

Quantitative histopathological analysis of cervical intra-epithelial neoplasia sections: Methodological issues

Martial Guillaud^a, Dennis Cox^b, Anais Malpica^c, Gregg Staerke^c, Jasenka Maticic^d, Dirk Van Niekirk^d, Karen Adler-Storthz^e, Neal Poulin^a, Michele Follen^{f,g,*} and Calum MacAulay^a

^a Dept. of Cancer Imaging, BC Cancer Agency, Vancouver, BC, Canada

^b Dept. of Statistics, Rice University, Houston, TX, USA

^c Dept. of Pathology, UT M.D. Anderson Cancer Center, Houston, TX, USA

^d Dept. of Pathology, BC Cancer Agency, Vancouver, BC, Canada

^e Dept. of Research Dentistry, University of Texas Health Science Center at Houston, Houston, TX, USA

^f Dept. of Gynecologic Oncology, Center for Biomedical Engineering, UT M.D. Anderson Cancer Center, Houston, TX, USA

^g Dept. of Obstetrics, Gynecology and Reproductive Sciences, University of Texas Health Science Center at Houston, Houston, TX, USA

Received 31 March 2003

Accepted 21 October 2003

Abstract. *Objectives:* As part a Program Project to evaluate emerging optical technologies for cervical neoplasia, our group is performing quantitative histopathological analysis of biopsies from 1800 patients. Several methodological issues have arisen with respect to this analysis: (1) Finding the most efficient way to compensate for staining intensity variation with out losing diagnostic information; (2) Assessing the inter- and intra-observer variability of the semi-interactive data collection; and (3) the use of non-overlapping cells from the intermediate layer only.

Methods: Non-overlapping quantitatively stained nuclei were selected from 280 samples with histopathological characteristics of normal (199), koilocytosis (37), CIN 1 (18), CIN 2 (10) and CIN 3 (16). Linear discriminant analysis was used to assess the diagnostic information in three different feature sets to evaluate and compare staining intensity normalization methods. Selected feature values and summary scores were used to evaluate intra- and inter-observer variability.

Results: The features normalized by the internal subset of the imaged cells had the same discriminatory power as those normalized by the control cells and by both normalization methods seem to have additional discriminatory power over the set of features which do not require normalization. The use of the internal subset decreased the image acquisition time by ~50% at each center, respectively. The intra- and inter-observer variability was of a similar size. Good performance was obtained by measuring the intermediate layer only.

Conclusion: The use of intensity normalization from a subset of the imaged non-overlapping intermediate layer cells works as well as or better than any of the other methods tested and provides a significant timesaving. Our intra- and inter-observer variability do not seem to affect the diagnostic power of the data. Although this must be tested in a larger data set, the use of intermediate layer cells only may be acceptable when using quantitative histopathology.

Keywords: Semi-interactive quantitative histopathology, stoichiometric staining, stain intensity normalization, inter- and intra-observer variability, CIN, SIL, morphology

*Corresponding author: Michele Follen, MD, PhD, University of Texas M.D. Anderson Cancer Center, Center for Biomedical Engineering, Unit 193, 1515 Holcombe Blvd., Houston, TX 77030, USA.

Tel.: +1 713 745 4780; Fax: +1 713 792 4856; E-mail: mfolle@mdanderson.org.

1. Introduction

Cervical cancer is the third most common cancer in women worldwide and the leading cause of cancer mortality for women in developing countries [17]. Mortality has decreased in every country in which a screening program has been instituted [16]. Our group has decided to focus our research on optical technologies for the screening and detection of cervical neoplasia. In our Program Project, we are evaluating the biological plausibility, technical feasibility, clinical effectiveness, patient satisfaction, and cost effectiveness of quantitative cytology, quantitative histopathology, localized fluorescence and reflectance spectroscopy, and whole organ multispectral digital colposcopy [5]. In these studies our group is performing quantitative histopathological analysis of biopsies from 1800 patients. The purpose of this report is to address several methodological issues that have arisen as part of this effort [2,10,20].

The principal methodological issues that have arisen for quantitative histopathology are: (1) many features need to be normalized to eliminate the effects of slide-to-slide variability in the stoichiometric staining. Finding the most efficient way to compensate for this stain intensity variation without losing diagnostic information is critical. The current practice in cytology utilizes a control cell population such as lymphocytes, chicken erythrocytes, trout red blood cells, etc. In sectioned material, lymphocytes or other cell populations not involved in the neoplastic process are used. We compare the use of lymphocyte normalization, normalization with a subset of the epithelial cells, and the use of only stain intensity independent features. (2) Since sectioned material cannot currently be fully automatically segmented, it is necessary for the technologist to select non-overlapping nuclei and segment adjacent nuclei. We give an interim analysis of the inter- and intra-observer variability of the semi-interactive data collection. A more comprehensive analysis will be reported at the end of the data collection for both clinical trials. (3) Even using semi-interactive data collection, the process of selecting non-overlapping nuclei from the crowded area of the basal membrane is very labor intensive. Also, we have found that 10% of the samples are missing the layer of superficial cells. Thus, we have considered using nuclei from the intermediate layer only.

2. Materials and methods

2.1. Materials

The study described was conducted at the University of Texas M.D. Anderson Cancer Center (Houston, TX) and at the British Columbia Cancer Agency. Women age eighteen and older were enrolled in the study and an informed consent was obtained from each subject who participated. The protocols were approved by the Internal Review Boards at M.D. Anderson Cancer Center and BC Cancer Agency/UBC.

In the screening study, patients enter with a history of no abnormal Pap smears. If their colposcopy is normal, biopsies of two normal areas are obtained. If their colposcopy is abnormal, they are treated like a patient in the diagnostic study. In the diagnostic study, patients are referred with an abnormal Pap smear and usually have lesions on their colposcopic exam. In this case, a biopsy from the abnormal area as well as a biopsy from a normal area of the cervix is taken. The biopsies are fixed in buffered formalin and embedded in paraffin blocks. Each patient has specimens collected for hybrid capture (Gaithersburg, Maryland) and for quantitative PCR for HPV DNA and RNA for 16, 18 and other high risk types. Three adjacent sections are cut at 4 μm and stained with Haematoxylin and Eosin (H&E). These sections are used for conventional readings. The adjacent sections are stained with the Thionin–Feulgen staining procedure. This stain has been shown to be stoichiometric and related to the biochemical measurement of DNA content.

