). It also indicates that the dysplastic cells have larger size (i.e., larger perimeter, area, maximum length, and maximum width), higher nuclear proportion (i.e., larger N/C ratio), greater irregularity in nuclear shape (i.e., larger MAP, AAP, MGP, and AGP), and higher nuclear irregularity (i.e., greater contrast). These findings are all compatible with previous human observation [72] and computed morphometric data [16], i.e., the dysplastic squamous cells have larger nuclear size, more irregular nuclear shape, hyperchromatia, and higher N/C ratio.

As indicated in Tables V and VI, with regard to the differentiation of four different types of cells, the outcome shows that the total accuracy only slightly raises from 96.58% to 96.91% accompanied with an increase of classification accuracy for parabasal (from 93.06% to 95.83%) and dysplastic (from 92.05% to 92.72%) cells and a slight decrease in discriminating superficial (from 98.58% to 98.37%) and intermediate (from 97.94% to 97.35%) cells after excluding the five variables that fail to discriminate two of the four different types of cells. Furthermore, concerning the discrimination between the dysplastic and the normal cells, the diagnostic performance is slightly improved as well, with an increase of accuracy from 97.91% to 98.16%, sensitivity from 87.42% to 88.74%, specificity from 99.43% to 99.52%, PPV from 95.65% to 96.40%, and NPV from 98.21% to 98.39%.

Filter methods, such as F-score and entropy, were also applied for feature selection. The diagnostic performance (accuracy: 98.33%) was only slightly improved compared to the traditional statistical method (accuracy: 98.16%) to discriminate dysplastic from normal cells; moreover, the classification performance for features selected using entropy in discriminating four different types of cells was decreased. Hence, the filter method
TABLE VIII

<table>
<thead>
<tr>
<th>Exp</th>
<th>Method</th>
<th>Class</th>
<th>Peri</th>
<th>Area</th>
<th>L</th>
<th>W</th>
<th>N/C</th>
<th>MAP</th>
<th>AAP</th>
<th>MGP</th>
<th>AGP</th>
<th>ECM</th>
<th>CCM</th>
<th>Coar</th>
<th>Cont</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV</td>
<td>RFE</td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RFA</td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT</td>
<td>RFE</td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RFA</td>
<td>4-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2-C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 7. Comparison of accuracy for different numbers of features selected using filter (F-score and entropy) and wrapper (SVM-RFE and SVM-RFA) methods.

is not good at selecting suitable features. On the other hand, the diagnostic performance and classification performance for two-cluster and four-cluster models constructed using features selected with wrapper methods, i.e., SVM-RFE and SVM-RFA, were shown to be better than the aforementioned methods. In addition, the performance of both classifiers constructed using features selected by the SVM-RFA outperforms the SVM-RFE. Fig. 7 compares the accuracy of the two-cluster predictive model against the number of selected variables using methods including entropy, F-score, SVM-RFE, and SVM-RFA. As shown in this figure, F-score is unable to select optimal combination of features, while the other three methods are able to obtain combinations with fewer features.

Many previous studies [16], [22] indicated that successful selection of morphometric variables will affect the accuracy of cancer diagnosis. The most widely used variables for cellular classification include mean nuclear area, perimeter, shape, and N/C ratio [20], [21], [73]. In addition to these basic morphometric variables, Murata et al. [22] reported 27 additional morphological nuclear variables (convex area, convex perimeter, coefficient of variation, variance, entropy, etc.). In this study, regarding the classifiers built with cross validation in the first experiment, only nine morphometric variables were used to construct the two-cluster predictive model and eight variables applied for designing the four-cluster classifier to achieve best performance.

