

Review Article

Intelligent Screening Systems for Cervical Cancer

Yessi Jusman, Siew Cheok Ng, and Noor Azuan Abu Osman

Department of Biomedical Engineering, Faculty of Engineering Building, University of Malaya, 50603 Kuala Lumpur, Malaysia

Correspondence should be addressed to Siew Cheok Ng; siewcng@um.edu.my and Noor Azuan Abu Osman; azuan@um.edu.my

Received 24 December 2013; Accepted 11 February 2014; Published 11 May 2014

Academic Editors: S. Balochian, V. Bhatnagar, and Y. Zhang

Copyright © 2014 Yessi Jusman et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Advent of medical image digitalization leads to image processing and computer-aided diagnosis systems in numerous clinical applications. These technologies could be used to automatically diagnose patient or serve as second opinion to pathologists. This paper briefly reviews cervical screening techniques, advantages, and disadvantages. The digital data of the screening techniques are used as data for the computer screening system as replaced in the expert analysis. Four stages of the computer system are enhancement, features extraction, feature selection, and classification reviewed in detail. The computer system based on cytology data and electromagnetic spectra data achieved better accuracy than other data.

1. Introduction

Cervical cancer is a leading cause of mortality and morbidity, which comprises approximately 12% of all cancers in women worldwide according to World Health Organization (WHO). In fact, the annual global statistics of WHO estimated 470 600 new cases and 233 400 deaths from cervical cancer around the year 2000. As reported in National Cervical Cancer Coalition (NCCC) in 2010, cervical cancer is a cancer of the cervix which is commonly caused by a virus named Human Papillomavirus (HPV) [1]. The virus can damage cells in the cervix, namely, squamous cells and glandular cells that may develop into squamous cell carcinoma (cancer of the squamous cells) and adenocarcinoma (cancer of the glandular cells), respectively. Squamous cell carcinoma can be thought of as similar to skin cancer because it begins on the surface of the ectocervix. Adenocarcinoma begins further inside the uterus, in the mucus-producing gland cells of the endocervix [2].

Cervical cancer develops from normal to precancerous cells (dysplasia) over a period of two to three decades [3]. Even though the dysplasia cells look like cancer cells, they are not malignant cells. These cells are known as cervical intraepithelial neoplasia (CIN) which is usually of low grade, and they only affect the surface of the cervical tissue. The majority will regress back to normal spontaneously. Over time, a small proportion will continue to develop into cancer.

Based on WHO system, the level of CIN growth can be divided into grades 1, 2, and 3. It should be noted that at least two-thirds of the CIN 1 lesions, half of the CIN 2 lesions, and one-third of the CIN 3 lesions will regress back to normal [3]. The median ages of patients with these different precursor grades are 25, 29, and 34 years, respectively. Ultimately, a small proportion will develop into infiltrating cancer, usually from the age of 45 years onwards.

In 1994, the Bethesda system was introduced to simplify the WHO system. This system divided all cervical epithelial precursor lesions into two groups: the Low-grade Squamous Intraepithelial Lesion (LSIL) and High-grade Squamous Intraepithelial Lesion (HSIL). The LSIL corresponds to CIN1, while the HSIL includes CIN2 and CIN3 [4].

Since a period of two to three decades is needed for cervical cancer to reach an invasive state, the incidence and mortality related to this disease can be significantly reduced through early detection and proper treatment. Realizing this fact, a variety of screening tests have therefore been developed in attempting to be implemented as early cervical precancerous screening tools.

2. Methodology

This paper reviews 103 journal papers. The papers are obtained electronically through 2 major scientific databases:

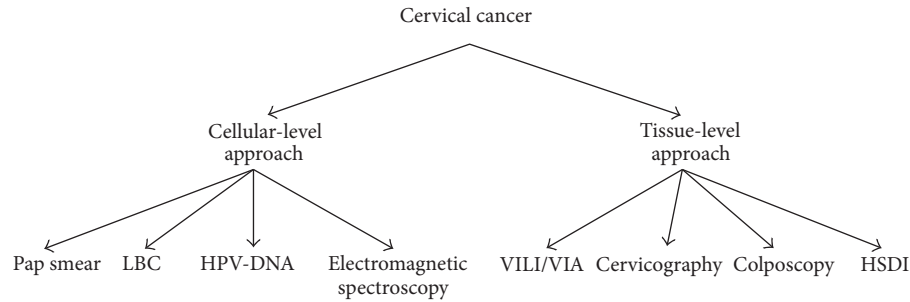


FIGURE 1: Taxonomy of cervical cancer screening.

TABLE 1: Comparison of the ability of the manual cervical screening methods.

Highlighted features	Cellular level				Tissue level			
	Pap smear	LBC	HPV-DNA	EMS	VILI/VIA	Cervicography	Colposcopy	HSDI
Low cost	V*	V*	V	V*	V	V	V*	V*
Short time	X	X	X	X	V	V	V	V
Not Subjective	X	X	V	V	X	X	X	X
Possible in real time	V	V	X	V	X	V	V	V

Google Scholar (<http://scholar.google.com.my/>) and Scopus (<http://www.scopus.com/home.url>). In the databases, the IEEE and Science Direct databases will be included already. Since there are various aspects being reviewed here, four sets of keywords have been used. The first set contains Cervical Cancer, Feature Extraction, and Intelligent System, which give an overview of an intelligent system for cervical cancer detection. The second set contains Cervical Cancer, Image Processing, and Intelligent System. The third set is made up of Cervical Cancer, Image Processing, and Classification. The final set contains Cervical Cancer, Features Extraction, and Image Processing.

In order to ensure a quality review, the academic papers reviewed here are limited to peer reviewed journal papers. Recent conference papers published in the year 2010 onwards are also considered as the work is up to date and the journal related to this work has yet to be published. However, certain conference papers that showed excellent results or used methods that are currently unpopular are also included to give a more complete perspective of the work done in this field.

3. Screening for Cervical Carcinoma

Screening programs for cervical cancer have been implemented in developing countries for decades and have shown to be effective in reducing the overall mortality from this disease. There are two main diagnostic screening approaches for cervical cancer as presented in Figure 1:

- (1) diagnostic screening approach based on cellular level (i.e., Pap smear, liquid based cytology (LBC), HPV-DNA testing, and electromagnetic spectroscopies);
- (2) diagnostic screening approach based on the tissue level (i.e., visual inspection after applying Lugol's iodine (VILI) or acetic acid (VIA), cervicography,

colposcopy, and hyperspectral diagnostic imaging (HSDI)).

For diagnostic screening based on cellular-level, the specimen collections are required before it is analyzed for the expert analysis results. In contrast, specimen collection is not required for diagnostic screening based on tissue-level. The expert analysis is required for cervix images visually after applying certain liquid into the cervix surface. Detail of standard procedure, advantages, and disadvantages for Pap smear, LBC, HPV-DNA, VILI/VIA, cervicography, and colposcopy techniques can be found in [5].

On the other hand, current technologies have investigated the cervical cell from the specimen under the spectroscopy equipment inducing an electromagnetic light. There are several techniques utilized for cervical cancer detection:

- (1) image results: fluorescent in situ hybridization (FISH) [6–11];
- (2) spectra results: Raman spectroscopy [12, 13], fluorescence spectroscopy [14, 15], and Fourier transform infrared (FTIR) spectroscopy [16–24].

On the other hand, there is an alternative technique based on tissue level known as hyperspectral diagnostic imaging (HSDI). The surface of the cervix is scanned with ultraviolet and white light for detecting lesions [25–27]. The scanning is achieved one line at a time, with the scan time varying from 12 to 24 seconds. By taking a series of scan lines, a hyperspectral data cube is obtained. This hyperspectral data cube contains spatial information (pixels) in two dimensions and spectral information (bands) in the third dimension [27]. This technique produces a 3D cervix image that is easier to interpret.

Based on the references, the techniques have several features required for considerations as summarized in Table 1. Each of the technique has advantages and disadvantages

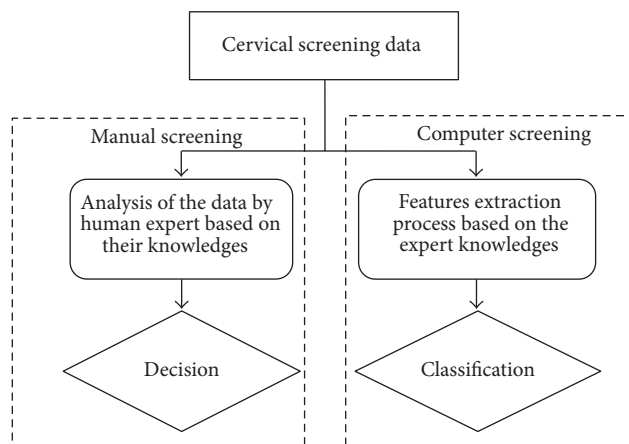


FIGURE 2: Comparison of analysis screening system by human expert and machine.

individually. Almost all of the techniques have on average nonexpensive or low cost features [5, 28, 29]. However, the EMS machines as well as microscope (for Pap smear and/or ThinPrep) and high resolution camera (for colposcopy and/or HSDI) are quite expensive to be bought for the beginning proses as screening technique but it is cost effective in the long run as no analysis from pathologist is required.

For the cellular-level techniques, the specimen collections require certain duration time and the results cannot be obtained spontaneously after specimen collection process due to need of the next process for the expert reading (i.e., image, spectrum, genetic material, etc.). As for the tissue-level, analysis of the experts could be obtained after reading the images captured by the camera.