2.2. Pathology review

The first pathology review is done by one of the gynecological pathologists on clinical duty. A second blinded review is performed by one of our study pathologists (A.M., G.S., J.M., D.V.N.). In cases of discrepancies between the two readings, the slide is read a third time by a study pathologist to provide the final consensus diagnosis. During the second review, the pathologist stores a digital picture of the diagnostic area in the study database. As part of the Program Project, all of the slides at the U.T. M.D. Anderson site will be read in Vancouver, and all of the slides at the Vancouver site will be read at U.T. M.D. Anderson. Additionally, two experienced gynecologic pathologists who are considered experts in cervical pathology have agreed to review all the slides. Inter- and intra-observer clinical data will be published.

2.2.1. Samples

A total of 1420 quantitatively stained sections have been analyzed with the imaging system. Only samples containing squamous epithelium are included here. Of these, a final histopathological interpretation is complete on 280 samples, which is the set of samples used for the studies reported here. These include the following diagnostic groups: (1) normal, 199 samples; (2) koilocytosis, 37 samples; (3) CIN 1, 18 samples, (4) CIN 2, 10 samples, and (5) CIN 3, 16 samples.

Each sample was mapped for image analysis. The basal layer is the one adjacent layer to the basement membrane and is characterized by the presence of oval shaped nuclei. The parabasal layer is defined as the groups of cells adjacent to the basal layer and consists of oval nuclei. The intermediate layer has polyhedral shaped nuclei and has more cytoplasm in each cell. As the cells move toward the surface and the cells become round and small. The superficial layer has small nuclei

and flat and lose cytoplasm; it is located at the very top of the epithelium.

2.3. Image analysis

Image analysis was performed using an in-house modified version of the Cyto-Savant automated quantitative system (Vancouver, BC, Canada). This system includes a 12 bit double correlated sampling MicroImager 1400 digital camera (pixels $6.8 \mu\text{m}^2$). This software was specially designed for semi-automatic analysis tissue sections. Thionin–Feulgen stained nuclei were measured with a monochromatic light at a wavelength of 600 nm, using a 20×0.75 NA Plan APO objective lens. With a printout of the diagnostic area on hand, a cyto-technologist locates the exact same area on the Feulgen-stained slide as on the H&E slide. As illustrated in Fig. 1, the cytotechnologist outlines with a mouse the basal membrane and the superficial

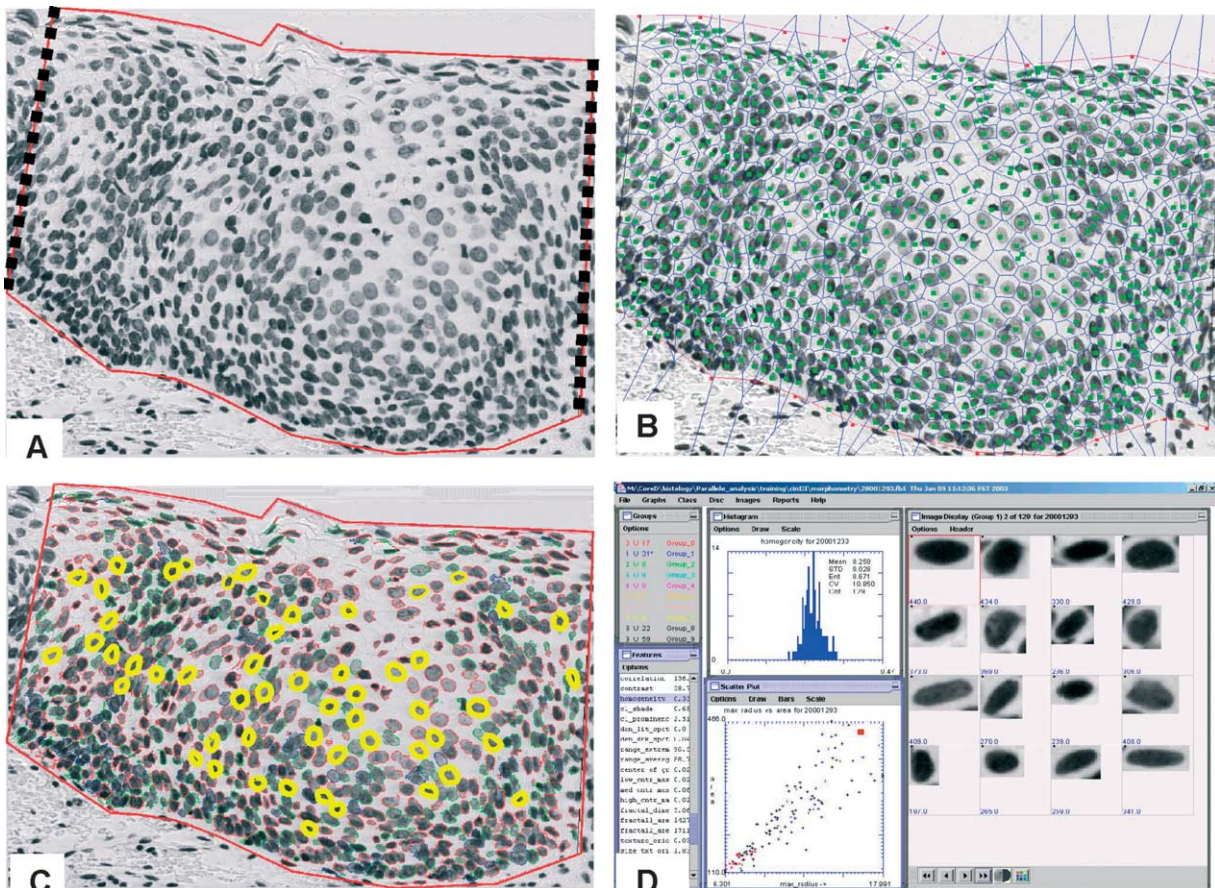


Fig. 1. Different steps of the semi-automated analysis of cervical lesions. A: Definition of the Region of Interest. B: Automatic locations of the positions of the nuclei. C: Automatic segmentation of the nuclei in the intermediate layers of the epithelium. D: Snapshot of the interface of our software used for post-analysis of the morphometric analyses (Quality Control).

membrane (Fig. 1A). These two membranes define the Region of Interest (ROI), or Sampling Window. Automatic detection of the nuclei (Fig. 1B) is then performed for further architectural analyses. This procedure is fully automated and requires only some minor manual changes. At a high magnification (20 \times), the nuclear segmentation is performed within the ROI (Fig. 1C). A nuclear segmentation algorithm has been described in detail elsewhere [12]. Briefly, a thresholding algorithm is used to separate the objects (nuclei) from the background, based on pixel intensity. A manual correction of the nuclear segmentation is made for touching objects. Auto-focusing and edge-relocation algorithms were finally applied to the nuclei to precisely and automatically place the edge of the object at the region of highest local gray level gradient [12]. The digital gray-level images of these nuclei were stored in a gallery (Fig. 1D).

