

Prognostic classification of early ovarian cancer based on very low dimensionality adaptive texture feature vectors from cell nuclei from monolayers and histological sections

Birgitte Nielsen^{a,b,*}, Fritz Albrechtsen^a,
Wanja Kildal^b and Håvard E. Danielsen^{b,c}

^a *Department of Informatics, University of Oslo, P.O.Box 1080 Blindern, N-0316 Oslo, Norway*

^b *Division of Digital Pathology, The Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway*

^c *Division of Genomic Medicine, The University of Sheffield, Sheffield, S102TN England*

Accepted 13 November 2001

In order to study the prognostic value of quantifying the chromatin structure of cell nuclei from patients with early ovarian cancer, low dimensionality adaptive fractal and Gray Level Cooccurrence Matrix texture feature vectors were extracted from nuclei images of monolayers and histological sections. Each light microscopy nucleus image was divided into a peripheral and a central part, representing 30% and 70% of the total area of the nucleus, respectively. Textural features were then extracted from the peripheral and central parts of the nuclei images.

The adaptive feature extraction was based on Class Difference Matrices and Class Distance Matrices. These matrices were useful to illustrate the difference in chromatin texture between the good and bad prognosis classes of ovarian samples. Class Difference and Distance Matrices also clearly illustrated the difference in texture between the peripheral and central parts of cell nuclei. Both when working with nuclei images from monolayers and from histological sections it seems useful to extract separate features from the peripheral and central parts of the nuclei images.

*Corresponding author: Dr. B. Nielsen, Department of Informatics, University of Oslo, P.O.Box 1080 Blindern, N-0316 Oslo, Norway. Fax: +47 22 85 24 01; E-mail: birgittn@ifi.uio.no.

1. Introduction

Most women undergoing treatment for early ovarian cancer have a good prognosis, but about 20% will eventually die of the disease. Identifying patients with increased risk of relapse is important, as it could be used to select patients in need for adjuvant treatment after surgery. In this study, we have extracted low dimensionality adaptive texture feature vectors for two different texture analysis methods in order to quantify the chromatin structure of cell nuclei from patients with early ovarian cancer.

Texture features used for classification in digital pathology are often selected from a large number of rather *ad hoc* features, where uncertainty related to parameter settings also contribute to a relatively high dimensionality of the feature space. Whatever sophisticated feature selection algorithms we use, the risk of purely coincidental “good” feature sets may become alarmingly high, if the available data set is limited. This is particularly a problem if separate training and test sets are not used [16]. The ordinary Gray Level Cooccurrence Matrix (GLCM) method [6] is a good example of a texture method with a large number of *pre-defined* features combined with a number of free parameters (number of gray levels, inter-pixel distance, orientation). Identifying a few consistently valuable features is important for many applications as it improves reliability and enhances our understanding of the phenomena that we are modelling. We have therefore proposed a small number of *adaptive* texture features that can be extracted by applying the same approach to several texture methods [2,3,13,14]. These adaptive texture features are based on *Class Difference Matrices* and *Mahalanobis Class Distance Matrices*.

The class difference matrix for a given texture method (i.e., the Gray Level Cooccurrence Matrix method [6]) contains the difference between the average frequency matrix (i.e., the average GLCM) of each of the two classes. The class distance matrix contains information on which matrix element that – when taken one matrix element at a time – will provide high class discrimination. Earlier [2,3,13], we have found that the class distance matrices between different pairs of mouse liver nuclei classes contained large areas of consistently high values. These areas corresponded to separate areas of consistently positive or negative class difference values. For each nucleus image, we could therefore combine information from the entries of the normalized frequency matrix (i.e., the GLCM) based on the class difference and distance matrices, to obtain a small number of adaptive texture features for classification.

There are many reasons to believe that the chromatin distribution in the nucleus is by no means a random process, resulting in a stationary texture across the whole cell nucleus. Thus, measuring texture for a biologically relevant subpart of the nucleus can be expected to give better results than simply doing it for the nucleus as a whole. Of particular interest is quantification of the tendency of condensed heterochromatin to be located adjacent to the nuclear envelope [22]. Based on the Statistical Geometric Features (SGF) of Chen et al. [4], Walker and Jackway [19] proposed a small set of additional features tailored to detect some of these radial differences. However, the features designed to measure the average displacement and average inertia of chromatin clumps from the center of gravity of the nucleus, were not reported to capture chromatin changes during cell dysplasia, when tested on a set of cervical cell images [19]. Using a partitioning of the cell nuclei into a peripheral and a central part, we have repeatedly seen that the classification based on chromatin texture will benefit from a radial dichotomy of the texture feature extraction [2,3,11–13]. In transmission electron micrographs (TEM) of ultrathin sections of liver cell nuclei we have found that there is a change in the amount and arrangement of condensed heterochromatin close to the nuclear envelope from normal to malignant cells. The heterochromatin structures are fewer and larger in the malignant cell nuclei. This is reflected in changes in texture parameters, particularly if these are extracted from the peripheral part of the nuclei.

The images utilized in the earlier studies [2,3,11–13] were obtained through electron microscopy and

were manually segmented. The potential of using these methods in clinical work becomes much greater if they can be applied to light microscopy images. In this study we have therefore applied two of the methods described earlier [13,14] on light microscopy images of cell nuclei from ovarian cancer classified as FIGO stage 1. The material consisted of two groups. The patients in the “good prognosis” group survived the follow-up period without a relapse, while the patients in the “bad prognosis” group died of cancer-related disease or relapsed during the follow-up period.

The distribution and organization of condensed chromatin within the cell nuclei may be studied either in histological sections or in monolayers. There is no a priori knowledge about which of the two modalities that will give the optimal prognostic information when this particular texture based classification is performed. The present study therefore includes and compares texture analysis of light microscopy images of nuclei from both histological sections and monolayers.

The textural phenomena that are observable close to the nuclear membrane in images of slices through the cell nuclei, will not necessarily be observable in the extremely peripheral part of projection images of the whole nuclear volume, as in monolayer images. Thus, the radial differentiation of the textures may be very different in the two sets of images. In addition, we know that the monolayer material has gone through a preparation process quite different from that of the histological sectioned material. Therefore, the visual texture observed in the digital images is quite different. This is even reflected in the average size of the cell nuclei. So the optimal texture analysis method is not necessarily the same, and the optimal parameters within a single method may also differ. This all contributes to making a comparison of textures in sections and monolayers and the differentiation into central and peripheral nuclear segments an interesting issue.

We have chosen to compare the widely known and very popular GLCM method to the new and powerful fractal signature/lacunarity matrix method [13]. Earlier [13,14], we have seen that these methods will benefit from the class distance approach. The lacunarity matrix method also provides a very useful tool for quantitative description of structural differences between the textures [11].

2. Material and methods

2.1. Early ovarian cancer

The cases of early ovarian cancer included in this study were selected from a larger material of ovar-

ian cancer classified as FIGO stage 1 [9]. The patients were treated at The Norwegian Radium Hospital during 1982–1989. The patients who survived the follow-up period without a relapse were included in the good prognosis group. The minimum length of follow-up for patients alive without a relapse was ten years. The patients who died of cancer-related disease or relapsed during this period were included in the bad prognosis group.

2.1.1. Monolayers

Twenty cases were included in each of the two groups. Paraffin-embedded tissue samples fixed in 4% buffered formalin were sectioned ($2 \times 50 \mu\text{m}$) and enzymatically digested (Sigma protease, type XXIV, Sigma Chemical C., St. Louis, Missouri, USA) for the preparation of isolated nuclei (monolayers) [7]. The nuclei were Feulgen–Schiff stained according to an established protocol [17]. Blocks were selected by the pathologist, who selected the tumour tissue to be prepared.