Based on Table 1, the HPV-DNA is not subjective due to the genetic material for chemistry analysis on the cell. Similarly, the EMS techniques are also not subjective. They have quantitative results used for analysis. However, the HPV-DNA and the VILI/VIA techniques are not possible to interface in real time so they cannot be developed into an intelligent system. Therefore, intelligent systems for cervical precancerous is limited to the six possible techniques in real time as presented in Table 1.

4. Intelligent System Approach to Cervical Cancer

The cervical screening methods mentioned in Section 3 are highly dependent on the skill of the experts. However, their judgment may be subjective and often leads to considerable variability [5]. Aside from that, the limited number of experts and the large number of patients resulted in a long queue for the screening process. To overcome these problems, computational tools have been developed for automated cancer diagnosis as drawn in Figure 2. The automated cancer diagnosis facilitates objective judgment complementary to expert's decision.

Figure 2 shows the principle comparison of the computer screening technique and the human expert. The feature's

extraction and classification by the computer replace the analysis and decision of human experts. Currently, the requirement for analysis based on computer screening increases. A number of researches were carried out specifically with the attempts to automate the classification [30, 31]. The results of several research to indicate that computer-imaging-assisted screening significantly increases the detection of cervical abnormalities compared to the manual screening [32, 33]. Consequently, automated screening devices would be a tremendous improvement for reducing the likelihood of human errors.

A typical computer screening system involves four stages, namely, data enhancement, features extraction, features selection, and classification as shown in Figure 3. Aside from visual inspection after applying Lugol's iodine (VILI) or acetic acid (VIA) and HPV-DNA Testing, the data from the other screening techniques can be digitalized and fed into the intelligent computer screening system. These data can be categorized as images or spectra.

In the enhancement stage, the image or spectra will be processed in order to eliminate the noise to increase the signal to noise ratio. For images, this stage also involves determination of the region of interest to be segmented out for further processing. For the images, features are extracted either at the cellular or at the tissue-level. Basically, the morphology, texture, shape, and/or intensity of the cell/tissue image are extracted as features. For spectra, the features are height of intensity, shift of wave number, and corrected area and area under peaks of the spectra.

The main purpose of feature's selection is to reduce the number of features used in classification while maintaining acceptable classification accuracy. Feature's selection includes methods such as sequential backward selection [34], sequential forward selection [35], sequential floating search method [36], discriminant analysis [37], and principal component analysis [17].

After features selection step, several classifiers can be employed to obtain classification performance based on the used features. Different classification results can be performed by the different features used [38]. The aim of

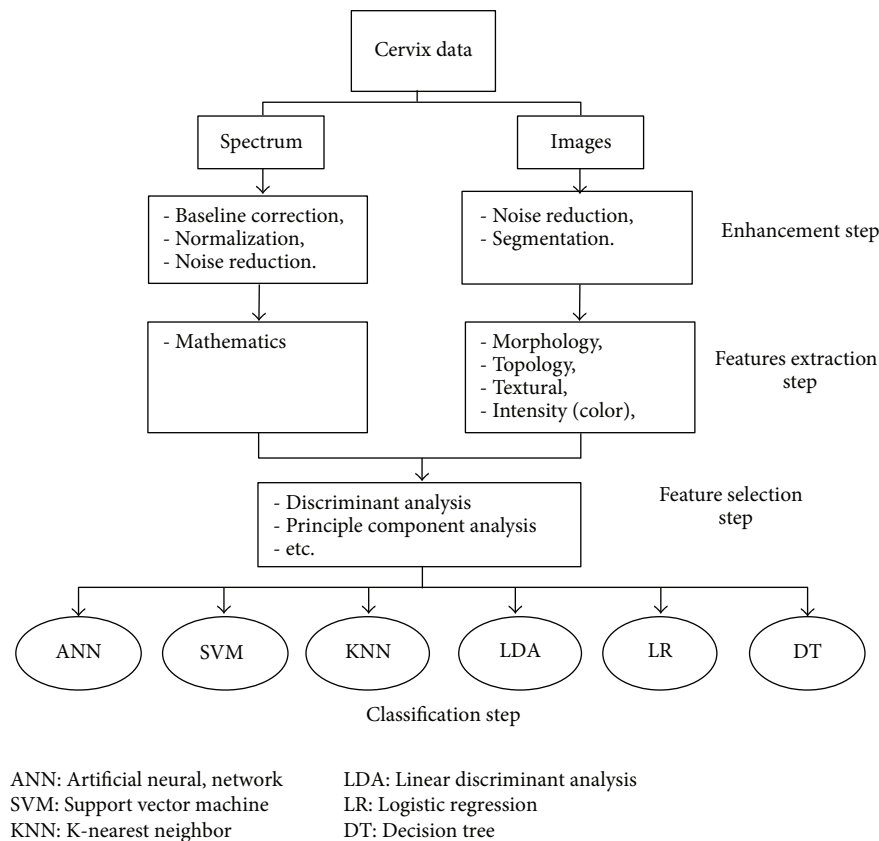


FIGURE 3: Intelligent cervical cancer classification systems.

TABLE 2: Information about cervical screening instruments.

Information	PAPNET	AutoPap 300	FocalPoint	TIS
Input data	Pap smear only	Pap smear only	Pap smear and ThinPrep	ThinPrep only
Characteristic	Semiautomatic system	Automatic system	Automatic system	Automatic system
USFDAapproval	Secondary screening	Primary screening	Primary screening	Primary screening

diagnosis step is to distinguish benignity and malignancy or to classify different malignancy levels by making use of extracted features. This step uses statistical analysis of the features and machine learning algorithms to reach a decision. An overview of these four stages is given in Figure 3. In the following sections, we will study each of these steps in detail.

Nowadays, there are several instruments which have been used to screen for abnormal cervical cells such as semi-automated or interactive system (PAPNET) and automated systems (AutoPap 300, FocalPoint, and ThinPrep Imaging System (TIS)) [30, 33, 39–41]. These instruments have been approved by United States Food and Drug Administration (USFDA) for screening system. These instruments utilize algorithmic image analysis to extract morphological features. Most of these systems help the expert to perform better diagnosis by improving cervical cell images quality so that the morphological features can be seen easily. Table 2 summarizes the instruments to view their advantages and disadvantages.

In fact, to build the current intelligent cervical screening system, two types of raw data (i.e., digital images and spectra) as presented in Section 3 can be used for the purposes. To construct the intelligent system, data enhancement (optional), features extraction, and classification steps are applied to the raw data to obtain good screening results approach of the human expert knowledge in some areas of their expertise [42, 43]. Therefore, here we review some current features extraction techniques and classification of two types of cervical data.

4.1. Data Enhancement. As stated in earlier section, there are two types of cervical cancer data, which are spectrum and image as presented in Figure 4. The main aim of the enhancement stage is to reduce noise and for the image data to determine the area of interest as well. Due to a considerable amount of noise that arises from the staining process, it is usually necessary to reduce the noise prior to the segmentation process. In some studies, noise reduction and segmentation are carried out at the same time.

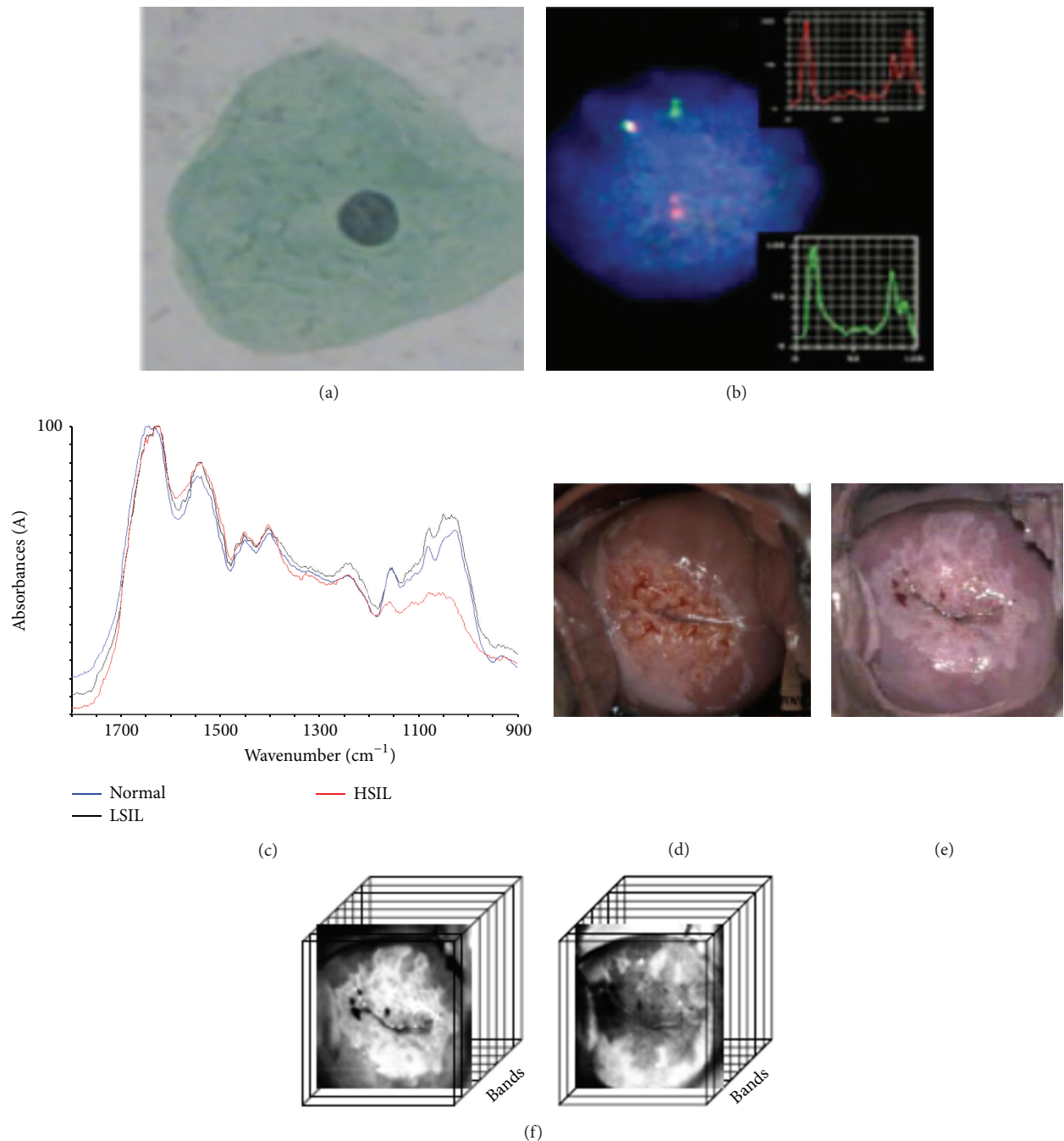


FIGURE 4: Cervical data used for intelligent classification. Cellular-level features; (a) cytology image, (b) FISH image, and (c) optical spectra. Tissue-level features; (d) cervicography, (e) colposcopy, and (f) optical image (HSDI).