2.3.1. Lymphocyte collection

Between 10 and 100 lymphocyte nuclei are collected from the underlying stromal compartment in the vicinity of the region of interest or in the adjacent fields if required. The same steps described above in the image analysis section for the epithelial nuclei are applied to the lymphocyte collection.

2.3.2. Quality control

The cytotechnologist manually reviews each object in an image gallery of all the selected cells (Fig. 1D)

and removes any object which does not fulfill the minimum requirements (bad mask, out of focus, pale nucleus, pycnotic nucleus, etc.). Special attention is given to the lymphocytes in order to obtain a homogeneous population: only dark, dense, round objects are accepted.

2.3.3. Feature calculation

Nuclear features are extracted from the digitized nuclear images of each selected cell. Table 1 gives the list of the features organized into different categories; approximately 120 features are calculated [4]. Morphological features describe the nuclear size, shape, and boundary irregularities. The eight photometric features estimate the absolute intensity, optical density levels of the nucleus, and the intensity distribution characteristics. DNA amount is the raw measurement of the Integrated Optical Density (IOD) from which all the photometric features are derived. The IOD norm is the mean value of the DNA amount of the reference population. The DNA Index is the normalized measure of the integrated optical density of the object, i.e. DNA amount divided by IOD norm.

2.3.4. Texture features

Discrete texture features are based on thresholded segmentation of the object into regions of low, medium, and high optical density. The thresholds are scaled to the sample staining intensity as represented by the IOD

Table 1
List of features used in this study per categories (see text)

Category	Features
<i>Cytometric Features</i>	
Morphometry (42)	
Size (3)	Area, mean_radius, variance_radius
Shape (5)	Eccentricity, sphericity, elongation, compactness, inertia_shape
Boundaries (34)	Low_freq_fft, Freq_low_fft, Harm01-32_fft
Photometric (5)	DNA_index, OD_max, OD_var, OD_skew, OD_kurt
Discrete texture (24)	Low, medium, and high-DNA_amount Low, medium, and high-DNA_area Low, medium, high, and medium_high-DNA compactness Low, medium, high, and medium_high-DNA average distance Low, medium, medium-high and high, density object Low, medium, and high centre mass Low_vs_medium, low_vs_high, low_vs_medium_high DNA
Markovian texture (7)	Entropy, energy, contrast, correlation, Homogeneity, cl_shade, cl_prominence
Non-Markovian texture (5)	Density_light_spots, density_dark_spots, center_of_gravity
Fractal texture (3)	Fractal_Area1, Fractal_Area2, Fractal_dimension
Run length texture (10)	Short_runs_mean, Short_run_stdv Long_runs_mean, Long_run_stdv Gray_level_mean, Gray_level_stdv Run_length_mean, Run_length_stdv Run_percent_mean, Run_percent_stdv

norm value determined from the reference population. Details of algorithms are described elsewhere [4]. Discrete texture features are by definition dependent on the normalization. Markovian texture features characterize gray level correlation between adjacent pixels in the image. Non-Markovian texture features describe the texture in terms of local maxima and minima of gray level differences in the object. Fractal texture features describe the texture using local differences integrated over the object at multiple resolutions. Run-Length Texture features describe chromatin distribution in terms of the length of consecutive pixels with the same compressed gray level value along different orientations (0° , 45° , 90° , 135°). In order to make the run length features rotationally invariant, for each run length feature we only use the mean and standard deviation over the four directions.

For the purpose of this interim analysis, three sets of features were created. Feature Set 1 is composed of all the features that do not require normalization by a reference population, i.e. all features except the DNA Index and the Discrete Texture features. Feature Set 2 consists of all features and the mode of the DNA amount distribution of the lymphocytes was manually determined and used as IOD norm. Feature Set 3 also includes all the features, but a different method of normalization is used. This method, called “epithelium normalization”, utilizes the mean of the DNA amount of the epithelial cells as the IOD norm.

2.4. Statistical analyses

Linear discriminant analysis was used to assess the diagnostic information in the three different feature sets on a cell-by-cell and sample-by-sample basis. For the cell-by-cell analysis, three diagnostic comparisons were made: normal vs. CIN 3, normal vs. koilocytosis, and koilocytosis vs. CIN 1. All three feature sets were included in the cell-by-cell comparisons. Because of the small numbers of samples with high-grade abnormalities, the analysis on the sample-by-sample basis was different. For the sample-by-sample analysis, three diagnostic comparisons were made: normal vs. abnormal [CIN 1, 2, 3], normal vs. koilocytosis, and koilocytosis vs. abnormal [CIN 1, 2, 3]. Only Feature Sets 2 and 3 were included in the sample-by-sample analysis. All discriminant analyses were performed using the same F -to-enter and F -to-remove value of 20 for a forward stepwise analysis. The main focus of this study is to compare the relative discriminant power of the different feature sets. We performed the analysis in this

fashion rather than pre-selecting the number of features to be included in the discriminant function. F -values were chosen in order to avoid over-training, especially for the sample-by-sample analyses, in which the initial number of features is about the same as the number of cases. All statistical analyses were performed with the STATISTICA™ package produced by StatSoft, Inc., Tulsa, OK.

Selected feature values and summary scores were used to evaluate intra- and inter-observer variability. Visually apparent feature values [area, shape, DNA amount] and the discriminant scores from the cell by cell comparisons were compared.

3. Results

3.1. Stain intensity compensation

3.1.1. Cell-by-cell analysis

To compare the discriminant power of the three sets of features, we created 5 groups of cells by sampling cells from the five different pathology grades. The normal group was composed of an average of 10 cells from each of the 199 normal biopsies (sampled randomly), for a total of 1773 cells. The Koilocytosis cell group was composed of an average of 30 cells randomly sampled from each of the 37 koilocytotic biopsies, for a total of 1224 cells. The CIN 1 (497 cells), CIN 2 (654 cells) and CIN 3 group (933 cells) were composed of all the cells from the respective CIN 1, CIN 2 and CIN 3 biopsies (exhaustive). These different sampling methods provide us with groups of a similar size.

As seen in Table 2A for Feature Set 1, the stepwise-forward discriminant analysis selected Max_Radius, Fractal_Area1, Gray_Level_Mean, Harm_fft3 and Energy as the variables contributing most to the discrimination of the two groups. The percentages of cells correctly classified as normal and CIN 3 were 92% and 74%, respectively. Similarly, using the Feature Set 2¹, the correct classification of cells in the normal group and in the CIN 3 group were 93% and 83%, respectively. 95% of cells from the normal group were correctly classified and 87% of cells from the CIN 3 group were correctly classified using the Feature Set 3²,

¹ Selected features: Sphericity, Harm_fft3, Cl_shade, Energy, DNA_Index, Low_Den_obj, Med_Av_Dst, Low_Den_Area.