The Fairfield DNA Ploidy System (Fairfield Imaging LTD, Kent, England), which consisted of a Zeiss Axioplan microscope equipped with a 40/0.75 objective lens (Zeiss), a 546 nm green filter and a black and white high resolution digital camera (C4742-95, Hamamatsu Photonics K.K., Hamamatsu, Japan) was used. A shade correction was performed for each image field and the image was stored in 1024×1024 pixels with a gray level resolution of 10 bits per pixel. The pixel resolution was 166 nm per pixel on the cell specimen.

Trained personnel performed a screening of the cells in the microscope and selected tumour cells for the analysis. Stromal cells, necrotic cells, doublets or cut cells were disregarded. The nuclei were segmented from the background by using a global threshold. The histograms of all nuclei images were normalized to the same mean value (650.0) and standard deviation (120.0). This was done to normalize the first order statistics while utilizing the whole range of gray levels, and avoid clipping of the histograms. Figure 1 shows examples of cell nuclei from each of the two groups. The mean number of measured tumour nuclei per sample was 256, ranging from 185 to 294 nuclei. This material was also included in [14].

2.1.2. Histological sections

In the experiment on histological sections, seventeen cases were included in the good prognosis group and eighteen cases in the bad prognosis group. Due to practical preparation problems only twenty-nine of

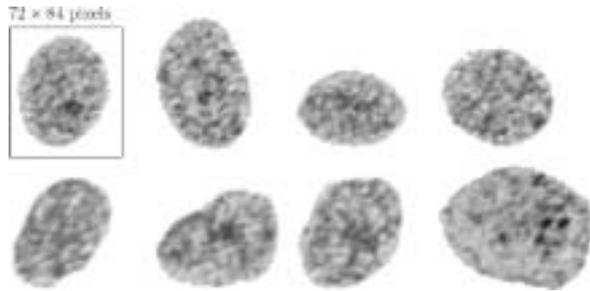


Fig. 1. Four cell nuclei from a good prognosis sample (upper) and four nuclei from a bad prognosis sample (lower). The nuclei were selected from monolayers.



Fig. 2. Four cell nuclei from a good prognosis sample (upper) and four nuclei from a bad prognosis sample (lower). The nuclei were selected from histological sections.

the forty cases that were used in the monolayer analysis were also included in the study of histological sections (twelve good prognosis samples and seventeen bad prognosis samples). Paraffin embedded tissue samples fixed in 4% buffered formalin were sectioned ($2 \mu\text{m}$), put on slides, prepared and Feulgen–Schiff-stained as described for monolayers with the following exceptions: enzymatic digestion (Sigma protease) were used on the sections for 25 minutes. Afterwards the sections were post-fixed in 4% formalin for 60 minutes. The histograms of all nuclei images were normalized to the same mean value (640.0) and standard deviation (145.0). Thus, the whole range of gray levels is used and no histogram clipping occurs, but the normalization parameters are slightly different from those of the monolayer data. Figure 2 shows examples of cell nuclei from each of the two groups. The mean number of measured nuclei per sample was 277, ranging from 110 to 329 nuclei.

2.2. Peel-off scanning

In order to extract separate estimates of texture features in the periphery and center of the 2D cell nu-

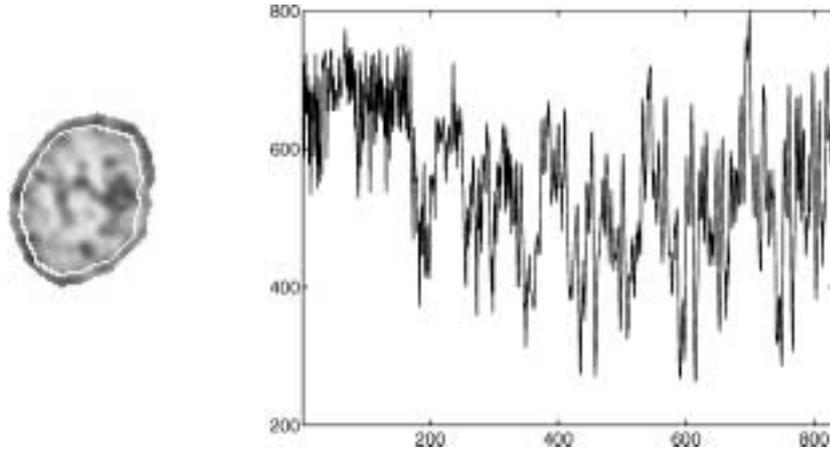


Fig. 3. Left: A cell nucleus from a good prognosis sample. The border between the 30% peripheral and 70% central part is outlined as a thin white line. Right: The 1D gray level signal corresponding to the peripheral part of the 2D nuclei image. The nucleus is selected from a histological section.

cleus image, an 8-neighbour backtracking bug follower [15] has been used in a spiral scanning algorithm [11]. Starting with the 2D image of the segmented cell nucleus, we follow the (outer) contour of the nucleus, and spiral inwards as we peel off pixels, layer by layer, from the nucleus, forming a discrete 1D gray level signal. In this type of spiral scanning, which we call “peel-off scanning”, the resulting 1D gray level curve only reflects the size and contrast of structures inside the nucleus, not the shape of the nuclear membrane. We also note that the pixels of the 1D signal are assumed to be equidistant, regardless of whether they were 4- or 8-neighbours in the original 2D context.

In TEM images of ultrathin sections of liver cell nuclei we have previously found that there is a change in the amount and arrangement of condensed chromatin close to the nuclear envelope from normal to malignant cells. The condensed chromatin structures are fewer and larger in the malignant cell nuclei. A simple geometric modeling of this observed 2D phenomenon will indicate that the condensed chromatin structures close to the nuclear membrane will overlap in a projected 2D image of the 3D monolayer structure. Thus, the textural phenomena that are observable close to the nuclear membrane in images of sections through the cell nuclei will not necessarily be observable in the extreme peripheral part of projection images of the whole nuclear volume, as in monolayer images. As it is very hard to predict the optimal way of partitioning the nucleus for the two modalities, we have chosen to use the same area fraction for both the histological sections and the monolayers images. Therefore, the 1D gray level signal resulting from the “peel-off scanning” of each cell

nucleus image was divided into a peripheral 30% and a central 70% segment. Since each pixel of the image is visited only once in this scanning process, we also get a 30%/70% split in terms of nuclear area. The texture features in the present study were extracted from these two segments. Figure 3 shows an example of a 2D cell nucleus image and the resulting 1D signal from the peripheral 30% segment. The peripheral segment correspond to about 5–6 pixel layers of the nucleus in the “peel-off scanning”. The repetitive pattern in the 1D signal correspond to one circumference of the remaining area of the cell nucleus in the “peel-off scanning” process.

2.3. Gray Level Cooccurrence Matrices

The Gray Level Cooccurrence Matrix (GLCM) method [6] is a way of extracting second order statistical texture features. A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G , in the image. In the case of our 1D signal, the matrix element $P(i, j|d)$ contains the probability for changes between gray levels i and j at a particular pixel displacement distance d . In this study we have used $G = 16$ gray levels and $d = 3$. Separate GLCM were accumulated from the peripheral 30% and the central 70% 1D segments.