The aim of noise reduction for the spectrum is to reduce high frequency noise contained in the spectrum that can be from either noise conducted through power lines or radiated through the hot air in the electromagnetic spectroscopy equipment [44]. Savitzky-Golay (SG) filter is currently being used widely for smoothing the spectroscopy spectra [45–52]. The SG filter has boundary problems which can be solved by using other techniques such as Binomial and Chebyshev filters [53–55].

For image data, image noise is random (not present in the object image) variation of brightness or color information in images and is usually an aspect of electronic noise. The noise is an undesirable by-product of image capture that adds spurious and extraneous information. It can compromise the level of detail in cervix image, and so reducing this noise can greatly enhance the image. There are several noise reduction techniques offered by many researchers for the automated cervical cancerous applications system as follows.

- (i) Based on pixel intensity: thresholding [43, 56–60] and filtering techniques [57, 60, 61].
- (ii) Based on shape: mathematical morphology [58, 60, 62].
- (iii) Based on the gradient: [63, 64].

Thresholding and filtering are to reduce the noise by making use of the pixel intensities. In threshold, the intensity histogram of an image is employed to determine the threshold value where the pixels are considered to be noise. For example, the Otsu method determines an optimal threshold which minimizes the within-class variance [62]. This method yields satisfactory results when the numbers of pixels in each class are close to each other. One weakness of threshold is that all pixels under the threshold value can be noise even the pixel information which is important. Conversely, the pixels over the threshold value can be information even the pixels which are noise. In filtering, the value of a pixel is transformed to a new value which is computed as a function of the values of pixels located in a selected neighborhood around this particular pixel. This is an improvement over the threshold method.

Another method for noise reduction which reduces the noise based shape characteristics of the input image is to use mathematical morphology. The basic morphological operators are the erosion and dilation of the set with a structuring element. These two basic transformations give two other transformations known as opening and closing. Opening is the erosion of an image followed by the dilation; it breaks narrow isthmuses and eliminates small objects and sharp peaks in the image. On the other hand, closing is the dilation of an image followed by the erosion; it fuses narrow breaks and fills tiny holes and gaps in the image [58, 65]. This technique can enhance region of interest (ROI) of the images perfectly by removing and adding small shape in the focused images.

Meanwhile, the segmentation process is used to detect the region of interest in the cervical image. The process is a key procedure in automating computer-aided diagnostic systems, because accurate images segmentation could help to reduce the processing time and increase the sensitivity rates. The segmentation method should be chosen depending on the type of the features to be extracted. Several segmentation techniques have been proposed and applied in cervix images as follows.

- (i) Based on shape: [57, 58, 66].
- (ii) Based on color: [61, 67–70].
- (iii) Based on texture: [61, 71].
- (iv) Based on contour: [59, 72–74].

4.2. Features Extraction. Automated cervical cancer diagnosis relies on using the information obtained from (i) the abnormalities in the cell structures (*cellular-level*) and (ii) the abnormalities in the cell distribution across the tissue (*tissue-level*). Many researchers have applied various captured techniques for the automated classification of cervical cancer. The techniques are cytology, FISH, and electromagnetic

scanner for cellular level while cervicography, colposcopy, and HSDI are used for tissue level. Features are then extracted from data of the techniques as presented in Table 3.

The features are extracted to quantify these changes in a given tissue. In order to measure the abnormalities at the cellular/tissue level, size and shape, ratio, topology, texture, and color intensity can be used as features listed in Table 3. The features are extracted and represented by a value to be used in the intelligent system.

4.2.1. Size and Shape Feature. A cell includes a nucleus surrounded by cytoplasm. As a traditional way, a pathologist evaluates the cytoplasm and the background of slide. The abnormality features are described as size (i.e., there is an increased size of the nucleus compared to the cytoplasm), shape (i.e., smooth, circular, and oval outline belongs to a normal nucleus), texture (i.e., rough textures belong to an abnormal nucleus), chromaticity (i.e., abnormal nucleuses are darker than normal ones) [62]. The quantification of these properties enables differentiating the malignant cells from those of benign and normal.

The size is expressed by the radius, area, and perimeter of the cell. Suppose that $S = \{s_1, \dots, s_n\}$ is a set of the boundary points of a segmented cell/nucleus and C is the centroid of these boundary points, a sample of a nucleus with its boundary points. On the other hand, the shape is expressed by the length of the major and minor axes, symmetry, and circularity. The size and shape features defined on the set of the boundary points, S , are given as follows.

- (i) Radius r is defined as the average length of the radial lines towards every boundary point. Mathematically,

$$r = \frac{\sum_{i=1}^n |s_i C|}{n}. \quad (1)$$

- (ii) Area is the number of pixels within the boundary.
- (iii) Perimeter P is measured as the sum of the distances between every consecutive boundary point. Mathematically,

$$P = |s_n s_1| + \sum_{i=1}^{n-1} |s_i s_{i+1}|. \quad (2)$$

- (iv) Major axis is the longest chord that goes through the center and minor axis is the line that is perpendicular to the major axis and that goes through the center.
- (v) Circularity is quantified by drawing chords between nonadjacent boundary points and checking whether or not the boundary points lie inside these chords.

Several researchers have identified capability of the size and shape features to classify the cervix using the cytology image [72, 73, 75], FISH image [60, 76, 77], and electromagnetic spectrum [24, 78]. Besides these features, the ratio of the same feature for different parts of a biological structure is used as another feature. For example, the nuclear area/cytoplasm area ratio [73] and the corrected area under peak A/under peak B ratio [78] are such a kind of features.

TABLE 3: The list of features that are extracted by different data.

	Cellular-level based features			Tissue-level based features		
	Cytology	FISH	Electromagnetic spectra	Cervicography	Colposcopy	HSDI image
Size	(i) Area of Cell [72], (ii) Area of Nucleus [72, 73, 75] (iii) Area of Cytoplasm [75].	(i) Area for each coloured spot [60, 76, 77]. (ii) Radius of each coloured spot [60].	Shift of peak frequency [24]	Perimeter of anatomical features [63, 79]	Perimeter of anatomical features [80]	Perimeter of acetowhite [27]
Shape	(i) Circularity of cytoplasm [75] (ii) Circularity of nucleus [24, 75]	Circularity of each coloured spot [60, 77].		(i) Circularity of cervix [66] (ii) Circularity or elliptical shape of Os region [66]		
Ratio	(i) Percentage of cell coverage [72] (ii) Ratio of nucleus to cytoplasm size [72, 75] (iii) Percentage of empty cells [72].		(i) Ratio of peak intensities [24, 78] (ii) Ratio of area under peaks [78]			
Topology	(i) Distribution of cell [72] (ii) Distribution of nucleus [72].	(i) Distances between the same color spots [60, 77]. (ii) Distance between the centers of the two spots [60, 77]. (iii) Center of gravity for each coloured spot [60]. (iv) Number of red and green spots [60, 76, 77].				
Texture	(i) Multinucleus cells [72]. (ii) Halos in cells [72].			Acetowhite region [86–88].	Acetowhite region [61, 71, 89–91]	
Color intensity	(i) Cell [72, 83] (ii) Nucleus [73] (iii) Cytoplasm [73]	Intensity of each coloured spot [60].		Anatomical features [86, 93, 94]	Anatomical features [61, 67, 91, 95–97].	

From cytology images as presented in Figure 4(a), the specific features as listed in Table 3 (i.e., size, shape, and ratio), namely, average nucleus size [72, 73, 75], average cytoplasm size [75], average cell size [72], cytoplasm circularity [75], nucleus circularity [24, 75], percentage of cell coverage [72], ratio of a nucleus to cytoplasm size [72, 75], and percentage of empty cells [72], are partially used to be an input attribute to the classification system.

For FISH image, the features from labeled biomarker spots of chromosomes 3 (red spot) and X (green spot) are the size of each colored spot [60, 76, 77], the effective radius of each red or green spot computed as the radius of a circle that had the same size as the colored spot [60, 76, 77], and the circularity of each colored spot [60, 77].

Meanwhile, from the electromagnetic spectra, the features are shift of peak frequency [24], absorbance value, and area under the spectra. For the absorbance features, the corrected absorbance value and ratio of the

absorbance/corrected absorbance values for certain regions in a spectrum are derived from the features [24, 78]. Then, from area under the spectra, the features can be taken as corrected area and ratio of the area/corrected area values for certain regions in one spectrum [78].