² Sphericity, Harm_fft3, Harm_fft4, Fractal_Area2, Run_length_Mean, High_DNA_Area, Low_vs_Med_Area, Low_Avg_Distance, MhDNA_Comp.

Table 2A

Cells-by-cells analysis: Percentage of correct classification after the stepwise discriminant analysis using the three features sets. F -to-enter, F -to-removed: 20

Pathology	Leukocytes aim 1		
	Feature Set 1 (no normalization required)	Feature Set 2 (leukocyte normalization)	Feature Set 3 (epithelial normalization)
Normal	92%	93%	95%
CIN 2	74%	83%	87%
Total	78%	87.5%	91%

Table 2B

Classification matrix cells-by-cells analysis: Results of the stepwise discriminant analysis of negative and HPVAC lesions using the three features sets. F -to-enter, F -to-removed: 20

Pathology	Feature Set 1 (no normalization required)	Feature Set 2 (leukocyte normalization)	Feature Set 3 (epithelial normalization)
Normal	66%	68%	75%
Koilocytosis	66%	69%	69%
Total	66%	68.5%	72%

Table 2C

Cells-by-cells analysis: Results of the stepwise discriminant analysis of koilocytosis and CIN 1 lesions using the three features sets. F -to-enter, F -to-removed: 20

Pathology	Feature Set 1 (no normalization required)	Feature Set 2 (leukocyte normalization)	Feature Set 3 (epithelial normalization)
HPVAc	67%	70%	72%
CIN 1	61%	57%	61%
Total	64%	63.5%	66.5%

which is slightly higher than the percentage obtained with Feature Set 2. Adding the discrete texture features (as seen in Table 2A Feature Sets 2 and 3) significantly improves the discrimination between these two groups (about 10% improvement), independently of the normalization method. As seen in Tables 2B and 2C, normalization methodology does not change the discriminant ability of the feature sets^{3,4} as we consider other important histopathological diagnosis in the cervix.

3.1.2. Morphometric scores

Using the features selected during the cell-by-cell discriminant analysis of the normal group versus the CIN 3 group, we calculated two morphometric scores for each biopsy, one derived from the analysis with

³Selected Features, Table 2B, Set 1: Max_Radius, Fract_Dimension, Energy, Contrast; Set 2: Max_Radius, Contrast, Long_Run, Fract_Dimension, Medium_Av_Distance, Set 3: Max_radius, Energy, Fractal_Dimension, Medium_DNA_Amount, Medium_DNA_Compactness, Low_DNA_Compactness, High_DNA_Density_Object, Mh_DNA_Compactness.

⁴Selected Features, Table 2C, Set 1: Fractal_Area1, Correlation, Short_run, Cl_shade; Set 2: Fractal_Area1, Short_Run, DNA_Index, Set 3: Fractal_Area1, Low_DNA_Compactness, DNA_Index.

Feature Set 2 and one derived from the analysis with Feature Set 3. Mean and standard errors of these scores are plotted for each pathology grade (Fig. 2). As expected from the previous results, the two curves show very similar trends. The score derived from Features Set 3 (the new normalization method) gives a better separation between the CIN 2 group and the CIN 3 group than does the score derived from Feature Set 2 (classical lymphocyte normalization).

3.1.3. Sample-by-sample analysis

In each specimen, for each feature, the mean and the standard deviation were calculated. Results of the stepwise discriminant analysis of normal, koilocytosis and abnormal lesions (CIN 1, CIN 2 and CIN 3 were pooled together to form the Abnormal group) are shown in Table 3A. The F -to-enter and F -to-remove were set at a much lower level than in the cell-by-cell analysis due to the lower number of cases.

The correct classification of normal cases and abnormal cases is respectively 94% and 70% using the Feature Set 2⁵ and 92% and 66% using Feature

⁵Selected Features: Max_Radius (stdv), Harm03 (mean) Con-

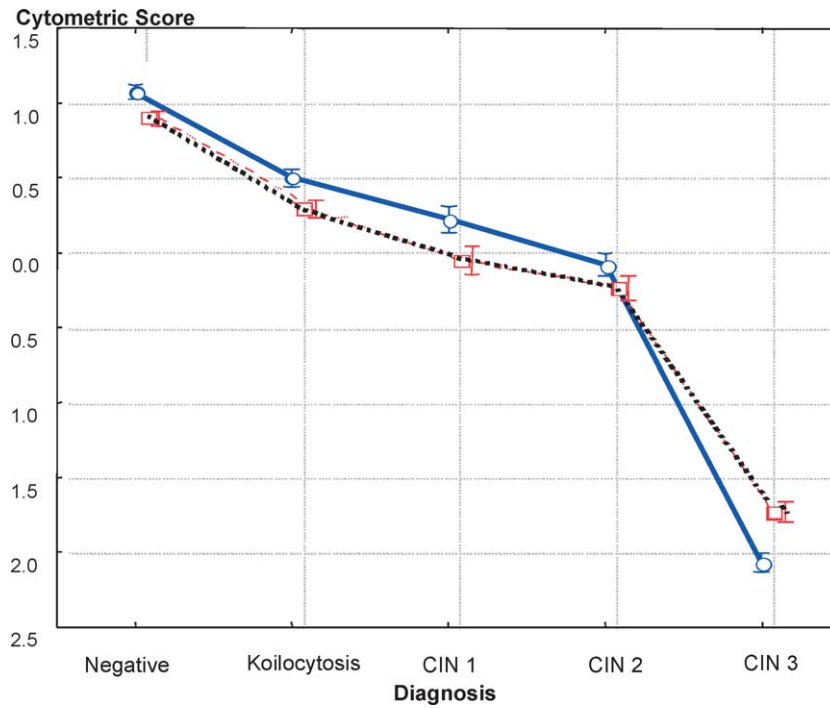


Fig. 2. Correlation between the morphometric scores and the pathology diagnosis. —○—: Normalisation by the mean DNA amount of epithelial cells. - -□- - -: Normalisation by the mode of the DNA amount of the lymphocytes.

Table 3A

Percentage of correct classification of normal and abnormal specimen given by a stepwise discriminant analysis using features set 2 and 3. *F*-to-enter: 4

Pathology	Features set	
	Set 2 (leukocyte normalization)	Set 3 (epithelial normalization)
Normal	94%	92%
Abnormal	70%	66%
Total	82%	79%

Set 3⁶ (Table 3A). Only one discrete texture feature (LowVsMed_DNA_2) was selected by the analysis with the Feature Set 2 and none with the Feature Set 3. Interestingly, in both analyses all but one of the features selected were the standard deviations. Stepwise discriminant analysis of normal and koilocytosis biopsies gives similar results for the two features sets; the respective overall classification was 69.5% with the

trast (stdv), Energy (stdv), OD_Kurt (stdv), Fract_Area1 (stdv), LowVsMedium_DNA (stdv).

