The prognostic value of adaptive nuclear texture features from patient gray level entropy matrices in early stage ovarian cancer

Birgitte Nielsen[a,c], Fritz Albregtsen[b,a], Wanja Kildal[a,c], Vera M. Abeler[d], Gunnar B. Kristensen[e,a] and Håvard E. Danielsen[a,b,c,*]

[a]*Institute for Medical Informatics, Oslo University Hospital, Oslo, Norway*
[b]*Department of Informatics, University of Oslo, Oslo, Norway*
[c]*Centre for Cancer Biomedicine, University of Oslo, Oslo, Norway*
[d]*Department of Pathology, Oslo University Hospital, Oslo, Norway*
[e]*Department of Gynecologic Oncology, Oslo University Hospital, Oslo, Norway*

Abstract. *Background:* Nuclear texture analysis gives information about the spatial arrangement of the pixel gray levels in a digitized microscopic nuclear image, providing texture features that may be used as quantitative tools for prognosis of human cancer. The aim of the study was to evaluate the prognostic value of adaptive nuclear texture features in early stage ovarian cancer.

Methods: 246 cases of early stage ovarian cancer were included in the analysis. Isolated nuclei (monolayers) were prepared from 50 μm tissue sections and stained with Feulgen-Schiff. Local gray level entropy was measured within small windows of each nuclear image and stored in gray level entropy matrices. A compact set of adaptive features was computed from these matrices.

Results: Univariate Kaplan-Meier analysis showed significantly better relapse-free survival ($p < 0.001$) for patients with low adaptive feature values compared to patients with high adaptive feature values. The 10-year relapse-free survival was about 78% for patients with low feature values and about 52% for patients with high feature values. Adaptive features were found to be of independent prognostic significance for relapse-free survival in a multivariate analysis.

Conclusion: Adaptive nuclear texture features from entropy matrices contain prognostic information and are of independent prognostic significance for relapse-free survival in early stage ovarian cancer.

Keywords: Adaptive texture features, early stage ovarian cancer, nuclear texture analysis, pattern classification, prognostic marker

1. Introduction

Most women undergoing treatment for early stage ovarian cancer have a favorable prognosis, but about 20% will eventually relapse and 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 digital pathology, the field of nuclear texture analysis gives information about the spatial arrangement of the pixel gray levels in a digitized microscopic nuclear image, providing statistical texture measures

*Corresponding author: Prof. H.E. Danielsen, Institute for Medical Informatics, Radiumhospitalet, Oslo University Hospital, Montebello, Oslo N-0310, Norway. Tel.: +47 22935617; Fax: +47 22935627; E-mail: havard.danielsen@oslo-universitetssykehus.no.

that may be used as quantitative tools for diagnosis and prognosis of human cancer. At present, there are several hundred published peer-reviewed studies in which image texture-based methods have been applied to digital microscopy for screening, diagnosis, and prognosis of human cancer. These studies relate to a wide range of organ systems, encompassing brain, head and neck, breast, pulmonary, gastrointestinal, kidney, prostate, bladder, ovary, and the uterine cervix [15].

In several frequently used texture analysis methods, second or higher order statistics on the relation between pixel gray level values are stored in matrices, e.g., gray-level co-occurrence matrices [9] and gray-level run length matrices [8]. Texture features are then extracted that directly describe the probability distribution within the matrix and indirectly describe the image texture. Each feature may be seen as a weighted sum of the normalized matrix element values. By varying the weighting function, different aspects of the texture can be extracted. These static weighting functions extract information from predefined parts of e.g., the co-occurrence matrix, indifferent to whether or not these parts actually contain useful information for discrimination between the classes. Thus, a large number of static, predefined features are usually extracted from a number of matrices stemming from a combination of free parameters (number of gray levels, inter-pixel distance, orientation).