At the case of tissue-level image, the shape feature is applied to differentiate the cervix images. The anatomical region features of the cervix (as marked by the medical experts) can be characterized by their elliptical or circular shapes; hence, the ellipse and the circle are chosen for the shape models. A vast amount of work was done to embed prior-shape information into a segmentation task. A popular approach is to use prior models based on allowable deformation of a template shape [66]. In addition, for tissue level case, the AW perimeter obtained after Lugol's solutions was assessed by examining the topography of the perimeter lines cut across the image contour with lines positioned in radial direction [27, 63, 79, 80].

Several techniques are applied to extract the size and shape features:

- (i) thresholding technique [60, 62, 77, 81];
- (ii) clustering technique [70, 73];
- (iii) fuzzy technique [69];
- (iv) wavelet technique [82, 83];
- (v) statistic techniques [13, 22, 78, 84].

At cellular-level, the size and shape features in cytology images are extracted using thresholding [62], clustering [70, 73], fuzzy [69], and wavelet techniques [83]. In the FISH images, the features are extracted using thresholding [60, 77, 81]. Besides, in the electromagnetic spectra, the features are extracted using statistical techniques [13, 22, 78, 84] and wavelet technique [82]. At tissue-level, perimeters were analyzed in terms of their topology changes such as perimeter' peaks [80]; van Raad et al. [68] used landmark technique of the closed contours to extract the perimeter features which differentiate normal and abnormal cervix. In the HSDI image case, the perimeter feature is extracted using landmark technique after an enhancement process [27]. Automated landmark extraction, including the extraction of the cervix boundary, detection of the Os (one of the anatomic region), and detections (and elimination) of specular reflections are used by [63, 79, 85].

4.2.2. Topology Features. The topological features provide information on the structure of a tissue by quantifying the spatial distribution of its cells. For that, this approach encodes the spatial interdependency of the cells prior to the feature extraction. The features are applied for cellular-level case. The specific features implemented for cytology images are distribution of cell [72] and distribution of the nucleus [72], while the distances between the same color spots [60, 77], the distance between the centers of the two spots [60, 77], the gravity center of each colored spot [60], and the total number of red spots and green spots [60, 76, 77] have been implemented in the FISH images. The thresholding techniques are applied to extract the features in the cellular-level case [60, 72].

4.2.3. Textural Features. Texture is a connected set of pixels that occurs repeatedly in an image. It provides information about the variation in the intensity of a surface by quantifying properties such as smoothness, coarseness, and regularity. At the cellular level, the existence of multinuclear cells [72] and the existence halos in cells [72] are used as features in the cytology image. Meanwhile, at the tissue level, the texture features are extracted from the AW region of the cervix image [61, 71, 86–91]. The texture is formed after giving the acetic acid or Lugol's iodine to the cervix surface as a sign of the abnormality.

There are several techniques applied for extracting the textural features in cervix images as follows.

- (i) Wavelet technique [89].
- (ii) Mathematical morphological operations [71, 90].

(iii) Clustering technique [86].

(iv) Thresholding technique [62, 88].

At the cervix image of tissue-level, van Raad [89] demonstrated Gabor wavelet for extracting the textural features which outline the area of metaplastic changes, known as the transformation zone (TZ). The performances of the Gabor wavelet scheme achieve close to 80% accuracy in discrimination on the ROI. On the other hand, textural features (i.e., mosaic pattern) within the AW region are obtained from skeletonized vascular structures uniquely. The skeletonized vascular structures represented typical vascularity embedded in the normal and abnormal regions extracted by a series of mathematical morphological operations [71]. The series of mathematical morphological operations are gray-scale method, top hat transform, morphological opening with a rotating structuring element (ROSE), thresholding, and skeletonizing. Similarly, the textural features are extracted based on iterative morphological operations with various sizes of structural elements, in combination with adaptive thresholding [90]. Furthermore, combination of mathematical morphology and clustering based on Gaussian mixture model (GMM) is proposed to extract the textural features in the cervix image [86]. The algorithms are used to segment macro regions of the textural cervix images. Thresholding technique is used to segment tissues and nucleus as the texture for each application, respectively [62, 88, 92].

4.2.4. Color Intensity Based Features. The color intensity-based features are extracted from the gray-level or color histogram of the image. This type of features does not provide any information about the spatial distribution of the pixels. The intensity histogram in a cell is employed to define features. In the case of cellular level images, the difference of color intensity can be used as features for the cancerous cells [72, 73]. Cytology image has a relatively darker color intensity composition than normal cells. The distinguishable patterns can be analyzed using the corresponding image's color intensity histogram [83]. Meanwhile, another feature to differentiate the abnormality of cervix using FISH image is the average intensity of each colored spot [60]. At the case of tissue level images, the changes in color and intensity correlate closely with changes in tissue type, severity of cervical neoplasia, and vessel patterns [61, 67, 86, 91, 93–97].

Several techniques used for extracting the intensity features are as follows.

- (i) Clustering technique [67, 86, 95, 97].
- (ii) Watershed technique [93, 94].
- (iii) Statistical technique [96].

van Raad [67] used a clustering technique (i.e., GMM) based MAP algorithm probability model in cervical images to extract color information features belonging to each of the tissue types in the cervix, such as the cervical canal (CC), the transformation zone (TZ), the squamous epithelium (SE), and the artifact named specular reflection (SR). Besides, mean-shift clustering is used to extract color and texture features of a tissue type [95]. Clustering based on the GMM

TABLE 4: The list of classifiers that are used by different studies.

	Cellular-level based features			Tissue-level based features		
	Cytology	FISH	Electromagnetic spectra	Cervicography	Colposcopy	HSDI image
Artificial Neural network	(3/1241/10/78.7) [56], (2/400/10/99) [72], (3/550/4/97.5) [73], (5/78/5/91.4) [75].		(2/361/13/74.4) [84], (2/201/3/87) [82], (3/780/22/97.4) [78].		(2/283/7/95.8) [80].	
Support vector machine			(3/63/10/72) [116]			
Logistic regression			(4/145/—/88) [12]			
K-nearest neighbors					(2/283/7/68.9) [80], (2/48/10/76.5) [97].	(7/371/5/95.96) [117]
Linear discriminant analysis	(5/230/15/60.4) [85].		(2/324/—/78) [118], (2/275/8/96.4) [22], (2/150/5/99.5) [13], (4/800/7/90) [24], (2/92/3/97.6) [119]	(2/100/—/78.5) [88].	(2/40/4/87.2) [61].	
Decision trees	(3/1241/10/77) [56], (—/61/2/96.7) [120].	(2/325/—/93.6) [60]		(2/211/—/78) [94].	(2/29/—/86) [121], (2/99/—/88.5) [96].	

The values given in bracket are number of classes/number of data/number of features used/accuracy.

is used in a joint color and geometric feature space to segment macro regions [86]. Similarly, [97] used a clustering technique (i.e., *K*-means clustering (KMC)) to generate an anatomical feature map for each cervical tissue type. The tissue regions defined by the anatomical feature map are further clustered into subregions. Watershed technique is used for a specific focus on the detection of lesion regions in uterine cervix images [93, 94]. Meanwhile, the spatial change of the AW lesion is extracted using color and texture information based on an opacity index that indicates the grades of temporal change [96].

As presented in Figure 4, possibility of the ratio and texture features can be extracted from FISH image for future works. As listed in Table 3, the features of the FISH image are area and radius for each colored spot. The ratio of the area for one colored spot and other colored spot can be possibly extracted. The ratio of the radius of one colored spot and other colored spots can be also possibly extracted. The texture of one FISH image integrally can be also extracted to differentiate the abnormality of the images.

4.3. Features Selection. After all the possible features for classification had been extracted, the selection of significant or dominant features can be conducted. Besides feature's extraction systems, the classification performance also depends on the selected features and the classification technique used. Feature selection is an important stage in classification, especially if it involves a large dimension of input features. By applying this feature selection stage, the original high

dimensional inputs could be transformed and reduced into new lower dimensional features [98].

Generally, all possible extracted features can be used as the inputs for a classification system. However, irrelevant or noisy features could deteriorate between classes and increase the overlap in a non-linear manner. The noisy features can mix up the boundaries for the generalization performance of the classification system [99]. A classifier with fewer inputs needs fewer weights to be adjusted, leading to better generalization and faster training [100]. Adding newer features can significantly lead to a reduction in the performance of the classification system [100].

Many researchers in computer vision based spectroscopy data applied the features selection techniques for cervical cells and other cell features [13, 51, 58, 84, 98, 101–110]. Generally, good performances in classification are achieved after applying the features selection techniques. Since the spectral data is heavily redundant, the selection of the significant wavelength as features in this case is vital.

For the image processing application, discriminant analysis (DA) and principle component analysis (PCA) are methods commonly used to find a linear combination of features which characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier or, more commonly, for dimensionality reduction before later classification [111].

The DA works by creating a new variable called the discriminant function score which is used to predict to which group a case belongs. The discriminant function scores are

computed similarly to factor scores (i.e., using eigenvalues). The computations find the coefficients for the independent variables (features) that maximize the measure of distance between the groups defined by the dependent variable. The disadvantages of the DA are the distribution of distance matrices in the same class to be singular if the dimension of the data is much higher than the number of training samples [112].

Besides, the PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system. In the PCA, the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinates, and so on. The first principal component corresponds to a line that passes through the multidimensional mean and minimizes the sum of squares of the point's distances from the line. The second principal component corresponds to the same concept after all correlations with the first principal component have been subtracted out from the points.

There are several researchers in cervical cancer application who use the PCA [13, 17, 113]. In their researches, the PCA is used as dimensionality reduction to improve the classification performance and decrease the training time of classifier. However, the disadvantages of the PCA consist in the fact that the directions maximizing variance do not always maximize information. In case, a great disadvantage of PCA is that it does not consider any class information [114]. This can lead to a loss of important discriminating information. In fact, the analysis showed that it was practically impossible to improve the classification error by this method [114]. Another disadvantage of the PCA is that it has high memory and computational requirements [115].