⁶Selected Features: Max_Radius (stdv), Harm03 (mean) Contrast (stdv), Energy (stdv) OD_Kurt (stdv), Fract_Area1 (stdv).

Features Set 2⁷ and 68% with the Feature Set 3⁸ (Table 3B). The results of the discriminant analysis between koilocytosis samples and abnormal samples are almost identical for the Feature Sets 2⁹ and 3¹⁰ (85.5% and 85%) (Table 3C). Cell by cell variances were found to be the most discriminant features.

3.2. Reproducibility study

For 18 different biopsies and two identical imaging systems, two observers collected image data multiple times. The image acquisition sessions for the same observer were temporally well separated. Figure 3A shows the number of cells collected by each observer for each sample. This graph shows that one of the 2 observers collected systematically more cells than

⁷Selected Features: Max_Radius (mean), Mean_Radius (mean), Harm_fft1 (stdv), Harm_fft4_(mean).

⁸Selected Features: Max_radius (mean), Sphericity (stdv), Fract_Area2 (stdv), MH_DNA_Comp (mean), MH_Density_Objects (stdv), Long_Run_Mean (stdv), Run_Length_mean (mean).

⁹Selected Features: Max_Radius (stdv), Harm_fft4 (stdv), Run_Length1 (stdv).

¹⁰Max_Radius (stdv), Harm_fft4 (stdv), Run_Length_stdv (stdv), Fract_Dimension (stdv).

Table 3B

Percentage of correct classification of normal and koilocytosis specimen given by stepwise discriminant analysis using features set 2 and 3. *F*-to-enter: 2

Pathology	Features set	
	Set 2 (leukocyte normalization)	Set 3 (epithelial normalization)
Normal	71%	74%
Koilocytosis	65%	65%
Total	68%	69.5%

Table 3C

Percentage of correct classification of koilocytosis and abnormal specimen using given by stepwise discriminant analysis using features set 2 and 3. *F*-to-enter: 5

Pathology	Features set	
	Set 2 (leukocyte normalization)	Set 3 (epithelial normalization)
Koilocytosis	86%	89%
Abnormal	84%	82%
Total	85%	85.5%

the other one. Nevertheless, this difference between the two observers is much smaller for the Mean DNA amount and in the Mean area (Figs 3B, 3C). As well as the mean discriminant function scores for the three cell-by-cell classifiers generated for Feature Set 3. The inter observer variability is about the same as the intra observer variability (Figs 3D, 3E and 3F).

3.3. Specimen quality

The intrinsic nature of intra-epithelial lesions leads to sections of different quality. An “ideal” lesion is shown in Fig. 1. In this specimen, the integrity of the epithelium is preserved as well as the adjacent stroma compartment. In Fig. 4, we show different examples of specimen of lower quality. Figure 4A shows a lesion in which the upper part of the epithelium was stripped off, very likely during the sectioning process. No architectural features are extracted from this type of specimen, but morphometric analysis may still be performed. In some cases, the epithelial structure is completely lost and only individual epithelial cells can be segmented, but as illustrated by the Fig. 4B the stromal compartment is completely absent, making it impossible to collect lymphocytes. Other specimen quality issues arise when biopsies show obvious signs of tangential sectioning, or other artifacts that make the morphological assessment impossible. A strict visual quality control is performed for each sample. Among 1420 biopsies

analyzed so far 2% of the biopsies were disregarded as not being analyzable. In 10% of the cases, the architectural analysis could not be performed (no basement membrane or superficial layer), and among the remaining 90%, 4% were judged to be of poor quality (tangential sectioning or other artifacts).

4. Discussion

The opportunity to correlate the findings of quantitative pathology to fluorescence and reflectance spectroscopy is challenging. Several prominent studies have indicated that systems that attempt to define sub-categories of high and low grade lesions are neither reproducible nor comparable among institutions or among different observers [7,10,14,20,21]. These discrepancies are caused by the subjective nature of visual interpretation and difficulties in expressing and teaching the set of rules and techniques that constitute the “art” of clinical pathology [2,3,11]. This manuscript discusses different aspects of quantitative analysis of pre-neoplastic lesions of the cervix in a framework of a research study [1]. As part of the Program Project, more than 8000 biopsies will be analyzed. This large volume has forced us to scrutinize each step of our methodology in order to optimize the procedure without losing any essential biological or statistical information.

As expected, the inclusion of normalized features that include information on the distribution of genetic material within nuclei adds diagnostic power. In particular, the inclusion of discrete texture features showed the largest increases in pairwise discriminant power in most of the analyses performed. Table 2A shows an increase of approximately 10% using the feature sets that include normalized features over the feature subset that does not include normalized features when discriminating normal versus CIN 3 on a cell-by-cell basis. This is biologically plausible given the known genetic instability associated with neoplastic progression.

When it comes to the method of normalization, our data show that we can reduce the duration as well as the subjectivity of the analysis without losing any biological information or discrimination power. First, we showed that we could possibly replace the use of lymphocytes as internal reference cells. The extremely time consuming step of collecting the lymphocyte data is justified only if it increases discrimination power among the different diagnostic grades over the simpler method which does not require the separate collection