For the classifiers designed with independent training and testing sets in the second experiment, eight variables were selected with SVM-RFA, while there are nine and eight features selected with SVM-RFD for four-cluster and two-cluster classifiers, respectively. Table VIII compared the features selected with different feature selection methods in two different experiments. It can be found that four features, including perimeter, AGP, ECM, and CCM were all selected regardless of variation in feature selection methods or experiments. On the other hand, three features, i.e., area, N/C, and coarseness, were selected by the four-cluster classifier, while contrast was selected by the two-cluster classifier no matter the feature selection methods or experimental designs. Among these selected features, nuclear area, nuclear perimeter, and N/C ratio were also selected by Genctav et al. [37] in cell classification.

After running Spearman’s correlation test using statistical software package (SPSS ver. 17), most of the adopted variables demonstrate significant correlation. For example, the Tamura features (coarseness versus contrast) or Haralick features (ECM versus CCM) show strong correlations ($p < 0.0001$). Among the features adopted in this study, only the coarseness of the Tamura features and ECM of the Haralick features do not show significant dependence ($p > 0.05$), and can be used to effectively discriminate abnormal dysplastic from normal cells and to differentiate four different cell types. For designing classifiers with selected features, features that are statistically dependent with each other can be used together to reinforce the performance of classification. As demonstrated in Genctav et al. [37], nuclear features including nuclear area (same as Area in this study), nuclear perimeter (same as Perimeter in this study), nuclear longest diameter (similar to L or MGP), and nucleus shortest diameter (similar to W), which are statistically dependent, are used together for classification. In addition, the N/C ratio and nuclear area used in both their and our studies are also statistically dependent.

Although Nunobiki et al. [19] reported the usefulness of RGB color specification in analyzing the variation of color properties for Papnicolaou-stained cervical smears, in this study, only gray-scale information was used to calculate cell textures for automated cytologic diagnosis with very good performance. In addition to appropriate morphometric variables for single cellular analysis, Zahniser et al. [15] proposed that a combination of three analytic methods, contextual analysis, single cell
analysis, and intermediate cell analysis could achieve a better overall classification of cervical smears, which is consistent to our investigation. Thus, the reliability of microscopic findings can be validated by an automated computed measurement system.

V. CONCLUSION

In this study, an SVM-based classifier was successfully used to train and predict the normality of the cervical squamous cells, which is better than the LDA classifier implemented using SPSS program. The SVM algorithm was implemented using the LIBSVM program (http://www.csie.ntu.edu.tw/~cjlin/libsvm). By using the model, we can quickly, easily, and automatically perform analysis, classification, and regression on the datasets. After having compared several feature-selection methods, SVM-RFA was found to be capable of selecting the optimal combination of salient features to achieve best classification accuracy and diagnostic performance. The classification accuracy of the SVM-based multiple-cluster classifier constructed based on the eight selected variables in discriminating four different types of cells is impressive by achieving a total accuracy as high as 97.16%. On the other hand, the two-cluster predictive model designed using seven selected salient features enabled us to effectively distinguish the dysplastic from normal cells with a very high diagnostic accuracy (98.83%), sensitivity (91.40%), specificity (99.90%), and AUC (0.995). Its classification outcomes are much better than the results predicted by the model constructed based on the features selected using statistical or filter methods.

In conclusion, an automated image analysis system has been developed for morphologic measurement and evaluation of cervical cells. By using the designed image analysis system, we can establish objective quantitative evaluation from images obtained from different laboratories. More subsequent investigation with statistical measurements is needed to elucidate its practical utility in a laboratory and its ability to improve the diagnostic performance of a laboratory. It is expected to help diminish the number of the gray-zone diagnosis, especially between the atypical squamous cells with undetermined significance (ASCUS) and LSIL, as well as between the ASCUS and HSIL. This approach could also assist cellular classification in other cytological fields. Through a combination of findings from clinical, cytomorphological, and morphometric variables, we can resolve additional problems related to interobserver variation and ambiguous diagnosis.

ACKNOWLEDGMENT

The authors would like to thank Prof. J. L. Semmlow of Rutgers University for his help in editing this paper.