4.4. Classification. The effectiveness of the automatic cervical precancerous screening system is evaluated in this section. The classifiers mostly used for cervical cancer study in detail are artificial neural networks or neural network (NN) [56, 72, 73, 75, 78, 80, 82, 84], support vector machine (SVM) [116], logistic regression [12], K -nearest neighborhood (KNN) [80, 97, 117], linear discriminant analysis (LDA) [13, 22, 24, 61, 85, 88, 118, 119], and decision trees [56, 60, 94, 96, 120, 121], as listed in Table 4. The performances of the classifiers generally showed good results as presented in Figure 5.

Generally, each type of classifiers can be employed for all types of data. For example, the FISH or cervicography data can be classified using NN, SVM, logistic regression, KNN, LDA, and decision tree. However, it is important to know the advantages and disadvantages of the classifiers that might be considered as alternatives. Logistic regression is attractive for probability prediction, because it is mathematically constrained to produce probabilities in the range $[0, 1]$ and generally converges on parameter estimates relatively easily [122]. The disadvantages of the logistic regression are not designed to deal with high-dimensional data and cannot approximate any smooth polynomial function, regardless of the order of the polynomial or the number of interaction terms [122].

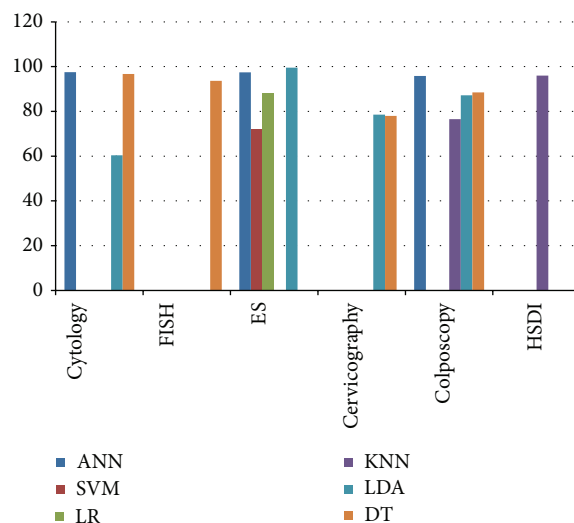


FIGURE 5: Performances of six classifiers generally for cervical precancerous data.

SVM's execution speed is very fast and there are no parameters to tune except the constant C . It is remarkably intolerant of the relative sizes of the number of training examples of the two classes. Since the technique is not directly trying to minimize the error rate, but trying to separate the patterns in high dimensional space, the result is that SVM is relatively insensitive to the relative numbers of each class. The possible disadvantages are large memory requirement [123] and the training time can be very large if there are large numbers of training examples [124].

Meanwhile, the NN architecture is initially not structured and the learning algorithm is responsible for the extraction of the regularities present in the data by finding a suitable set of synapses during the process of observation of the examples. Thus, NNs solve problems by self-learning and self-organization [125]. However, the neural network required long training time, and the results depend on the initialization parameters. It consisted of an arbitrary number of layers, and parameters [122]. Different combinations of number of hidden neurons, learning rate, momentum rate, activation function, epoch size, and initial weights have to be tried in order to produce better results [125].

Decision tree is relatively easy to interpret and to implement. Like SVMs and NNs, many methods for decision trees do not provide a probability of class membership, although some variants, in particular, classification and regression trees, do provide such probabilities. However, performance of all decision trees is dependent on both their method of construction and the amount of pruning (removal of highly specific nodes) performed [122].

KNN and LDA are methods implemented in numerous programs and easy to be implemented as classification tools. Both techniques have direct analytical solution and very good at detecting global phenomena (whereas decision tree detects local phenomena). However, they are simply defined and implemented, especially if there is insufficient data to

adequately define sample means and covariance matrices. Both techniques only detect linear phenomena and are sensitive to individuals outside the norm.

From Table 4 and Figure 5, the NN results showed constantly higher performance results in terms of accuracy than the other classifiers. The result ranges are 78.7–99% of accuracy. Mostly, the accuracy results are higher than 90%. In detail, [56] achieved 78.7% of accuracy in their preliminary study in classifying more than 1000 data to be two classes. [72, 73, 75, 78, 80] successfully achieved more than 90% of accuracies (i.e., 99%, 97.5%, 91.4%, 95.8%, and 97.4%) to classify 400 data to be 2 classes, 550 data to be 3 classes, 78 data to be 5 classes, 780 data to be 3 classes, and 283 data to be 2 classes.

As presented in Table 4, six types of data are used for classification purpose. All data have good capability to be used as intelligent classification data. The classification performances of the data are spread from range 60 to 99% of accuracies. Overall performance shows that cytology features and the electromagnetic spectra features give the higher accuracy than the other data. Many of the researchers that use the data gain accuracy values more than 90% such as performances using cytology data: 96.7% [120], 99% [72], 97.5% [73], and 91.4% [75] and performances using electromagnetic spectra data: 96.4% [22], 99.5% [13], 90% [24], 97.6% [119], and 97.4% [78]. Only few have performance less than 90% of accuracy.

The cytology combined with neural network gives the accuracy of up to 99% of accuracy to classify 400 data to be 2 classes, followed by neural network using the electromagnetic spectra features at 97.4% for classifying 780 data to be 3 classes. Greatly, the electromagnetic spectra features could achieve the higher accuracy only using discriminant analysis at 99.5% of accuracy. Therefore, based on Table 4, the better cervix data used for the automated diagnosis are the cytology and the electromagnetic spectra features and the best classifier used for the automated diagnosis system is neural network.

As reviewed, the intelligent classification system for cervical precancerous cells has been attempted and developed using two types of input attributes; cervical cell/tissue images and cervical cell spectra. Therefore, the systems have employed image and signal processing techniques for extracting features as the input attributes, respectively. Both systems could classify the cervical precancerous cells with high performances. The applications of image and spectra processing and classifier for cervical precancerous classification have been developed by many researchers in the world. The screening techniques have been proven to have better performance than the manual techniques. Thus, the intelligent classification system for cervical precancerous using the image and/or optical spectra as input is believed to have better classification performance and could be used as a second opinion to pathologists.

5. Summary

Six types of cervical precancerous data (i.e., cytology, FISH, electromagnetic spectra, cervicography, colposcopy,

and HSDI) generally can be used for the intelligent screening of cervical cancer. Computer screening system for cervical cancer based on cellular level data, namely, cytology, FISH, and electromagnetic spectroscopy, achieved better results as compared to tissue level data such as cervicography and colposcopy.

Classification tools (i.e., ANN, SVM, logistic regression, KNN, LDA, and decision tree) generally can achieve good performances to classify the cervical precancerous data. The screening systems based on neural network technique are frequently applied due to the better results and potential of the technique to build a real time system.

The long training time of the neural network can be reduced by using the features selection stage in the computer screening system. The dimensionality reduction popularly done by using discriminant analysis and principal component analysis can be developed using new techniques that can be proposed as future work in this research field. The developed techniques will reduce the training time and improve the classification result of the neural network.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research is supported by UM High Impact Research Grant UM-MOHE UM.C/625/1/HIR/MOHE/14 from the Ministry of Higher Education Malaysia.

References

- [1] NCCC, "Cervical cancer," <http://www.nccc-online.org/index.php/cervicalcancer>.
- [2] ACS, "What is cervical cancer?" 2011, American Cancer Society, <http://www.cancer.org/Cancer/CervicalCancer/Detailed-Guide/cervical-cancer-what-is-cervical-cancer>.
- [3] H. S. Cronjé, "Screening for cervical cancer in the developing world," *Best Practice and Research: Clinical Obstetrics and Gynaecology*, vol. 19, no. 4, pp. 517–529, 2005.
- [4] K. Frankel and M. K. Sidawy, "Formal proposal to combine the papanicolaou numerical system with Bethesda terminology for reporting cervical/vaginal cytologic diagnoses," *Diagnostic Cytopathology*, vol. 10, no. 4, pp. 395–396, 1994.
- [5] R. A. Kerkar and Y. V. Kulkarni, "Screening for cervical cancer: an overview," *Obstetrics and Gynecology of India*, vol. 56, no. 2, pp. 115–122, 2006.
- [6] P. Segers, S. Haesen, P. Castelain et al., "Study of numerical aberrations of chromosome 1 by fluorescent in situ hybridization and DNA content by densitometric analysis on (pre)-malignant cervical lesions," *Histochemical Journal*, vol. 27, no. 1, pp. 24–34, 1995.
- [7] C. Mian, D. Bancher, P. Kohlberger et al., "Fluorescence in situ hybridization in cervical smears: detection of numerical aberrations of chromosomes 7, 3, and X and relationship to HPV infection," *Gynecologic Oncology*, vol. 75, no. 1, pp. 41–46, 1999.