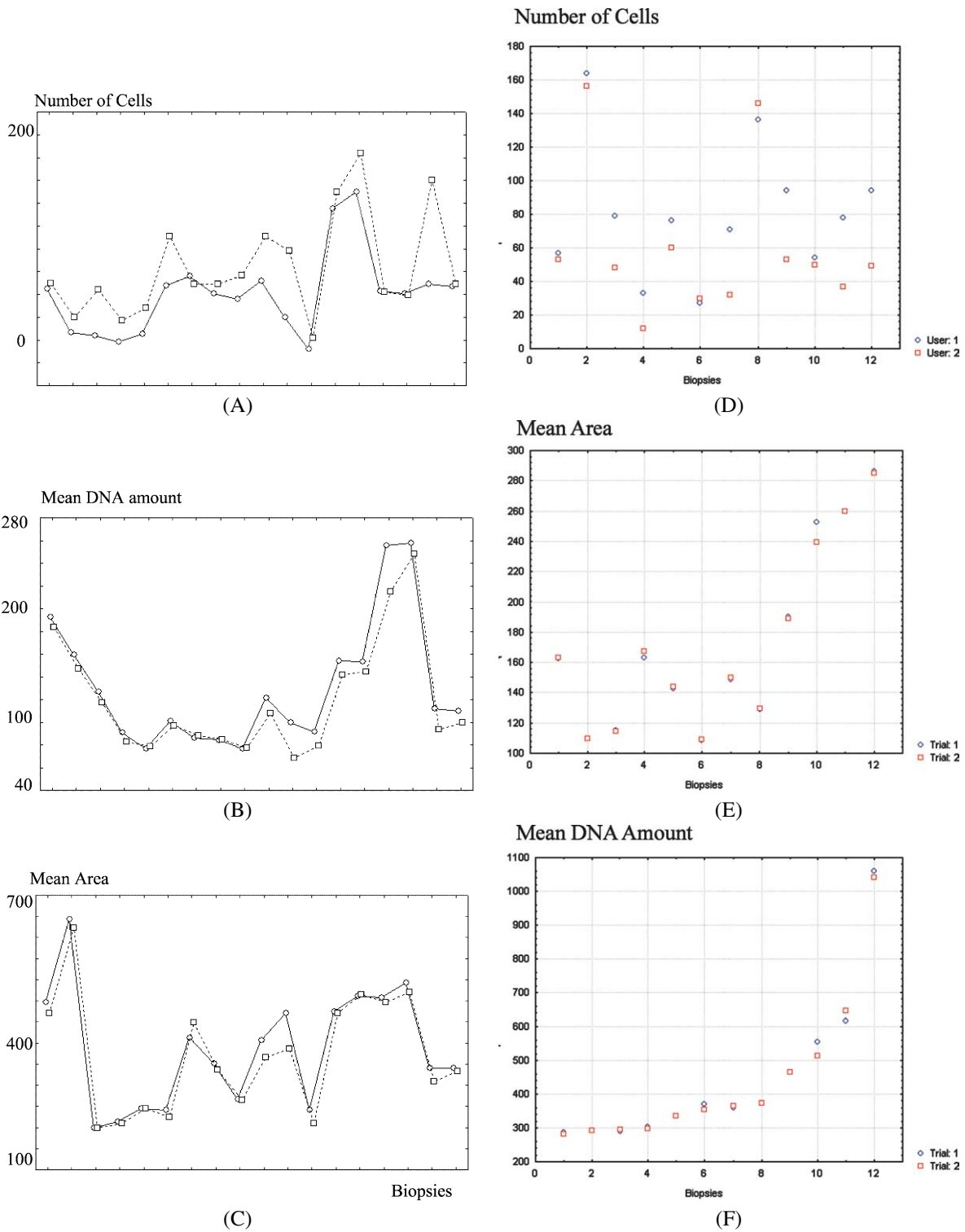


Fig. 3. Inter-observer and Intra-observer Variability (2 users). A: Number of cells collected: inter-observer variability. B: Mean DNA amount of the epithelial cells: inter-observer variability. C: Mean area measured of the epithelial cells: inter-observer variability. D: Number of cells collected: intra-observer variability. E: Mean area of the epithelial cells: intra-observer variability. F: Mean DNA amount of the epithelial cells: intra-observer variability.

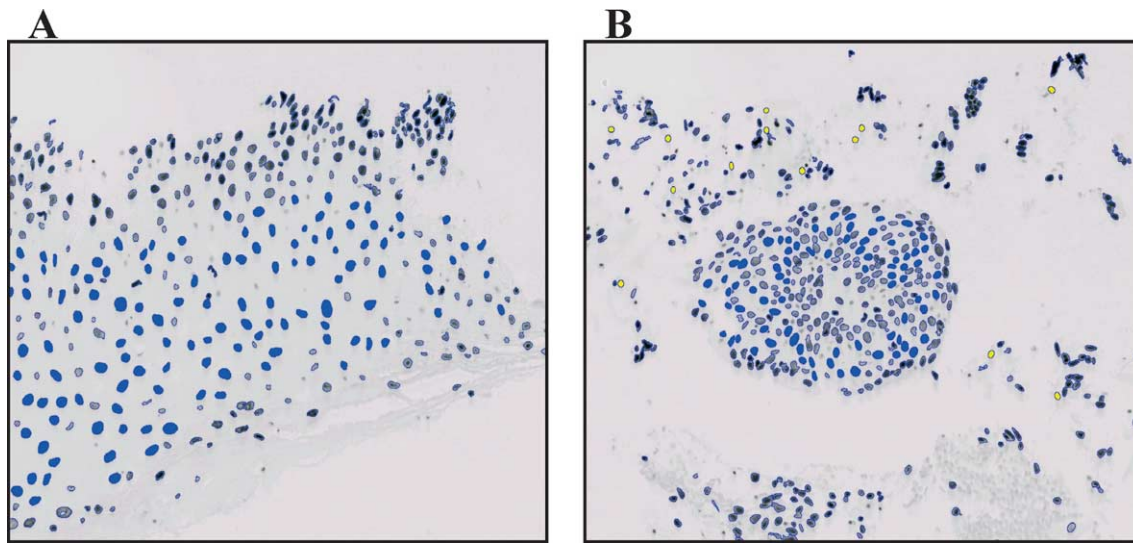


Fig. 4. Examples of different cervical sections. A: The superficial membrane is difficult to recognise and it is impossible to locate any lymphocytes. B: The stroma compartment is absent: few leukocytes are present (yellow) and the structure of the epithelium is completely lost, making impossible the architectural analysis.

of a control set of cells. Our data suggest that it is the opposite case: there is the same or better performance using the mean DNA of epithelial cells than normalization by the mode of the lymphocytes. This needs to be validated with a larger sample size.

One can argue that there is a loss robustness or precision using the normalization by the mean DNA of epithelial cells. The same argument can be made against the robustness and precision of the estimation of the mode of a distribution from a sample of 10 or 20 lymphocytes. Indeed, the estimation of the mode a distribution is a difficult statistical problem and is very sensitive to the number of bins of the histogram. Furthermore, in a significant number of samples, the quality of the biopsy does not allow the collection of lymphocytes because of insufficient stroma. All of our data suggests that the mean DNA amount of the epithelial cells is acceptable.

By not normalizing with a control set of cells (lymphocytes), we lose the possibility of the assessment of ploidy in these lesions. Ploidy measurement could only be performed after a very time-consuming and tedious process of precisely measuring the section thickness, which is impractical to perform in a routine practice. Mairinger et al. [13] compared different mathematical methods to correct histograms obtained from DNA-measurements on thin sections. They concluded that DNA measurements on thin paraffin sections are possible only if the actual section thickness is known. The DNA Index, which is a relative measure of the to-

tal nuclear DNA amount, had a significant discriminant power in some of our analyses [18,19].