In order to find nuclear features that discriminate robustly between cases from different diagnostic or prognostic classes, a statistical evaluation of features must be performed, and this requires careful experimental design. Often, a feature selection is performed to obtain the best reduced set of features, either by exhaustively searching through all possible combinations, or following some suboptimal scheme, see e.g., [11]. In [21, 22] we showed that if the number of samples is low, a low number of feature candidates is important in order to select the correct features. In the field of nuclear texture analysis, it is quite common to evaluate a large number of features on a limited learning set of clinical cases, without testing the chosen classifier on an independent validation data set. This easily leads to false or overoptimistic results [15].

Only a few studies are published on diagnostic or prognostic classification of patients with ovarian cancer based on nuclear texture analysis. Deligdisch et al. [4] used the slope of the line corresponding to the declining rate of the autocorrelation as a measure of nuclear texture. Geisler et al. [7], Werness et al. [24],

Protopapa et al. [20] and Brewer et al. [2, 3] used features extracted from co-occurrence and run-length matrices for analyzing nuclear texture. However, none of these studies evaluated their findings on an independent validation set.

Through a series of methodical papers, we have established a unified approach to extracting a compact set of superior features [1, 16–19]. Instead of searching through a large number of possible feature combinations, we have developed methods to incorporate as much class discriminating information as possible into as few features as possible, preferably including adaptivity into the computation of the features. In [17], we found that the new adaptive feature extraction scheme as applied to four relevant texture analysis methods (e.g., co-occurrence and run-length matrices) outperformed the classical static texture features when applied to the most difficult set of 45 Brodatz texture pairs. In a retrospective study including 134 cases of early stage ovarian cancer, a single adaptive feature extracted from either the co-occurrence, run-length, or co-occurrence of run-length matrices [1] discriminated the cases into two prognostic classes with a correct classification rate of 70% [19].

Yogesan et al. [25] and Jørgensen et al. [13] introduced the use of non-adaptive entropy-based textural features as a new prognostic tool in an effort to predict which patients with metastatic prostate cancer that were most likely to respond to hormone treatment. In a recent study we applied adaptive entropy-based textural features for the assessment of dysplasia in Barrett's oesophagus and found that nuclear texture features (Nucleotyping) differentiated dysplastic and non-dysplastic cases with a greater correct classification rate than image cytometric DNA analysis [6]. The aim of the present study has been to apply the methodology of adaptive feature extraction using the gray-level entropy matrix method on a learning set of 134 cases of early stage ovarian cancer, and then to estimate the prognostic value of a single adaptive texture feature on a validation set of 112 cases. This has been done both representing each patient by a mean value of adaptive texture feature values from the 2D entropy matrices of all nuclei within selected area groups, as well as representing the patient by a single 4D entropy matrix per patient, where nuclear area is one of the axes.

For the threshold value of the adaptive texture feature that discriminates best between the two prognostic classes of patients, we have also performed a survival

analysis, both on the learning, validation, and the complete data set.

2. Material and methods

2.1. Cell nuclei from early stage ovarian cancer

This retrospective study was performed on tissue samples from patients treated for early stage ovarian cancer during 1982–1989 [14]. 246 cases of ovarian cancer classified as International Federation of Gynecology and Obstetrics (FIGO) stage I were included in the analysis. A more detailed description of the material is given in [14].

Paraffin-embedded tissue fixed in 4% buffered formalin was used for preparation of nuclei suspension. The tumor tissue was selected by a pathologist [14]. Monolayers (isolated nuclei) were prepared from one or more 50 μm sections using a modification of Hedley's method [10]. The nuclei were Feulgen-Schiff stained according to an established protocol [23].

The Fairfield DNA Ploidy System (Fairfield Imaging Ltd, Nottingham, UK), which consisted of a Zeiss Axioplan microscope equipped with a 40/0.75 objective lens (Zeiss), a 546 nm green filter and a high-resolution digital camera (C4742-95, Hamamatsu Photonics K.K., Hamamatsu, Japan) with 1024×1024 pixels/image and a gray level resolution of 10 bits/pixel was used to capture each image field. Shading correction was performed for each such image field. The pixel resolution was 166 nm/pixel on the cell specimen. Trained personnel performed a screening of the nuclei at the microscope and selected tumor nuclei for the analysis. Stromal nuclei, necrotic nuclei, doublets or cut nuclei were discarded. Each nucleus was segmented from the background using a global threshold. The segmented nuclei were stored in galleries in each case. The mean number of measured tumor nuclei per case was about 300.