- [8] H. Vrolijk, W. C. R. Sloos, F. M. van de Rijke et al., "Automation of spot counting in interphase cytogenetics using brightfield microscopy," *Cytometry*, vol. 24, no. 2, pp. 158–166, 1996.
- [9] G. Méhes, N. Speich, M. Bollmann, and R. Bollmann, "Chromosomal aberrations accumulate in polyploid cells of High-grade Squamous Intraepithelial Lesions (HSIL)," *Pathology and Oncology Research*, vol. 10, no. 3, pp. 142–148, 2004.
- [10] J. T. Bryan, F. Taddeo, D. Skulsky et al., "Detection of specific human papillomavirus types in paraffin-embedded sections of cervical carcinomas," *Journal of Medical Virology*, vol. 78, no. 1, pp. 117–124, 2006.
- [11] A. H. N. Hopman, W. Theelen, P. P. H. Hommelberg et al., "Genomic integration of oncogenic HPV and gain of the human telomerase gene TERC at 3q26 are strongly associated events in the progression of uterine cervical dysplasia to invasive cancer," *Journal of Pathology*, vol. 210, no. 4, pp. 412–419, 2006.
- [12] E. M. Kanter, E. Vargis, S. Majumder et al., "Application of raman spectroscopy for cervical dysplasia diagnosis," *Journal of Biophotonics*, vol. 2, no. 1-2, pp. 81–90, 2009.
- [13] C. M. Krishna, N. B. Prathima, R. Malini et al., "Raman spectroscopy studies for diagnosis of cancers in human uterine cervix," *Vibrational Spectroscopy*, vol. 41, no. 1, pp. 136–141, 2006.
- [14] S. K. Chang, Y. N. Mirabal, E. N. Atkinson, A. Malpica, M. Follen, and R. Richards-Kortum, "Combination of fluorescence and reflectance spectroscopy for in vivo detection of cervical pre-cancers," in *Proceedings of the IEEE Engineering in Medicine and Biology 24th Annual Conference and the Fall Meeting of the Biomedical Engineering Society (BMES/EMBS '02)*, pp. 2265–2266, Houston, Tex, USA, October 2002.
- [15] Z. Huang, J. Mo, W. Zheng, J. Low, J. Ng, and A. Ilancheran, "Combining near-infrared autofluorescence and raman spectroscopy improves the in vivo detection of cervical precancer," in *Proceedings of the Conference on Quantum Electronics and Laser Science Conference on Lasers and Electro-Optics (CLEO/QELS '08)*, San Jose, Calif, USA, May 2008.
- [16] P. T. T. Wong, R. K. Wong, T. A. Caputo, T. A. Godwin, and B. Rigas, "Infrared spectroscopy of exfoliated human cervical cells: evidence of extensive structural changes during carcinogenesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 24, pp. 10988–10992, 1991.
- [17] M. A. Cohenford, T. A. Godwin, F. Cahn, P. Bhandare, T. A. Caputo, and B. Rigas, "Infrared spectroscopy of normal and abnormal cervical smears: evaluation by principal component analysis," *Gynecologic Oncology*, vol. 66, no. 1, pp. 59–65, 1997.
- [18] M. F. K. Fung, M. Senterman, P. Eid, W. Faught, N. Z. Mikhael, and P. T. T. Wong, "Comparison of fourier-transform infrared spectroscopic screening of exfoliated cervical cells with standard papanicolaou screening," *Gynecologic Oncology*, vol. 66, no. 1, pp. 10–15, 1997.
- [19] L. Chiriboga, P. Xie, H. Yee, D. Zarou, D. Zakim, and M. Diem, "Infrared spectroscopy of human tissue. IV. Detection of dysplastic and neoplastic changes of human cervical tissue via infrared microscopy," *Cellular and Molecular Biology*, vol. 44, no. 1, pp. 219–229, 1998.
- [20] M. Diem, L. Chiriboga, P. Lasch, and A. Pacifico, "IR spectra and IR spectral maps of individual normal and cancerous cells," *Biopolymers—Biospectroscopy Section*, vol. 67, no. 4-5, pp. 349–353, 2002.
- [21] M. J. Romeo, B. R. Wood, M. A. Quinn, and D. McNaughton, "Removal of blood components from cervical smears: implications for cancer diagnosis using FTIR spectroscopy," *Biopolymers—Biospectroscopy Section*, vol. 72, no. 1, pp. 69–76, 2003.
- [22] R. Sindhuphak, S. Issaravanich, V. Udomprasertgul et al., "A new approach for the detection of cervical cancer in Thai women," *Gynecologic Oncology*, vol. 90, no. 1, pp. 10–14, 2003.
- [23] B. R. Wood, L. Chiriboga, H. Yee, M. A. Quinn, D. McNaughton, and M. Diem, "Fourier transform infrared (FTIR) spectral mapping of the cervical transformation zone, and dysplastic squamous epithelium," *Gynecologic Oncology*, vol. 93, no. 1, pp. 59–68, 2004.
- [24] S. G. El-Tawil, R. Adnan, Z. N. Muhamed, and N. H. Othman, "Comparative study between Pap smear cytology and FTIR spectroscopy: a new tool for screening for cervical cancer," *Pathology*, vol. 40, no. 6, pp. 600–603, 2008.
- [25] M. F. Parker, J. P. Karins, and D. M. O'Connor, "Hyperspectral diagnostic imaging of the cervix: initial observations," in *Proceedings of the IEEE Pacific Medical Technology Symposium*, pp. 144–148, Honolulu, Hawaii, USA, August 1998.
- [26] C. Balas, "A novel optical imaging method for the early detection, quantitative grading, and mapping of cancerous and precancerous lesions of cervix," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 1, pp. 96–104, 2001.
- [27] H. Lange, *Reflectance and Fluorescence Hyperspectral Elastic Image Registration*, STI Medical Systems, S.M. Systems, Honolulu, Hawaii, USA, 2007.
- [28] A. Stafl, "Cervicography: a new method for cervical cancer detection," *The American Journal of Obstetrics and Gynecology*, vol. 139, no. 7, pp. 815–825, 1981.
- [29] K. E. Hartmann, K. Nanda, S. Hall, and E. Myers, "Technologic advances for evaluation of cervical cytology: is newer better?" *Obstetrical and Gynecological Survey*, vol. 56, no. 12, pp. 765–774, 2001.
- [30] D. V. Coleman, "Evaluation of automated systems for the primary screening of cervical smears," *Current Diagnostic Pathology*, vol. 5, no. 2, pp. 57–64, 1998.
- [31] J. Karnon, J. Peters, J. Platt, J. Chilcott, E. McGoogan, and N. Brewer, "Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis," *Health Technology Assessment*, vol. 8, no. 20, 2004.
- [32] R. Lozano, "Comparison of computer-assisted and manual screening of cervical cytology," *Gynecologic Oncology*, vol. 104, no. 1, pp. 134–138, 2007.
- [33] D. Schledermann, T. Hyldebrandt, D. Ejersbo, and B. Hoelund, "Automated screening versus manual screening: a comparison of the ThinPrep imaging system and manual screening in a time study," *Diagnostic Cytopathology*, vol. 35, no. 6, pp. 348–352, 2007.
- [34] P. Pudil, F. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proceedings of the IEEE 12th IAPR International Conference on Pattern Recognition: Conference A: Computer Vision and Image Processing*, pp. 279–283, Jerusalem, Israel, 1994.
- [35] P. Pudil, K. Fuka, K. Beranek, and P. Dvorak, "Potential of artificial intelligence based feature selection methods in regression models," in *Proceedings of the IEEE 3rd International Conference on Computational Intelligence and Multimedia Applications*, New Delhi, India, 1999.
- [36] S. J. Reeves and Z. Zhe, "Sequential algorithms for observation selection," *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 123–132, 1999.