In order for quantitative histopathology to be widely applicable for a given organ site, it is necessary that the operator specific bias is reducible to the extent that different observers are indistinguishable and hence that the results are objective and reproducible. Only if this can be shown will it be possible for the techniques to be transferred to a different laboratory. We have shown here that the inter- and intra-observer variability are of similar size and hence that we have achieved this goal. Clearly if this methodology is to be used routinely in clinical practice, we need to repeat this analysis on more observers and with more samples. We plan a larger inter-institutional inter- and intra-observer study to verify our preliminary results.

Basal and parabasal cells are morphologically different from intermediate and superficial cells. The dysplastic process is characterized by increasing dedifferentiation, nuclear enlargement, and hyperchromasia [6,7,9,21] as well as architectural disorganization of the epithelium. Exhaustive sampling is theoretically the only approach that would fully capture the heterogeneity of any phenotypic feature recorded from normal tissue going from the basal membrane to the surface of the superficial layer. In the study we are interested first and foremost in the detection of grades dysplasia, whose changes occur predominately in the distribution of cell phenotypes within the intermediate layer [7,9,21]. Concerning the sampling protocol,

we were facing two main options, both with advantages and limits. The first approach “exhaustive sampling” consists of collecting many cells as possible from the all layers within the Region of Interest, alternatively the method proposed here selects only non-overlapping (or slightly overlapping) cells from only the intermediate layer where the majority of dysplastic changes are manifest. Basal and parabasal cells are highly overlapping and exhaustive selection of these cells is extremely subjective, time consuming and would introduce bias from the separation of the overlapping or touching objects. Selection of superficial cells is a tedious task, since some of them are dying cells, flattened, and also in a significant number of sections the top layer of the epithelium is not present. By selecting only the cells in parabasal and intermediate layers, the number of overlapping cells is much lower and we look at a more morphologically and functionally homogeneous population. It allows us to measure, in a more reproducible and robust way, any dysplastic changes as a degree of deviation from the normal intermediate population of cells (Fig. 5). The

sampling strategy proposed here reflects a trade-off between consistency, objectivity, and speed of data selection on the one hand, and exhaustive sampling of the lesion on the other hand. In a planned future study, we will report on the combination of architectural and exhaustive morphometric analysis [cellular sociology] of more than 8000 cervical specimens. Clearly in these 280 specimens there are insufficient specimens to compare the classification of the samples using data from the superficial layer and without using data from the superficial layer. As the study progresses and at study termination there will be sufficient samples to compare these two methodologies. Thus we will see if it is necessary to measure the superficial layer with quantitative histopathology.

In this interim analysis, we found no adverse impact on diagnostic power from normalization with the mean DNA from the sampled epithelial cells rather than the mode of the lymphocytes. Our preliminary data show that the inter- and intra-observer variability of several important diagnostic features are essentially the same. We have argued that sampling from the intermediate

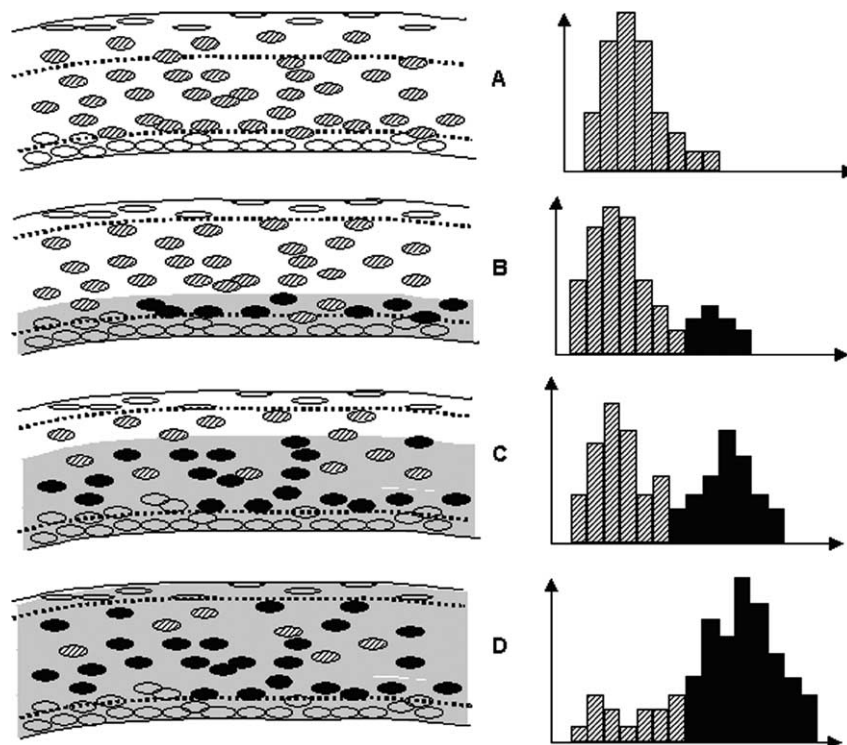


Fig. 5. Schematic illustration of the sampling process. A – Normal epithelium, D – CIN 3 epithelium. The two dashed lines delineate the sampling area, from which cells will be collected. White cells: not collected; dashed and black cells: collected. Dark cells represent cells with detectable nuclear abnormalities. As the number of “abnormal” move towards the surface of the epithelium, the proportion of collected abnormal cells increases. Such sampling reflects the dysplastic changes, while reducing the manual and subjective selection of cells <http://www.esacp.org/acp/2003/XX-Y/guillaud.htm>.

layer will provide more objective and accurate results. More data is needed to validate these observations; it is being collected, and will be reported on in the future. Optical technologies, like fluorescence and reflectance spectroscopy, hold great promise to change the way health care is practiced because they are portable and low cost. In our studies, while we are using clinical pathology as the gold standard and we are subjecting quantitative pathology to technology assessment, we will have the data to validate clinical pathology against quantitative pathology.

Acknowledgements

This research was supported by the National Cancer Institute under Program Project Grant 3PO1-CA82710-04. The authors thank Deanna Haskins, Anita Cairo, Tatiana Alexeenko, Iouri Boiko, Nan Earle, and Trey Kell for their contributions to this project.