2.3.1. GLCM Class Distance Matrices

For each of the $N(\omega_c)$ training set images of class ω_c we find the Gray Level Cooccurrence Matrix $P_n(i, j)$. For each element in this matrix we then estimate the class conditional probability distributions $P_P(i, j|\omega_c)$

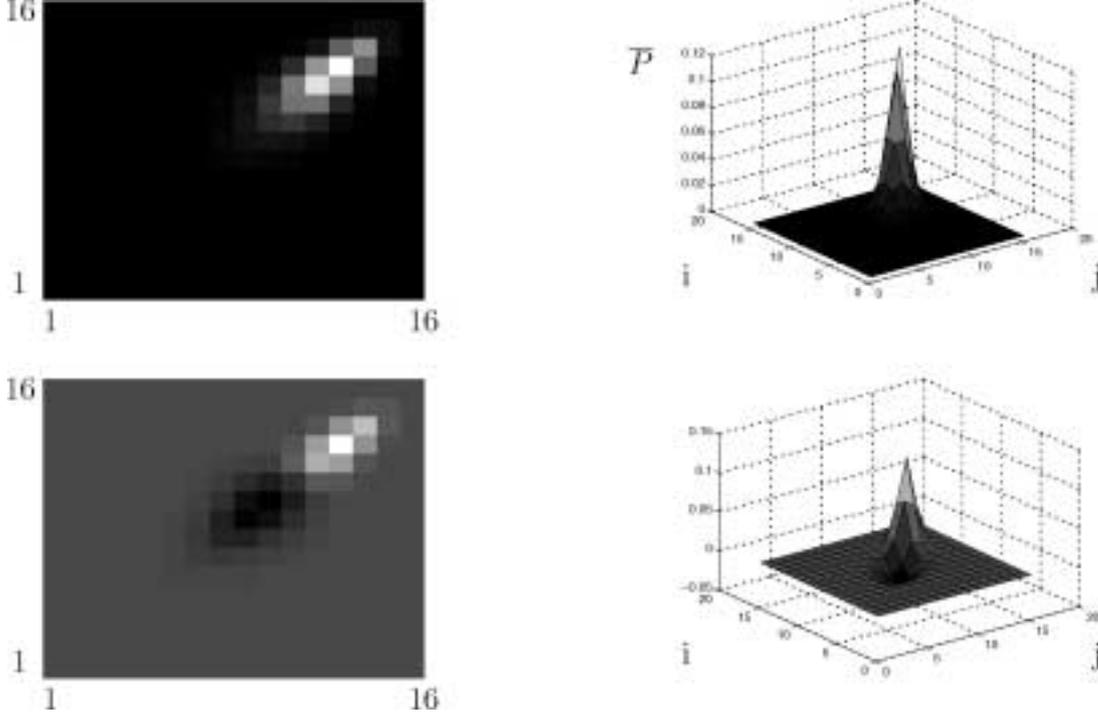


Fig. 4. Upper: The average normalized GLCM, $\bar{P}(i, j|\omega_c)$, computed from the peripheral part of all the monolayer nuclei images in the good prognosis group. Lower: The difference between the $\bar{P}(i, j|\omega_c)$ computed from the peripheral and central parts of monolayer nuclei images in the good prognosis group. The matrices are visualized both as gray-scale and surface plots.

of the normalized matrix value, based on all the $N(\omega_c)$ training set images of class ω_c . Based on these class conditional distributions, we compute the average normalized GLCM, $\bar{P}(i, j|\omega_c)$, for each class ω_c , the class variance matrix $\sigma_P^2(i, j|\omega_c)$, the Class Difference Matrix $\Delta_P(i, j|\omega_1, \omega_2)$, and finally the GLCM Mahalanobis Class Distance Matrix, $J_P(i, j|\omega_1, \omega_2)$ between the two classes ω_1 and ω_2 [14]:

$$\bar{P}(i, j|\omega_c) = \frac{1}{N(\omega_c)} \sum_{n=1}^{N(\omega_c)} P_n(i, j),$$

$$\Delta_P(i, j|\omega_1, \omega_2) = \bar{P}(i, j|\omega_1) - \bar{P}(i, j|\omega_2), \quad (1)$$

$$J_P(i, j|\omega_1, \omega_2) = \frac{(\bar{P}(i, j|\omega_1) - \bar{P}(i, j|\omega_2))^2}{(\sigma_P^2(i, j|\omega_1) + \sigma_P^2(i, j|\omega_2))/2}.$$

Figure 4 (upper) shows the average GLCM computed from the peripheral part of all the monolayer nuclei images in the good prognosis class. Examples of GLCM Class Difference and Distance Matrices are shown in Figure 5.

2.3.2. Adaptive GLCM features

For each nucleus image, we combine those GLCM matrix elements that contribute the most to the class

separability into adaptive GLCM features, simply by using the squared class distance matrix values as summation weights [14]. We utilize the fact that in the two-class problem, the class difference has a *sign*, depending on whether the first or the second class matrix element contains the highest average probability. So the weighted summation is performed over the two disjoint partitions of the class difference matrix Δ_P by using the binary decision functions:

$$\Omega_{(+)}(i, j|\Delta_P) = \begin{cases} 1 & \text{if } \Delta_P(i, j|\omega_1, \omega_2) \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

$$\Omega_{(-)}(i, j|\Delta_P) = 1 - \Omega_{(+)}(i, j|\Delta_P).$$

The two adaptive GLCM features for any given image of class ω_1 or ω_2 having a GLCM matrix $P(i, j|d)$ are then given by the matrix summation [14]:

$$F_{P(+)} = \sum_{i,j} P(i, j) \cdot J_P^2(i, j) \cdot \Omega_{(+)}(i, j|\Delta_P),$$

$$F_{P(-)} = \sum_{i,j} P(i, j) \cdot J_P^2(i, j) \cdot \Omega_{(-)}(i, j|\Delta_P). \quad (3)$$