- [37] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall, San Francisco, Calif, USA, 6th edition, 2007.
- [38] K.-L. Lin, C.-Y. Lin, C.-D. Huang et al., "Feature selection and combination criteria for improving accuracy in protein structure prediction," *IEEE Transactions on Nanobioscience*, vol. 6, no. 2, pp. 186–196, 2007.
- [39] R. Ashfaq, B. Solares, and M. H. Saboorian, "Detection of endocervical component by PAPNET(TM) system on negative cervical smears," *Diagnostic Cytopathology*, vol. 15, no. 2, pp. 121–123, 1996.
- [40] K. Losell and A. Dejmek, "Comparison of papnet-assisted and manual screening of cervical smears," *Diagnostic Cytopathology*, vol. 21, no. 4, pp. 296–299, 1999.
- [41] D. C. Wilbur, T. A. Bonfiglio, M. A. Rutkowski et al., "Sensitivity of the autoPap 300 QC system for cervical cytologic abnormalities: biopsy data confirmation," *Acta Cytologica*, vol. 40, no. 1, pp. 127–132, 1996.
- [42] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, Addison Wesley, Harlow, UK, 2nd edition, 2005.
- [43] P. Sobrevilla, E. Lerma, and E. Montseny, "An approach to a fuzzy-based automatic pap screening system-FAPSS-addressed to cytology cells detection," in *Proceedings of the IEEE 22nd International Conference of the North American Fuzzy Information Processing Society*, pp. 138–142, Chicago, Ill, USA, July 2003.
- [44] J. Kauppinen and J. Partanen, *Fourier Transforms in Spectroscopy*, Wiley-VCH, Weinheim, Germany, 2001.
- [45] H. Fabian, N. A. N. Thi, M. Eiden, P. Lasch, J. Schmitt, and D. Naumann, "Diagnosing benign and malignant lesions in breast tissue sections by using IR-microspectroscopy," *Biochimica et Biophysica Acta—Biomembranes*, vol. 1758, no. 7, pp. 874–882, 2006.
- [46] F. Bonnier, S. Rubin, L. Ventéo et al., "In-vitro analysis of normal and aneurismal human ascending aortic tissues using FT-IR microspectroscopy," *Biochimica et Biophysica Acta—Biomembranes*, vol. 1758, no. 7, pp. 968–973, 2006.
- [47] K. Das, N. Stone, C. Kendall, C. Fowler, and J. Christie-Brown, "Role of Fourier transform infrared spectroscopy (FTIR) in the diagnosis of parathyroid pathology," *Photodiagnosis and Photodynamic Therapy*, vol. 4, no. 2, pp. 124–129, 2007.
- [48] S. K. Majumder, M. D. Keller, and A. Mahadevan-Jansen, "Optical detection of breast tumors—a comparison of diagnostic performance of autofluorescence, diffuse reflectance, and raman spectroscopy," in *Progress in Biomedical Optics and Imaging*, T. Vo-Dinh, R. Raghavachari, W. S. Grundfest et al., Eds., Proceedings of SPIE, no. 6430, pp. 1–11, Munich, Germany, 2007.
- [49] S. K. Majumder, E. Kanter, A. R. Viehoveer, H. Jones, and A. Mahadevan-Jansen, "Near-infrared raman spectroscopy for in-vivo diagnosis of cervical dysplasia—a probability-based multi-class diagnostic algorithm," in *Progress in Biomedical Optics and Imaging*, T. Vo-Dinh, R. Raghavachari, W. S. Grundfest et al., Eds., Proceedings of SPIE, no. 6430, pp. 1–11, Munich, Germany, 2007.
- [50] S. K. Majumder, M. D. Keller, F. I. Boulos, M. C. Kelley, and A. Mahadevan-Jansen, "Comparison of autofluorescence, diffuse reflectance, and raman spectroscopy for breast tissue discrimination," *Journal of Biomedical Optics*, vol. 13, no. 5, Article ID 054009, 2008.
- [51] N. Baheri, M. Miranbaygi, and R. Malekfar, "Improved skin xerosis detection by combining extracted features from raman spectra," in *Proceedings of the 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL '09)*, Bratislava, Slovakia, November 2009.
- [52] K. Banas, A. Banas, H. O. Moser et al., "Multivariate analysis techniques in the forensics investigation of the postblast residues by means of fourier transform-infrared spectroscopy," *Analytical Chemistry*, vol. 82, no. 7, pp. 3038–3044, 2010.
- [53] C. Palacio, C. Pascual, F. Suarez, and I. Lloret, "Smoothing of digital spectroscopic data by using a Chebyshev filter," *Vacuum*, vol. 64, no. 3–4, pp. 481–485, 2002.
- [54] C. Battistoni, S. Kaciulis, G. Mattogno, and G. Righini, "Noise removal from Auger images by using adaptive binomial filter," *Journal of Electron Spectroscopy and Related Phenomena*, vol. 76, no. C, pp. 399–404, 1995.
- [55] C. Battistoni, G. Mattogno, and G. Righini, "Spectral noise removal by new digital smoothing routine," *Journal of Electron Spectroscopy and Related Phenomena*, vol. 74, no. 2, pp. 159–166, 1995.
- [56] J. S.-J. Lee, J.-N. Hwang, D. T. Davis, and A. C. Nelson, "Integration of neural networks and decision tree classifiers for automated cytology screening," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 257–262, Singapore, July 1991.
- [57] Part, P. Malm, and A. Brun, "Closing curves with Riemannian dilation: application to segmentation in automated cervical cancer screening," in *Advances in Visual Computing*, vol. 5875 of *Lecture Notes in Computer Science*, pp. 337–346, Springer, Berlin, Germany, 2009.
- [58] M. E. Plissiti, C. Nikou, and A. Charchanti, "Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images," *Pattern Recognition Letters*, vol. 32, no. 6, pp. 838–853, 2011.
- [59] H. Lange, *Automatic Glare Removal in Reflectance Imagery of the Uterine Cervix*, STI Medical Systems, S.M. Systems, Honolulu, Hawaii, USA, 2005.
- [60] X. Wang, B. Zheng, S. Li et al., "Automated detection and analysis of fluorescent in situ hybridization spots depicted in digital microscopic images of Pap-smear specimens," *Journal of Biomedical Optics*, vol. 14, no. 2, Article ID 021002, 2009.
- [61] W. Li, J. Gu, D. Ferris, and A. Poirson, *Automated Image Analysis of Uterine Cervical Images*, STI Medical Systems, S.M. Systems, Honolulu, Hawaii, USA, 2007.
- [62] B. Sokouti, S. Haghipour, and A. D. Tabrizi, "A pilot study on image analysis techniques for extracting early uterine cervix cancer cell features," *Journal of Medical Systems*, pp. 1–7, 2011.
- [63] H. Greenspan, S. Gordon, G. Zimmerman et al., "Automatic detection of anatomical landmarks in uterine cervix images," *IEEE Transactions on Medical Imaging*, vol. 28, no. 3, pp. 454–468, 2009.
- [64] C.-H. Lin, Y.-K. Chan, and C.-C. Chen, "Detection and segmentation of cervical cell cytoplasm and nucleus," *International Journal of Imaging Systems and Technology*, vol. 19, no. 3, pp. 260–270, 2009.
- [65] C. Demir and B. Yener, *Automated Cancer Diagnosis Based on Histopathological Images: A Systematic Survey*, Rensselaer Polytechnic Institute, New York, NY, USA, 2005.
- [66] S. Lotenberg, S. Gordon, and H. Greenspan, "Shape priors for segmentation of the cervix region within uterine cervix images," *Journal of Digital Imaging*, vol. 22, no. 3, pp. 286–296, 2009.
- [67] V. van Raad, *Image Analysis and Segmentation of Anatomical Features of Cervix Uteri in Color Space*, STI Medical Systems, S.M. Systems, Honolulu, Hawaii, USA, 2005.

- [68] V. van Raad, Z. Xue, and H. Lange, *Lesion Margin Analysis for Automated Classification of Cervical Cancer Lesions*, STI Medical Systems, S.M. Systems, Honolulu, Hawaii, USA, 2006.
- [69] P. Sobrevilla, E. Montseny, F. Vaschetto, and E. Lerma, "Fuzzy-based analysis of microscopic color cervical pap smear images: nuclei detection," *International Journal of Computational Intelligence and Applications*, vol. 9, no. 3, pp. 187–206, 2010.
- [70] S. N. Sulaiman, N. A. Mat Isa, and N. H. Othman, "Semi-automated pseudo colour features extraction technique for cervical cancer's Pap smear images," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol. 15, no. 3, pp. 131–143, 2011.
- [71] B. Tulpule, S. Yang, Y. Srinivasan, S. Mitra, and B. Nutter, "Segmentation and classification of cervix lesions by pattern and texture analysis," in *Proceedings of the 14th IEEE International Conference on Fuzzy Systems (FUZZ '05)*, pp. 173–176, Reno, Nev, USA, May 2005.
- [72] Z. Li and K. Najarian, "Automated classification of Pap smear tests using neural networks," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '01)*, pp. 2899–2901, Washington, DC, USA, July 2001.
- [73] N. A. Mat-Isa, M. Y. Mashor, and N. H. Othman, "An automated cervical pre-cancerous diagnostic system," *Artificial Intelligence in Medicine*, vol. 42, no. 1, pp. 1–11, 2008.
- [74] K. Zhang, L. Zhang, H. Song, and W. Zhou, "Active contours with selective local or global segmentation: a new formulation and level set method," *Image and Vision Computing*, vol. 28, no. 4, pp. 668–676, 2010.
- [75] M. E. Gómez-Mayorga, F. J. Gallegos-Funes, J. M. de-la-Rosa-Vázquez, R. Cruz-Santiago, and V. Ponomaryov, "Diagnosis of cervical cancer using the median M-type radial basis function (MMRBF) neural network," in *Proceedings of the 8th Mexican International Conference on Artificial Intelligence*, pp. 258–267, Guanajuato, México, 2009.
- [76] H. Netten, L. J. van Vliet, H. Vrolijk, W. C. R. Sloos, H. J. Tanke, and I. T. Young, "Fluorescent dot counting in interphase cell nuclei," *Bioimaging*, vol. 4, no. 2, pp. 93–106, 1996.
- [77] M. Gué, C. Messaoudi, J. S. Sun, and T. Boudier, "Smart 3D-FISH: automation of distance analysis in nuclei of interphase cells by image processing," *Cytometry A*, vol. 67, no. 1, pp. 18–26, 2005.
- [78] Y. Jusman, N. A. Mat Isa, R. Adnan, and N. H. Othman, "Intelligent classification of cervical pre-cancerous cells based on the FTIR spectra," *Ain Shams Engineering Journal*, vol. 3, no. 1, pp. 61–70, 2012.
- [79] Z. Xue, S. Antani, L. R. Long, and G. R. Thoma, "An online segmentation tool for cervicographic image analysis," in *Proceedings of the 1st ACM International Health Informatics Symposium (IHI '10)*, pp. 425–429, Arlington, Va, USA, November 2010.
- [80] I. Claude, R. Winzenrieth, P. Pouletaut, and J. C. Boulanger, "Contour features for colposcopic image classification by artificial neural networks," in *Proceedings of IEEE 16th International Conference on Pattern Recognition*, pp. 771–774, Quebec, Canada, August 2002.
- [81] B. Kajtár, G. Méhes, T. Lörch et al., "Automated fluorescent in situ hybridization (FISH) analysis of t(9;22)(q34;q11) in interphase nuclei," *Cytometry A*, vol. 69, no. 6, pp. 506–514, 2006.
- [82] S. Mark, R. K. Sahu, K. Kantarovich et al., "Fourier transform infrared microspectroscopy as a quantitative diagnostic tool for assignment of premalignancy grading in cervical neoplasia," *Journal of Biomedical Optics*, vol. 9, no. 3, pp. 558–567, 2004.
- [83] J. Suryatenggara, B. K. Ane, M. Pandjaitan, and W. Steinberg, "Pattern recognition on 2D cervical cytological digital images for early detection of cervix cancer," in *Proceedings of the World Congress on Nature and Biologically Inspired Computing (NABIC '09)*, pp. 257–262, Coimbatore, India, December 2009.
- [84] K. Turner, N. Ramanujam, J. Ghosh, and R. Richards-Kortum, "Ensembles of radial basis function networks for spectroscopic detection of cervical precancer," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 8, pp. 953–961, 1998.
- [85] S. J. Keenan, J. Diamond, W. Glenn McCluggage et al., "An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN)," *Journal of Pathology*, vol. 192, no. 3, pp. 351–362, 2000.
- [86] Y. Srinivasan, E. Corona, B. Nutter, S. Mitra, and S. Bhattacharya, "A unified model-based image analysis framework for automated detection of precancerous lesions in digitized uterine cervix images," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 101–111, 2009.
- [87] Z. Xue, L. R. Long, S. Antani, and G. R. Thoma, "Automatic extraction of mosaic patterns in uterine cervix images," in *Proceedings of the 23rd IEEE International Symposium on Computer-Based Medical Systems (CBMS '10)*, pp. 273–278, Perth, Australia, October 2010.
- [88] S. Zhang, J. Huang, D. Metaxas, W. Wang, and X. Huang, "Discriminative sparse representations for cervigram image segmentation," in *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging: from Nano to Macro (ISBI '10)*, pp. 133–136, Rotterdam, The Netherlands, April 2010.
- [89] V. van Raad, "Design of Gabor wavelets for analysis of texture features in cervical images," in *Proceedings of the 25th Annual International Conference of the IEEE*, vol. 801, pp. 806–809, Engineering in Medicine and Biology Society, 2003.
- [90] W. Li and A. Poirson, *Detection and Characterization of Abnormal Vascular Patterns in Automated Cervical Image Analysis*, STI Medical Systems, S.M. Systems, Honolulu, Hawaii, USA, 2006.
- [91] J. D. García-Arteaga and J. Kybic, *Geometric and Information Constraints for Automatic Landmark Selection in Colposcopy Sequences*, STI Medical Systems, S.M. Systems, Honolulu, Hawaii, USA, 2007.
- [92] K. Krishnaveni, S. Allwin, S. P. K. Kenny, and G. Mariappan, "Analysis for textural features in nuclei of cervical cyto images," in *Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICIC '10)*, pp. 943–947, Coimbatore, India, December 2010.
- [93] A. Alush, H. Greenspan, and J. Goldberger, "Lesion detection and segmentation in uterine cervix images using an ARC-level mRF," in *Proceedings of the IEEE International Symposium on Biomedical Imaging: from Nano to Macro (ISBI '09)*, pp. 474–477, Boston, Mass, USA, July 2009.
- [94] A. Alush, H. Greenspan, and J. Goldberger, "Automated and interactive lesion detection and segmentation in uterine cervix images," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 488–501, 2010.
- [95] X. Huang, W. Wang, Z. Xue, S. Antani, L. R. Long, and J. Jeronimo, "Tissue classification using cluster features for lesion detection in digital cervigrams," in *Medical Imaging 2008: Image Processing*, Proceedings of SPIE, no. 6914, pp. 69141Z–69141Z, San Diego, Calif, USA, 2008.
- [96] W. Li, S. Venkataraman, U. Gustafsson, J. C. Oyama, D. G. Ferris, and R. W. Lieberman, "Using acetowhite opacity