REFERENCES


Yung-Fu Chen received the Diploma degree from the National Taipei Institute of Technology, Taipei, Taiwan, in 1983, and the M.S. degree from the University of Missouri-Rolla, Rolla, MO, USA, in 1990, both in electrical engineering, and the Ph.D. degree in biomedical engineering from National Cheng Kung University, Tainan, Taiwan, in 2002. Since 1992, he has been with the Central Taiwan University of Science and Technology, Taichung, Taiwan, where he is currently a Professor in the Department of Healthcare Administration and an Adjunct Professor at the Institute of Biomedical Engineering and Material Science. He is also an Adjunct Professor in the Department of Health Services Administration, China Medical University, Taichung. His research interests include medical informatics, innovative applications of radiofrequency identification in business and medicine, and digital preservation.

Chung-Chuan Cheng received the B.S. degree in information management from the Chaoyang University of Technology, Taichung, Taiwan, in 1996. He is currently working toward the Ph.D. degree in the School of Electrical Engineering, National Chung Hsing University, Taichung.

His research interests include content-based image retrieval, information management, and web-based business system development.

Po-Chi Huang received the M.D. degree from China Medical University, Taichung, Taiwan, in 1997, and the M.S. degree in biomedical science from National Chung Hsing University, Taichung, in 2003. From 2008 to 2011, she was with the Department of Pathology, Taichung Hospital. She is currently with the Department of Pathology, Taipei Hospital, Department of Health, Executive Yuan, Taipei, Taiwan. Her research interests include cytopathology and medical education.

Tsung-Po Chen received the M.D. degree from National Yang Ming University, Taipei, Taiwan, in 1985. He is currently working toward the Ph.D. degree at the Institute of Medicine, Chung Shan Medical University, Tainan, Taiwan. From 1989 to 1994, he was with the Department of Surgery, Taichung Veterans General Hospital. He is currently the Superintendent of Tsung-Po Chen Clinics and the Director-General of Taichung City Clinic Association, Taichung. His research interests include cytopathology and clinical decision support system in colorectal practice.

Ker-Cheng Lin received the M.D. degree from National Defense Medical College, Taipei, Taiwan, in 1977, and the M.S. degree in public health from The Johns Hopkins University, Baltimore, MD, USA, in 1996. He was with Chishan Hospital, Department of Health, from 1998 to 2000 and from 2007 to 2010, serving as the Deputy Superintendent and the Superintendent, respectively. Since 2000, he has been with Nantou Hospital, Department of Health, Executive Yuan, Nantou, Taiwan, where he is currently the Superintendent. He is a gynecologist and an expert in breast cancer surgery.

Yung-Kuan Chan received the M.S. degree in computer science from the New Mexico Institute of Mining and Technology, Socorro, NM, USA, in 1991, and the Ph.D. degree in computer science and information engineering from National Chung Cheng University, Chiayi, Taiwan, in 2000.

From 2001 to 2002, he was an Assistant Professor in the Department of Information Management, Chaoyang University of Technology. From August 2002 to July, 2003, he was an Assistant Professor in the Department of Computer Science and Information Engineering, National Hwei Institute of Technology. From August 2003 to July 2005, he was an Assistant Professor, from August 2005 to July 2008, an Associate Professor, and currently a Full Professor and the Chairman of the Department of Management Information Systems, National Chung Hsing University, Taichung. His research interests include image processing, image retrieval, image compression, image hiding, and data mining.

Hsuan-Hung Lin received the Ph.D. degree in applied mathematics from National Chung Hsing University, Taichung, Taiwan, in 2010. He is currently an Associate Professor in the Department of Management Information Systems and the Director of Computer Center, Central Taiwan University of Science and Technology, Taichung. His research interests include biomedical information and bioinformatics.

John Y. Chiang received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from the Northwestern University, Evanston, IL, USA, in 1987 and 1990, respectively, all in electrical engineering. Since 1990, he has been with National Sun Yat-sen University, Kaohsiung, Taiwan, where he is currently a Professor in the Department of Computer Science and Engineering. His research interests include pattern recognition, image processing, and content-based image retrieval.