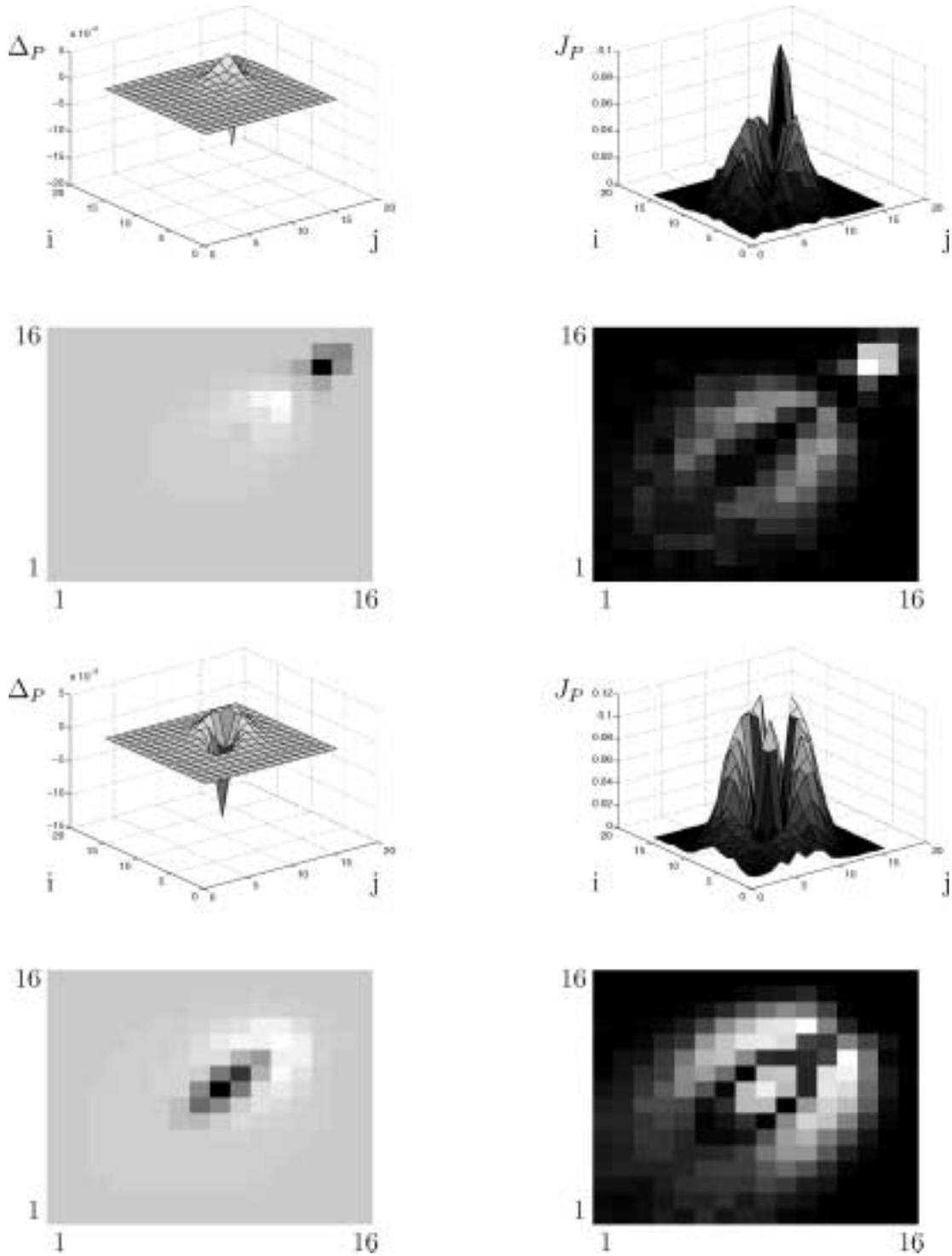


Fig. 5. Left: The GLCM Class Difference Matrix, $\Delta_P(i, j | \omega_1, \omega_2)$, computed from the peripheral 30% (upper) and the central 70% (lower) 1D segments of nuclei from monolayers. The positive values in the surface plots (and the lighter areas in the gray-scale plots) correspond to matrix elements that are more probable for the good prognosis class than for the bad prognosis class. Right: The Mahalanobis Class Distance Matrix, $J_P(i, j | \omega_1, \omega_2)$, between the good and bad prognosis classes, based on the GLCM matrices, computed from the peripheral 30% (upper) and the central 70% (lower) segments.

2.4. Fractal Signature Vectors

Our method for estimation of fractal features from 1D curves [1,11,13] is based on the polygonization method of Wall and Danielsson [21]. This method gives a linear approximation to planar curves by obtaining a subset of the original set of points in such a way that the new curve formed by a sequence of line segments joining the points in the subset does not “deviate” more than a certain amount from the original points. There are several ways of quantifying the “deviation”. Here we have used the area-to-line-length ratio of Wall and Danielsson [21]. Wall–Danielsson’s method steps from point to point through an ordered sequence of points (x_i, f_i) , and outputs the previous point as a new breakpoint if the area deviation A_i per unit length of the approximating line segment s_i exceeds a prespecified tolerance, T . If $|A_i|/s_i < T$, i is incremented and (A_i, s_i) is recomputed. Otherwise, the previous point is a new breakpoint and the previous value of s_i is stored. This method is purely sequential and very fast, particularly when the x -values are assumed equidistant ($\Delta x = 1$).

We approximate the 1D gray level signal by polygonization with several values of the tolerance, T . The set of tolerance values is computed from a Fibonacci sequence [1,11,13]. For each tolerance value the total length of the line segments that approximate the curve is summed up by $S_T = \sum_i s_i$.

For each cell nucleus we compute the slope of the best fitted straight line through three and three points in the $\{\log(T), \log(S_T)\}$ -domain. We call this the fractal signature, $\varphi_n(T)$ [13] of the central T -value.

2.4.1. Fractal Signature Class Distance Vectors

For each tolerance value T , the class conditional probability distributions $P_\varphi(T|\omega_c)$ of the fractal signature $\varphi(T)$ are estimated, based on the $N(\omega_c)$ cells of class ω_c . We compute the average Fractal Signature Vector $\bar{\varphi}(T|\omega_c)$ for each class ω_c , the class variance vector $\sigma_\varphi^2(T|\omega_c)$, the Class Difference Vector $\Delta_\varphi(T|\omega_1, \omega_2)$ as well as the Fractal Signature Mahalanobis Class Distance Vector, $J_\varphi(T|\omega_1, \omega_2)$, between the classes [13]:

$$\begin{aligned} \bar{\varphi}(T|\omega_c) &= \frac{1}{N(\omega_c)} \sum_{n=1}^{N(\omega_c)} \varphi_n(T), \\ \Delta_\varphi(T|\omega_1, \omega_2) &= \bar{\varphi}(T|\omega_1) - \bar{\varphi}(T|\omega_2), \\ J_\varphi(T|\omega_1, \omega_2) &= \frac{(\bar{\varphi}(T|\omega_1) - \bar{\varphi}(T|\omega_2))^2}{(\sigma_\varphi^2(T|\omega_1) + \sigma_\varphi^2(T|\omega_2))/2}. \end{aligned} \quad (4)$$

2.4.2. Adaptive Fractal Signature features

By using the binary decision functions:

$$\begin{aligned} \Omega_{(+)}(T|\Delta_\varphi) &= \begin{cases} 1 & \text{if } \Delta_\varphi(T|\omega_1, \omega_2) \geq 0, \\ 0 & \text{otherwise,} \end{cases} \\ \Omega_{(-)}(T|\Delta_\varphi) &= 1 - \Omega_{(+)}(T|\Delta_\varphi) \end{aligned} \quad (5)$$

the two adaptive fractal signature features for any given image of class ω_1 or ω_2 having a fractal signature vector $\varphi(T)$ are given by the vector summation [13]:

$$\begin{aligned} FS_{(+)} &= \sum_T \varphi(T) \cdot J_\varphi^2(T) \cdot \Omega_{(+)}(T|\Delta_\varphi), \\ FS_{(-)} &= \sum_T \varphi(T) \cdot J_\varphi^2(T) \cdot \Omega_{(-)}(T|\Delta_\varphi). \end{aligned} \quad (6)$$

2.5. Lacunarity Matrices

The fractal dimension and the fractal signature is based on the total length of the line segments that approximate the original gray level curve for each tolerance value. This simple measure does not take into account the distribution of line segment lengths, and may therefore not be sufficient for texture analysis and characterization. One may in fact produce textures having the exact same fractal dimension, but with different textures. Mandelbrot [10] introduced the term lacunarity, Λ , to characterize the deviation from texture homogeneity. In the case of line segments from polygonization, if the distribution of line lengths is narrow, the lacunarity is low, if the distribution is wide, lacunarity is higher. Λ is often estimated as a scalar, e.g., the variance of the distribution [18], for each tolerance value, giving a lacunarity vector. Instead, we have chosen to maintain all the information about the distribution of the length of the line segments. Thus, for each 1D gray level curve, we estimate the probability distribution $p(s)$ of the length, s_i , of the line segments approximating the curve, for each tolerance parameter T [11]. We store these into a Lacunarity Matrix, $\Lambda_n(s, T)$, covering the whole range of the tolerance parameter T [13]. The line lengths are grouped into ranges, 1–10, 11–20, etc.