- index for detecting cervical intraepithelial neoplasia," *Journal of Biomedical Optics*, vol. 14, no. 1, Article ID 014020, 2009.
- [97] S. Y. Park, D. Sargent, R. Lieberman, and U. Gustafsson, "Domain-specific image analysis for cervical neoplasia detection based on conditional random fields," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 867–878, 2011.
- [98] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 6, pp. 1517–1525, 2004.
- [99] A. Yildiz, M. Akin, and M. Poyraz, "An expert system for automated recognition of patients with obstructive sleep apnea using electrocardiogram recordings," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12880–12890, 2011.
- [100] J. C. B. Melo, G. D. C. Cavalcanti, and K. S. Guimarães, "PCA feature extraction for protein structure prediction," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 2952–2957, Portland, Ore, USA, July 2003.
- [101] K. Polat, S. Güneş, and A. Arslan, "A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square support vector machine," *Expert Systems with Applications*, vol. 34, no. 1, pp. 482–487, 2008.
- [102] E. Dogantekin, A. Dogantekin, and D. Avci, "Automatic hepatitis diagnosis system based on linear discriminant analysis and adaptive network based on fuzzy inference system," *Expert Systems with Applications*, vol. 36, no. 8, pp. 11282–11286, 2009.
- [103] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimization for support vector machines," *Expert Systems with Applications*, vol. 31, no. 2, pp. 231–240, 2006.
- [104] C.-L. Huang and J.-F. Dun, "A distributed PSO-SVM hybrid system with feature selection and parameter optimization," *Applied Soft Computing Journal*, vol. 8, no. 4, pp. 1381–1391, 2008.
- [105] L. A. Reisner, A. Cao, and A. K. Pandya, "An integrated software system for processing, analyzing, and classifying raman spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 105, no. 1, pp. 83–90, 2011.
- [106] G. M. Palmer, C. Zhu, T. M. Breslin, F. Xu, K. W. Gilchrist, and N. Ramanujam, "Comparison of multiexcitation fluorescence and diffuse reflectance spectroscopy for the diagnosis of breast cancer," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 11, pp. 1233–1242, 2003.
- [107] S. J. Baek, A. Park, J. Y. Kim, S. Y. Na, Y. Won, and J. Choo, "Detection of basal cell carcinoma by automatic classification of confocal raman spectra," in *Computational Intelligence and Bioinformatics*, vol. 4115 of *Lecture Notes in Computer Science*, pp. 402–411, Springer, Berlin, Germany, 2006.
- [108] F. M. Lyng, E. Ó. Faoláin, J. Conroy et al., "Vibrational spectroscopy for cervical cancer pathology, from biochemical analysis to diagnostic tool," *Experimental and Molecular Pathology*, vol. 82, no. 2, pp. 121–129, 2007.
- [109] C.-D. Huang, C.-T. Lin, and N. R. Pal, "Hierarchical learning architecture with automatic feature selection for multiclass protein fold classification," *IEEE Transactions on Nanobioscience*, vol. 2, no. 4, pp. 221–232, 2003.
- [110] W. Lin, X. Yuan, P. Yuen et al., "Classification of in vivo autofluorescence spectra using support vector machines," *Journal of Biomedical Optics*, vol. 9, no. 1, pp. 180–186, 2004.
- [111] V.-E. Neagoe, A.-C. Mugioiu, and I.-A. Stanculescu, "Face recognition using PCA versus ICA versus LDA cascaded with the neural classifier of concurrent self-organizing maps," in *Proceedings of the 8th International Conference on Communications (COMM '10)*, pp. 225–228, Bucharest, Romania, June 2010.
- [112] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [113] P. R. T. Jess, D. D. W. Smith, M. Mazilu, K. Dholakia, A. C. Riches, and C. S. Herrington, "Early detection of cervical neoplasia by raman spectroscopy," *International Journal of Cancer*, vol. 121, no. 12, pp. 2723–2728, 2007.
- [114] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification," in *Proceedings of the IEEE International Conference on Multimedia and Expo, ICME 2005*, pp. 940–943, Amsterdam, The Netherlands, July 2005.
- [115] G. Karypis and E. Han, "Concept indexing: a fast dimensionality reduction algorithm with applications to document retrieval and categorization," DTIC Document, U.o. Minnesota, Minneapolis, Minn, USA, 2000.
- [116] E. Njoroge, S. R. Alty, M. R. Gani, and M. Alkatib, "Classification of cervical cancer cells using FTIR data," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '06)*, pp. 5338–5341, New York, NY, USA, September 2006.
- [117] V. Margariti, M. Zervakis, and C. Balas, "Wavelet and physical parametric analysis of the acetowhitening optical effect: comparative evaluation of performances in non-invasive diagnosis of cervical neoplasia," in *Proceedings of the 10th International Conference on Information Technology and Applications in Biomedicine: Emerging Technologies for Patient Specific Healthcare (ITAB '10)*, Corfu, Greece, November 2010.
- [118] Y. N. Mirabal, S. K. Chang, E. N. Atkinson, A. Malpica, M. Follen, and R. Richards-Kortum, "Reflectance spectroscopy for in vivo detection of cervical precancer," *Journal of Biomedical Optics*, vol. 7, no. 4, pp. 587–594, 2002.
- [119] J. Mo, W. Zheng, J. J. H. Low, J. Ng, A. Ilancheran, and Z. Huang, "High wavenumber raman spectroscopy for in vivo detection of cervical dysplasia," *Analytical Chemistry*, vol. 81, no. 21, pp. 8908–8915, 2009.
- [120] R. F. Walker, P. Jackway, B. Lovell, and I. D. Longstaff, "Classification of cervical cell nuclei using morphological segmentation and textural feature extraction," in *Proceedings of the 2nd Australian and New Zealand Conference on Intelligent Information Systems*, pp. 297–301, December 1994.
- [121] S. Y. Park, M. Follen, A. Milbourne et al., "Automated image analysis of digital colposcopy for the detection of cervical neoplasia," *Journal of Biomedical Optics*, vol. 13, no. 1, Article ID 014029, 2008.
- [122] D. Westreich, J. Lessler, and M. J. Funk, "Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression," *Journal of Clinical Epidemiology*, vol. 63, no. 8, pp. 826–833, 2010.
- [123] J.-P. Zhang, Z.-W. Li, and J. Yang, "A parallel SVM training algorithm on large-scale classification problems," in *Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC '05)*, pp. 1637–1641, Guangzhou, China, August 2005.
- [124] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1048–1054, 1999.
- [125] C. Sivapragasam and N. Muttill, "Discharge rating curve extension—a new approach," *Water Resources Management*, vol. 19, no. 5, pp. 505–520, 2005.