Appendix. Categories and Features Available

Category	Features
	<i>Cytometric features</i>
Morphometry (43)	
Size (4)	Area, mean_radius, variance_radius, maximum_radius
Shape (5)	Eccentricity, sphericity, elongation, compactness, inertia_shape
Boundaries (34)	Low_freq_fft, Freq_low_fft, Harm01-32_fft
Photometric (5)	DNA_index OD_max, OD_Var, OD_skew, OD_kurt
Discrete texture (24)	Low, medium, and high DNA amount Low, medium, and high DNA area Low, medium, high and medium_high DNA compactness Low, medium, high and medium_high DNA average distance Low, medium, medium-high and high density object Low, medium, and high centre mass Low_vs_medium, low_vs_High and Low_vs_Medium-High DNA
Markovian texture (7)	Entropy, energy, contrast, correlation, homogeneity, cl_shade, cl_prommence
Non-Markovian texture (5)	Density_light_spots, density_dark_spots, center_of_gravity, range_extreme, range_average
Fractal texture (3)	Fractal_areal, fractal_area2, fractal_dimension
Run length texture (20)	Short_runs_mean, Short_run_stdv, Short_run_min, Short_run_max Long_runs_mean, Long_run_stdv, Long_run_min, Long_run_max Gray_level_mean, Gray_level_stdv, Gray_level_min, Gray_level_max Run_length_mean, Run_length_stdv, Run_length_min, Run_length_max Run_Percent_mean, Run_Percent_stdv, Run_Percent_min, Run_Percent_max
	<i>Architectural features</i>
Entropy	
Features derived from Voronoi polygons (18)	Area (mean), Area (standard deviation), Area (skewness), Area (kurtosis), Area disorder, Perimeter (mean), Perimeter (standard deviation), Perimeter (skewness), Perimeter (kurtosis), Roundness factor (mean), Roundness factor (standard deviation), Roundness factor (skewness), Roundness factor (kurtosis), Roundness factor heterogeneity, Number of sides (mean), Number of sides (standard deviation), Number of sides (skewness), Number of sides (kurtosis)
Features derived from Delaunay graph (4)	Nearest neighbor distance (mean), Nearest neighbor distance (standard deviation), Delaunay nearest neighbor distance (mean), Delaunay nearest neighbor distance (standard deviation)
Features derived from the minimum spanning tree (MST) (7)	Percentage of nuclei with one connected nucleus, Percentage of nuclei with two connected nuclei, Percentage of nuclei with more than two connected nuclei, Length of the MST edge (mean), Length of the MST edge (standard deviation)

References

- [1] E. Artacho-Perula, R. Roldan-Villalobos, J. Salas-Molina and R. Vaamonde-Lemos, Histomorphometry of normal and abnormal cervical samples, *Analyt. Quantit. Cytol. Histol.* **15**(4) (1993), 290–297.
- [2] J.P. Baak, The framework of pathology: good laboratory practice by quantitative and molecular methods, *J. Pathol.* **198**(3) (2002), 277–283.
- [3] H.K. Choi, T. Jarkrans, E. Bengtsson, J. Vasko, K. Wester, P.U. Malmstrom and C. Busch, Image analysis based grading of bladder carcinoma. Comparison of object, texture and graph based methods and their reproducibility, *Analyt. Cell. Pathol.* **15**(1) (1997), 1–18.
- [4] A.M. Doudkine, C. MacAulay, N. Poulin and B. Palcic, Nuclear texture measurements in image cytometry, *Pathologica* **87** (1995), 286–299.
- [5] M. Follen and R. Richards-Kortum, Emerging technologies and cervical cancer [editorial], *J. Nat. Cancer Inst.* **92**(5) (2000), 363–365.
- [6] A.G. Hanselaar, N. Poulin, M.M. Pahlplatz, D. Garner, C. MacAulay, J. Maticic, J. LeRiche and B. Palcic, DNA-cytometry of progressive and regressive cervical intraepithelial neoplasia, *Analyt. Cell. Pathol.* **16**(1) (1998), 11–27.
- [7] M.K. Heatley, How should we grade CIN?, *Histopathology* **40**(4) (2002), 377–379.
- [8] S.M. Ismail, A.B. Colclough, J.S. Dinnen, D. Eakins, D.M. Evans, E. Gradwell, J.P. O’Sullivan, J.M. Summerell and R. Newcombe, Reporting cervical intra-epithelial neoplasia (CIN): intra- and interpathologist variation and factors associated with disagreement, *Histopathology* **16**(4) (1990), 371–376.
- [9] S.M. Ismail and A.N. Fiander, How should we grade CIN? Grading cervical intraepithelial neoplasia, *Histopathology* **40**(4) (2002), 385–390.
- [10] S.J. Keenan, J. Diamond, W.G. McCluggage, H. Bharucha, D. Thompson, P.H. Bartels and P.W. Hamilton, An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN), *J. Pathol.* **192**(3) (2000), 351–362.
- [11] L. Koss, *Diagnostic Cytology and Its Histopathologic Basis*, J.B. Lippencott Co, Philadelphia, PA, 1992.
- [12] C. MacAulay and B. Palcic, An edge relocation segmentation algorithm, *Analyt. Quantit. Cytol. Histol.* **12**(3) (1990), 165–171.
- [13] T. Mairinger and A. Gschwendtner, Comparison of different mathematical algorithms to correct DNA-histograms obtained by measurements on thin liver tissue sections, *Analyt. Cell. Pathol.* **11**(3) (1996), 159–171.
- [14] W.G. McCluggage, M.Y. Walsh, C.M. Thornton, P.W. Hamilton, A. Date, L.M. Caughley and H. Bharucha, Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system, *Br. J. Obstet. Gynaecol.* **105**(2) (1998), 206–210.
- [15] B. Palcic, Nuclear texture: Can it be used as a surrogate endpoint biomarker?, *J. Cell. Biochem. (Suppl.)* **19** (1994), 40–46.
- [16] D.M. Parkin, ed., *Cancer Incidence in Five Continents*, Vol. VII, IARC Press, 1997.
- [17] D.M. Parkin, P. Pisani and J. Ferlay, Estimates of the worldwide incidence of 25 major cancers in 1990, *Int. J. Cancer* **80**(6) (1999), 827–841.
- [18] N. Poulin, I. Boiko, C. MacAulay, C. Boone, K. Nishioka, W. Hittelman and M.F. Mitchell, Nuclear morphometry as an intermediate endpoint biomarker in chemoprevention of cervical carcinoma using alpha-difluoromethylornithine, *Cytometry* **38**(5) (1999), 214–223.
- [19] R.G. Steinbeck, Proliferation and DNA aneuploidy in mild dysplasia imply early steps of cervical carcinogenesis, *Acta Oncologica* **36**(1) (1997), 3–12.
- [20] B. Weyn, W. Tjalma, G. Van De Wouwer, A. Van Daele, P. Scheunders, W. Jacob and E. Van Marck, Validation of nuclear texture, density, morphometry and tissue syntactic structure analysis as prognosticators of cervical carcinoma, *Analyt. Quantit. Cytol. Histol.* **22**(5) (2000), 373–382.
- [21] G.E. Wilson, How should we grade CIN? The classification of cervical intraepithelial neoplasia, *Histopathology* **40**(4) (2002), 380–384.