2.5.1. Lacunarity Class Distance Matrices

From the class-conditional distributions of each element of the lacunarity matrix, $P_\Lambda(s, T|\omega_c)$, we calculate the average Lacunarity Matrix $\bar{\Lambda}(s, T|\omega_c)$ for each class ω_c , the class variance matrix $\sigma_\Lambda^2(s, T|\omega_c)$, the Class Difference Matrix $\Delta_\Lambda(s, T|\omega_1, \omega_2)$ and finally the Lacunarity Mahalanobis Class Distance Ma-

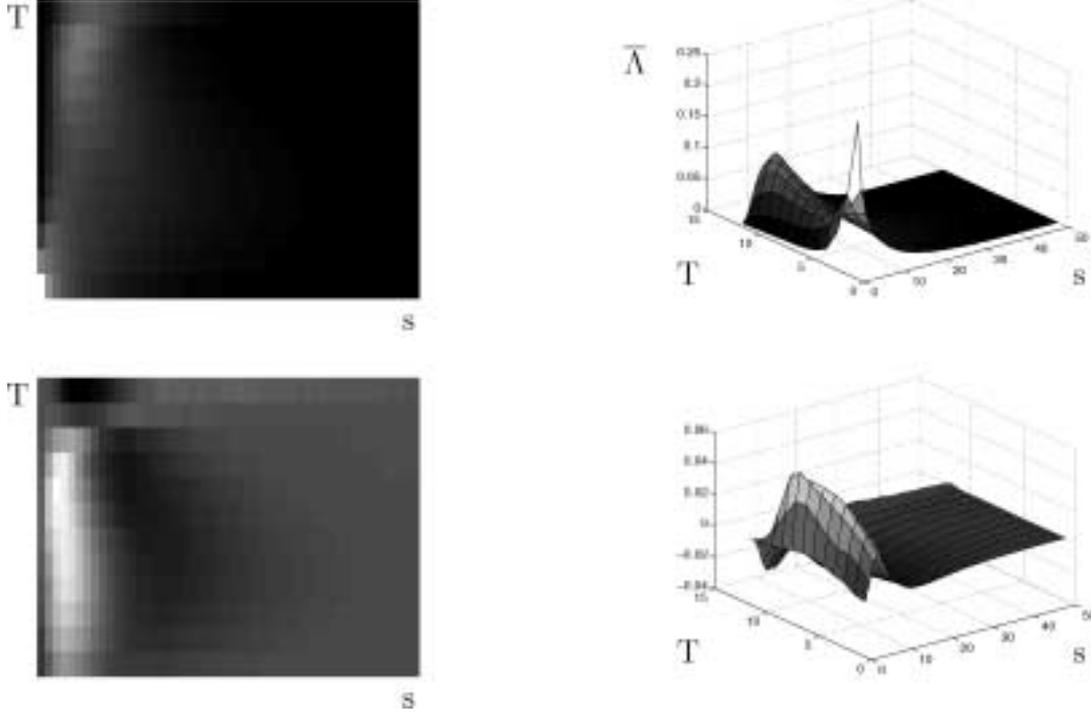


Fig. 6. Upper: The mean lacunarity matrix, $\bar{\Lambda}(s, T|\omega_c)$, computed from the central part of all the monolayer nuclei images in the good prognosis group. Lower: The difference between the $\bar{\Lambda}(s, T|\omega_c)$ computed from the peripheral and central parts of monolayer nuclei images in the good prognosis group. The matrices are visualized both as gray-scale and surface plots.

trix, $J_{\Lambda}(s, T|\omega_1, \omega_2)$, between the two classes ω_1 and ω_2 [13]:

$$\begin{aligned} \bar{\Lambda}(s, T|\omega_c) &= \frac{1}{N(\omega_c)} \sum_{n=1}^{N(\omega_c)} \Lambda_n(s, T), \\ \Delta_{\Lambda}(s, T|\omega_1, \omega_2) &= \bar{\Lambda}(s, T|\omega_1) - \bar{\Lambda}(s, T|\omega_2), \\ J_{\Lambda}(s, T|\omega_1, \omega_2) &= \frac{(\bar{\Lambda}(s, T|\omega_1) - \bar{\Lambda}(s, T|\omega_2))^2}{(\sigma_{\Lambda}^2(s, T|\omega_1) + \sigma_{\Lambda}^2(s, T|\omega_2))/2}. \end{aligned} \quad (7)$$

Figure 6 (upper) shows the mean lacunarity matrix computed from the central part of all the monolayer nuclei images in the good prognosis class. The figure (lower) also shows that there is a marked difference between the lacunarity matrices from the two segments of cell nuclei within the good prognosis class. Figure 7 shows examples of Lacunarity Class Difference and Class Distance Matrices.

2.5.2. Adaptive Lacunarity features

By using the decision matrices (i.e., bit maps):

$$\begin{aligned} \Omega_{(+)}(s, T|\Delta_{\Lambda}) &= \begin{cases} 1 & \text{if } \Delta_{\Lambda}(s, T|\omega_1, \omega_2) \geq 0, \\ 0 & \text{otherwise,} \end{cases} \\ \Omega_{(-)}(s, T|\Delta_{\Lambda}) &= 1 - \Omega_{(+)}(s, T|\Delta_{\Lambda}) \end{aligned} \quad (8)$$

the two adaptive lacunarity matrix features for any given image of class ω_1 or ω_2 having a lacunarity matrix $\Lambda(s, T)$ are given by the matrix summation [13]:

$$\begin{aligned} L_{(+)} &= \sum_{s, T} \Lambda(s, T) \cdot J_{\Lambda}^2(s, T) \cdot \Omega_{(+)}(s, T|\Delta_{\Lambda}), \\ L_{(-)} &= \sum_{s, T} \Lambda(s, T) \cdot J_{\Lambda}^2(s, T) \cdot \Omega_{(-)}(s, T|\Delta_{\Lambda}). \end{aligned} \quad (9)$$

2.6. Classification

Bayesian classification with equal prior probabilities for each class was used as the rule for classification. The feature distribution within each class was assumed to be multivariate normal and the within-class covariance matrices were assumed equal. Initially, the value of each texture feature used to classify each patient of the ovarian data set was the mean value of the distribution of feature values of the nuclei representing the sample (patient). We have verified that the distributions of each of the 6 features obtained from the approximately 250 cell nuclei per patient are very close to normal distributions. The mean feature value and the standard deviation are not only good parameters to

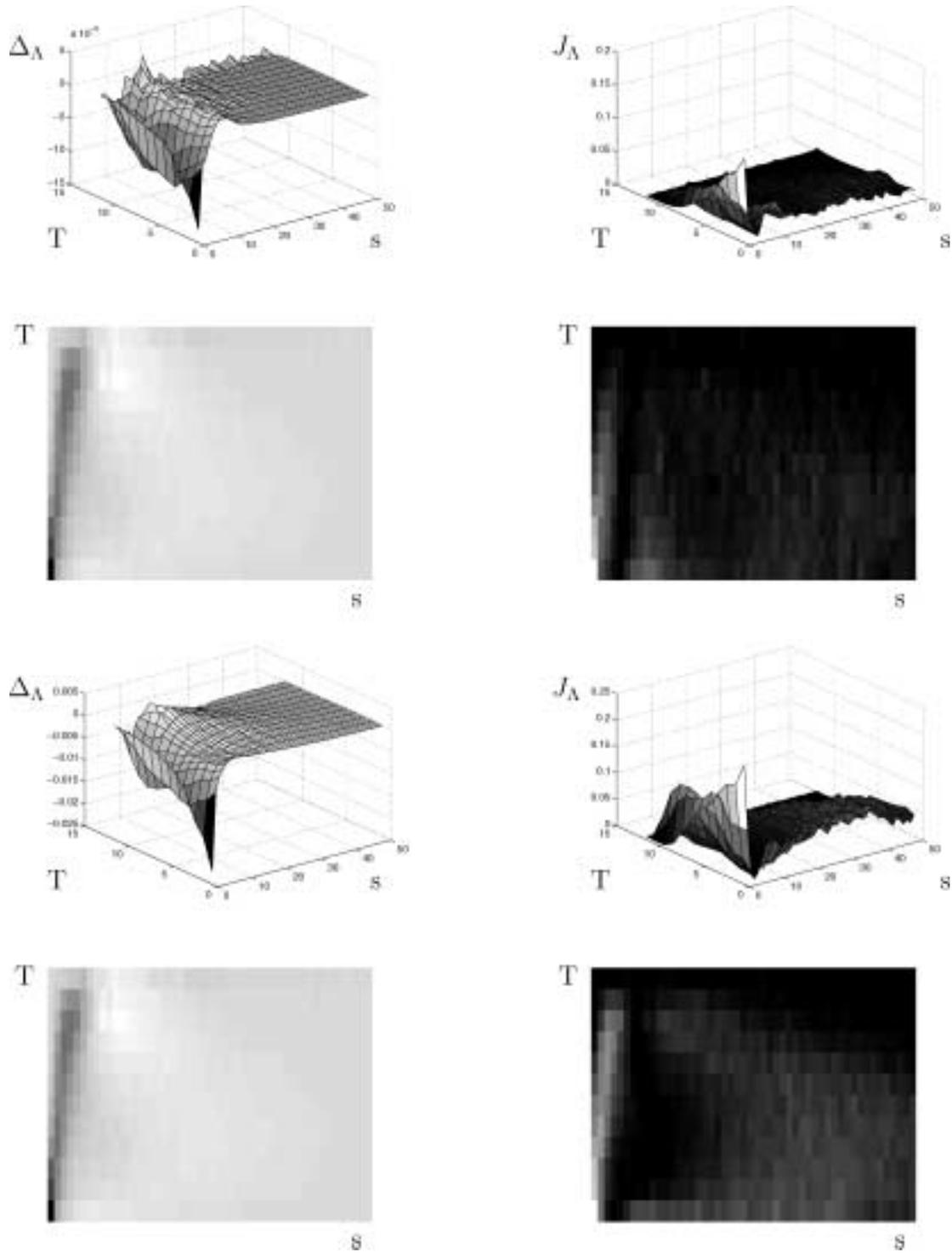


Fig. 7. Left: The Lacunarity Class Difference Matrix, $\Delta_{\Lambda}(s, T|\omega_1, \omega_2)$, computed from the peripheral 30% (upper) and the central 70% (lower) 1D segments of nuclei from monolayers. The positive values in the surface plot (and the lighter areas in the gray-scale plot) correspond to matrix elements that are more probable for the good prognosis class than for the bad prognosis class. Right: The Mahalanobis Class Distance Matrix, $J_{\Lambda}(s, T|\omega_1, \omega_2)$, between the good and bad prognosis classes, based on the lacunarity matrices, computed from the peripheral 30% (upper) and central 70% segments.

describe the statistical distribution of the feature value for a given patient, but may also be natural features to use when classifying the patients. However, this does not mean that these two distribution parameters are the optimal patient features for classification. In a previous study [14], we observed that the class distance was increased by about 40% by using the 10 or 90 percentile of an adaptive feature instead of the median or the mean. Therefore, the standard deviation and the 10 and 90 percentiles of the six feature value distributions for each patient have also been tried.

Because of the small number of ovarian samples available, we have used the leave-one-out method to estimate the misclassification rates. Thus, the same data set was used both to define and evaluate the linear discriminant functions. However, for each cycle of the leave-one-out, new class difference and class distance matrices were obtained.

3. Results

3.1. Monolayers

Figure 5 shows GLCM Class Difference and Class Distance Matrices obtained from the peripheral and central 1D segments of monolayer cell nuclei. The class distance matrices contain large areas of consistently relatively high values. These areas correspond to separate areas of consistently positive or negative class difference values. The mean feature value over all cells for each patient gave a 35–40% classification error for each of the two adaptive GLCM features of Eq. (3), both for the peripheral and central part of the cell nuclei. However, when we consider both the mean, standard deviation and the 10 and 90 percentiles of the feature value distributions, the features from the periphery tend to give lower errors (3%) than the center. As shown in [14], we find that a lower classification error is obtained by using the percentiles. The 90 percentile of the GLCM feature $F_{P(+)}(p)$ from the peripheral part of the nuclei gave an error rate of 27.5%.

The Fractal Signature features (Eq. (6)) gave slightly higher errors than the adaptive GLCM features. Again the peripheral features were better, and the best feature was found among the percentiles.

Figure 7 shows Lacunarity Class Difference and Distance Matrices obtained from the peripheral and central segments, demonstrating that the highest class differences and the highest class distances within each segment are localized into certain parts of the matrices.

As seen in Fig. 6, there is a marked difference between the lacunarity matrices from the two segments of cell nuclei within a given class, and these differences are also localized into consistent areas of the matrix. The classification performance of the adaptive lacunarity features (Eq. (9)) were slightly better than for the adaptive GLCM features. The mean feature value over all cells for each patient gave a 32.5% classification error for each of the single adaptive lacunarity features $L_+(p)$ and $L_-(p)$, and the 90 percentile of $L_+(p)$ gave an error rate of 27.5%. So again we see that the peripheral data have an edge, and that the best feature is found among the percentiles. Searching for the best pair of adaptive features from the same method (either GLCM, Fractal Signature or Lacunarity), the best combination was two lacunarity features, one from the peripheral and one from the central part of the cell nuclei, giving an error rate of 25.0%, as estimated by the leave-one-out method. The best combination of three and four adaptive fractal and adaptive GLCM features would give error rates of 20.0% and 15.0%, respectively. However, this would involve searching through 17.296 and 194.580 possible combinations, quite contrary to our aim of reducing the risk of selecting good feature sets by coincidence.

3.2. Histological sections

The difference in favour of the peripheral features is still present among the fractal signature and lacunarity features. The best single adaptive feature was the 10 percentile of the lacunarity feature $L_-(p)$, giving an error rate of 22.9% (Fig. 8).

The best pair of adaptive features from the same method (either GLCM, Fractal Signature or Lacunarity), again turned out to be a combination of two lacunarity features, the 10 percentile of the lacunarity feature $L_-(p)$ from the peripheral segment and the mean of $L_+(c)$ from the central segment, giving an error rate of 17.1% (Fig. 9).

4. Discussion

The pixels of the 1D signal are assumed to be equidistant, regardless of whether they were 4- or 8-neighbours in the original 2D context. Does this introduce a distortion in the texture measure? There are several aspects to this. It is common practice when using, e.g., GLCM or GLRLM to implicitly use the chessboard distance metric, and thus to mix the information

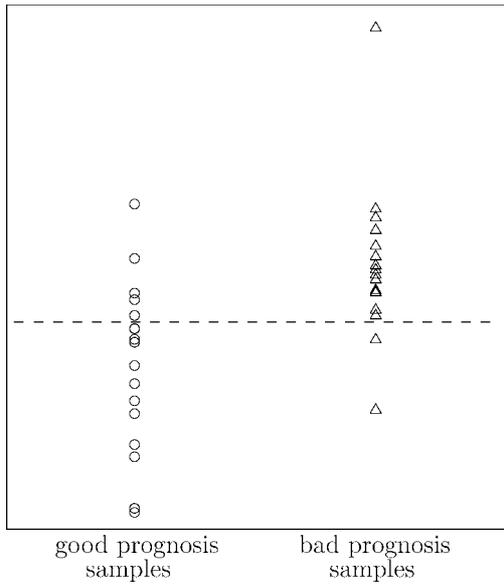


Fig. 8. The adaptive lacunarity feature $L_{(-)}(p)$ extracted from the 30% peripheral part of nuclei from histological sections are plotted for each sample of ovarian cancer. Each sample (patient) is represented by the 10 percentile of the distribution of the feature value extracted from about 280 cell nuclei per sample. The dotted line represents the decision boundary for the classifier computed from the whole data set.

obtained from horizontal/vertical and diagonal steps. Thus, the results of a 2D GLCM or GLRLM analysis and a similar 1D analysis of the peel-off-scanning gray level signal will be the same [12]. We have chosen to represent the results of the peel-off-scanning as consecutive pixel values, without storing the coordinates of the pixels or the Euclidian distances between neighboring pixels. This speeds up the subsequent computations during polygonization. A disadvantage is that some line lengths may be underestimated, thus affecting both the fractal signature and the lacunarity matrix. Our aim is not to obtain exact estimates of the true fractal signature or the true lacunarity, but merely to differentiate between the classes. And there does not seem to be any reason to believe that the underestimated line lengths will systematically affect one class of cells more than the other.

There is a marked difference in the GLCM class difference and distance matrices computed from the peripheral 30% and the central 70% segments of nuclei from monolayers (see Fig. 5). These subtle texture differences are very hard to see in the gray level images.

We note that the peripheral part of a 2D nucleus image from monolayers only reflects the chromatin structure close to the nuclear membrane, while the central

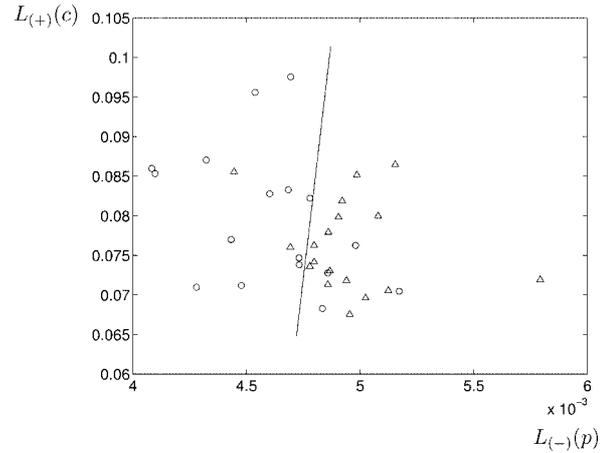


Fig. 9. The mean value of the adaptive lacunarity feature $L_{(+)}(c)$ is plotted versus the 10 percentile of the feature $L_{(-)}(p)$ for samples from the good prognosis (\circ) and bad prognosis (Δ) classes. The decision boundary for the classifier computed from the whole data set is also indicated. The nuclei are from histological sections.

segment will contain information about the chromatin structure throughout the whole nucleus.

This difficulty is not present in the images of histological sections. When working with sections we have observed a non-uniform dying throughout the material. Variable thickness of the sections may be one of the reasons for this. We have observed that some samples have a higher probability of high gray level values in the central part of the nuclei. Although the histograms of all nuclei are normalized, this may influence the location of the peaks of the difference and distance matrices of both the peripheral and central segments. It may also be responsible for the more noisy appearance of the class distance matrices from the histological sections than from the monolayers.

On a larger data set, it would be of interest to study the merits of different ways of normalizing the image histograms prior to the computation of class difference and class distance matrices. As we are performing separate feature extraction from the two parts of the nuclei, it may be best to also perform separate histogram transforms to the two segments. However, given the relatively small data set, it would be futile to test a multitude of combinations of strategies.

In order to understand some of the differences between the two classes expressed in terms of differences in the chromatin structure, it is obviously of interest to analyze the results of the polygonization-based fractal lacunarity method. Therefore, 2D normalized histograms of the line lengths versus distance between breakpoints in the x -direction of the 1D digi-

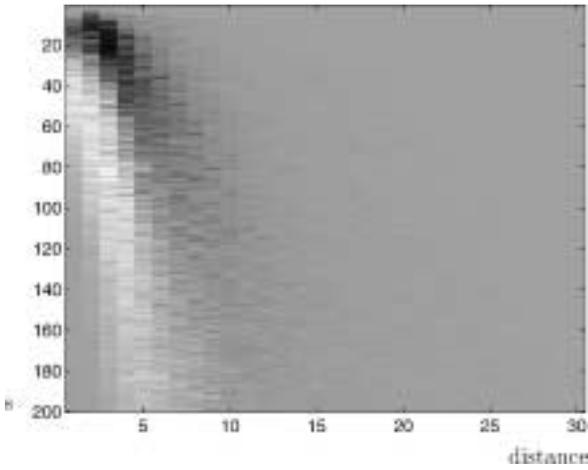


Fig. 10. 2D difference histogram of line length s versus distance in x -direction of the digital curve. This illustrates the difference between the mean normalized histograms extracted from all cells in the good and bad prognosis classes, computed for tolerance $T = 3.25$. The histograms were extracted from the 70% central segment of the monolayer cell nuclei. The lighter area corresponds to line segments that are more probable for the good prognosis class than for the bad prognosis class and the darker area corresponds to line segments that are less probable for the good prognosis class than for the bad prognosis class.

tal curve were generated. Mean histograms were computed over all cells in each class, and difference histograms (Fig. 10) for each polygonization tolerance value were used as a tool to analyze the difference in texture between the classes. We note that line segments of the same length may have widely different intensity gradients. From Fig. 10 one may observe that for a given horizontal distance between breakpoints in the 1D signal, shorter polygonization line segments are more probable in bad prognosis cell nuclei, generally implying lower gray level gradients. This leads directly to the observation that GLCM matrices would be more concentrated along the diagonal in the bad prognosis class than in the good prognosis class, as confirmed by Fig. 5.

Connors and Harlow [5] concluded that the co-occurrence matrix cannot discriminate between all visual texture pairs if only one intersample distance is utilized. This suggests that a number of intersample distances should be used for a more accurate classification. Similarly, the number of gray levels may be varied. Using few levels is equivalent to viewing the image on a coarse scale, whereas more gray levels give an image with more detail. The performance of a given pre-defined GLCM feature, as well as the ranking of the features, may depend on the number of gray levels used. Even in our method, a statistically reliable es-

timate of the GLCM joint probability distribution per cell is needed. So the matrix should contain a reasonably large occupancy level in those locations where the corresponding class distance matrix values are high. This can generally be achieved by restricting the number of gray value quantization levels. Based on previous experience we have chosen $G = 16$ gray levels and the distance $d = 3$ for the GLCM method in order to reduce the total number of features to be evaluated. We note that other (G, d) -combinations were not tested prior to this choice.

Previously we have only used the mean of the cell features obtained for each sample in the classification [2,3,11–13]. We have seen [14] that the use of percentiles from the feature distribution of each sample may improve the classification results. However, using a number of distribution parameters in addition to the mean will obviously increase the actual number of classification features. The reduction in feature space dimensionality accomplished by the new adaptive texture features is therefore an important point. This is particularly valuable when we extract the texture features from two separate segments (center and periphery) for each cell nucleus.

Thus, in the case of GLCM, if we had started out with only the 9 most popular pre-defined GLCM features, having two segments, and using only one (d, G) parameter setting, we would have had a total of 72 possible features when we use the mean, the standard deviation and the two percentiles (10 and 90), and a staggering 2556 possible pairs of features. This in contrast to the 16 adaptive features obtained when we combine 2 matrix features with 2 cell segments and 4 distribution parameters. The number of possible feature pairs is now only 120, reduced by a factor of 21.3, substantially reducing the risk of selecting good feature pairs by pure coincidence.

We use the mean, the standard deviation, or a percentile to represent the measured feature values from an average of $N = 256$ (277) cell nuclei. It is well known that for normally distributed data, the standard error of the estimated mean $\hat{\mu}$ is given by $\sigma_{\hat{\mu}} = \hat{\sigma}/\sqrt{N}$, where $\hat{\sigma}$ is the estimated standard deviation of the N observations. For the standard deviation and the 10/90 percentile, this should be multiplied by a factor $1/\sqrt{2}$ and 1.7094, respectively. In our case, the relative uncertainty of the mean, $\sigma_{\hat{\mu}}/\hat{\mu}$, for the adaptive features used here, is around 0.0125.

If we assume that the contents of each element of, e.g., the GLCM matrix are normally distributed, the same equation for the standard error of the mean ap-

plies here. However, we know that the population of such matrices is not at all even – it is usually concentrated around the main diagonal. So the texture matrix does not stabilize in the same manner as the mean feature value. But that is actually irrelevant, as we put most weight on those matrix elements that exhibit the greatest Mahalanobis distance, and these are also among the ones with the highest population. Thus, the stability of the computed very low dimensionality features is by far more important than the detailed stability of the entire texture matrices.

A possible extension of the adaptive feature extraction for the GLCM based method is to let a parameter value vary within the class distance matrix, as Walker et al. did for the GLCM matrices [20]. We would then have to work with 3- or 4-dimensional class distance matrices, depending on whether we included several values of the inter-pixel distance, d , or several number of gray levels in the image, G , or both. The size of the matrix would naturally taper off as G was decreased. It is important to note that the number of features extracted from each matrix in this way would remain the same, 2, regardless of the dimensionality of the matrix and the range of the parameters. GLCM class difference and distance matrices computed for $d = 5, 7$ and $G = 16$ have shown gradually and consistent variations in the matrices as d varies. Thus, this approach may be useful, as it will include a more complete description of the texture into the two adaptive features, and at the same time even make the statistics more reliable.

In a retrospective study of 260 cases of stage 1 epithelial ovarian cancer, Kildal et al. [8] found that image-based DNA ploidy measurements using Feulgen stained nuclear monolayer preparations gave excellent prognostic information. Almost the entire 10-year mortality in this patient group was within the group of patients whose tumours were judged “aneuploid/polyploid”. By contrast, almost 95% of the patients with “diploid/tetraploid” tumours were still alive after ten years. An interesting next step would be a fusion of the best chromatin texture features with robust quantitative features from image-based DNA ploidy measurements in a combined classifier. It would also be of interest to study details of the relation between texture and ploidy measurements, to further understand the mechanisms of chromatin organization as cancer develops.

In conclusion, the texture of the peripheral part of light microscopy images of ovarian cancer cell nuclei differs from that of the central part of the nuclei. Both

when working with nuclei from histological sections and monolayers it seems to be beneficial to perform a radial differentiation of the texture feature extraction. We have pointed out a procedure based on class difference and class distance matrices that both in the case of GLCM matrices and fractal lacunarity matrices give very low dimensionality adaptive texture feature vectors. The best pair of such features was able to discriminate between the two groups of good and bad prognosis (from histological sections) with an error rate of 17%.

Acknowledgements

We thank Marna Lill Kjørereng, Wenche Hansen and Ruth Puntervold of the Division of Digital Pathology, the Norwegian Radium Hospital, for their skillful technical assistance.

References

- [1] F. Albrechtsen and B. Nielsen, Fractal dimension and lacunarity estimated by sequential 1D polygonization of 2D images, in: *Theory and Applications of Image Analysis II – Selected Papers from the 9th Scandinavian Conference on Image Analysis*, G. Borgefors, ed., World Scientific Publ. Co., Singapore, 1995, pp. 79–88.
- [2] F. Albrechtsen, B. Nielsen and H.E. Danielsen, Adaptive gray level run length features from class distance matrices, in: *Proceedings, the 15th International Conference on Pattern Recognition*, Vol. 3, Barcelona, 2000, pp. 746–749.
- [3] F. Albrechtsen and B. Nielsen, Texture classification based on cooccurrence of gray level run length matrices, invited paper, *Australian Journal of Intelligent Information Processing Systems* **6**(1) (2000), 38–45.
- [4] Y.Q. Chen, M.S. Nixon and D.W. Thomas, Statistical geometric features for texture classification, *Pattern Recognition* **28**(4) (1995), 537–552.
- [5] R.W. Conners and C.A. Harlow, A theoretical comparison of texture algorithms, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **2**(3) (1980), 204–222.
- [6] R.M. Haralick, K. Shanmugam and I. Dinstein, Textural features for image classification, *IEEE Trans. on Systems, Man and Cybernetics* **3** (1973), 610–621.
- [7] D.W. Hedley, DNA analysis from paraffin-embedded blocks, *Methods Cell Biol.* **41** (1994), 231–240.
- [8] W. Kildal, G. Kristensen, V. Abeler, E. Pettersen and H.E. Danielsen, Prognostic significance and comparison between image cytometry and flow cytometry in DNA ploidy distribution in epithelial ovarian cancer stage 1, *Analytical Cellular Pathology* **18** (1999), 31–32.

- [9] G.B. Kristensen, W. Kildal, V.M. Abeler, E.O. Pettersen, J. Kaern, C.G. Tropé, I. Vergote and H.E. Danielsen, High-resolution image analysis – A better method for predicting prognosis in early ovarian cancer, *Gyn. Oncol.* **76**(2) (2000), 236.
- [10] B.B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, New York, 1983.
- [11] B. Nielsen, F. Albrechtsen and H.E. Danielsen, The use of fractal features from the periphery of cell nuclei as a classification tool, *Analytical Cellular Pathology* **19** (1999), 21–37.
- [12] B. Nielsen, F. Albrechtsen, S. Baheerathan and H.E. Danielsen, Peel-off-scanning to obtain radial differentiation of fractal and complexity features in cell nuclei, in: *Proceedings, VISION INTERFACE 2000*, Montreal, Canada, 2000.
- [13] B. Nielsen, F. Albrechtsen and H.E. Danielsen, Using fractal signature vectors and lacunarity class distance matrices to extract new adaptive features from cell nuclei, in: *Fractals in Biology and Medicine*, Vol. 3, Birkhäuser-Verlag, Basel, 2001 (in press).
- [14] B. Nielsen and F. Albrechtsen, Low dimensionality adaptive texture feature vectors from parametric class distance matrices, submitted to: *Pattern Recognition*, 2001.
- [15] W.K. Pratt, *Digital Image Processing*, Wiley-Interscience, New York, 1991.
- [16] H. Schulerud, G.B. Kristensen, K. Liestøl, L. Vlatkovic, A. Reith, F. Albrechtsen and H.E. Danielsen, A review of caveats in statistical nuclear image analysis, *Analytical Cellular Pathology* **16**(2) (1998), 63–82.
- [17] H.J. Tanke and E.M. van Ingen, A reliable Feulgen-acriflavine-SO₂ staining procedure for quantitative DNA measurements, *J. Histochem. Cytochem.* **28** (1980), 1007–1013.
- [18] R. Voss, Random fractals: Characterization and measurement, in: *Scaling Phenomena in Disordered Systems*, R. Pynn and A. Skjeltorp, eds, Plenum Press, New York, 1985, pp. 1–11.
- [19] R.F. Walker and P.T. Jackway, Statistical geometric features – extensions for cytological texture analysis, in: *Proceedings, the 13th International Conference on Pattern Recognition*, Vol. 2, Vienna, Austria, 1996, pp. 790–794.
- [20] R.F. Walker, P.T. Jackway and I.D. Longstaff, Recent developments in the use of the co-occurrence matrix for texture recognition, in: *Proceedings, the 13th International Conference on Digital Signal Processing*, Vol. 1, Santorini, Greece, 1997, pp. 63–65.
- [21] K. Wall and P.E. Danielsson, A fast sequential method for polygonal approximation of digitized curves, *Computer Vision, Graphics, and Image Processing* **28** (1984), 220–227.
- [22] I.T. Young, P.W. Verbeek and B.H. Mayall, Characterization of chromatin distribution in cell nuclei, *Cytometry* **7** (1986), 467–474.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