2.2. Learning and validation data sets

The data set was randomly divided into a learning set and a validation set. The 134 cases included in the learning data set were grouped into two different prognostic classes, which were used for designing classifiers based on single features. The 94 cases (70%) included in the good prognosis class survived for at least ten years without a relapse, whereas the 40 cases

(30%) included in the poor prognosis class relapsed or died of a cancer-related disease within ten years [19]. The validation set included 105 cases of which 69 cases (66%) could be grouped into the good prognosis class and 36 cases (34%) could be grouped into the poor prognosis class. An additional 7 cases did not relapse during a follow-up period of less than ten years, and could therefore not be defined into the good or the poor prognostic class. However, these were included in the survival analysis.

2.3. 2D gray level entropy matrices

The Gray Level Entropy Matrix (GLEM) element **P(i,j)** contains the estimated probability of a local first order gray level entropy value j within a window of size w × w centered around a pixel with gray level value i [25]. The first order entropy is defined as

$$j = -\sum_{i=1}^{G} P(i) \ln\{P(i)\}, \ P(i) > 0,$$

where **P(i)** is the estimated probability of occurrence of gray level i within the window of size $w \times w$, and G is the number of gray level quantization levels in the image. The entropy measures the gray level uniformity within the window. Homogeneous structures will give low entropy values whereas inhomogeneous structures will give high entropy values.

Yogesan et al. [25] computed GLEM matrices with eight different window sizes between $w \times w = 2 \times 2$ and 15×15, and defined nine texture features based on the GLEM. Each of the nine pre-defined non-adaptive features may be seen as a weighted sum of the normalized gray-level entropy matrix element values, where the weighting applied to each element is based on a given weighting function. By varying the weighting function, different aspects of the texture can be extracted. The weighting functions fall into two general categories:

1. *Weighting based on the value of the GLEM element.* Four features are examples of this: The entropy homogeneity (EH), entropy nonnormality (ENN), roughness (RH), and minimum gray-level entropy (MGE).
2. *Weighting based on the position in the GLEM.* Five features use position-dependent weights: The average entropy (AE), low-entropy emphasis (LEE), and high-entropy emphasis (HEE) use weights that only contain entropy j, while the low

gray-level entropy emphasis (LGEE) and high gray-level entropy emphasis (HGEE) use feature weights depending on both gray-level i and entropy j.

In the present study we have extracted these nine texture features from GLEM matrices computed for $G = 64$ and $w = 9$, resulting in 36 possible pairs of the nine predefined features.

2.4. Adaptive features from 2D gray level entropy matrices

As an alternative to the nine features defined in [25] we extract only two adaptive features from each gray level entropy matrix [18, 6]. The nuclear images were grouped into ten area groups $a = 1, 2, \dots, 10$, according to the number of pixels in the nucleus (where $a = 1$ corresponds to a nuclear area of $1000-1999$ pixels, $a = 2$; nuclear area of $2000-2999$ pixels, $\dots$, $a = 10$; nuclear area $>10\,000$ pixels) [19]. As discussed in [19], there is a clear relation between nuclear DNA content, area, first- order gray level statistics, and texture. We found that nuclei having an area between $2000-4999$ pixels contained most of the class distance information discriminating between the good and poor prognostic classes of patients with early stage ovarian cancer, and by using area groups we avoided problems caused by mixing data from cells having different nuclear area [19]. In the learning set, these three area groups contain 72.4% of all nuclei.

For each cell nucleus from a patient, the number of gray levels in the image was reduced by re-quantization to $G = 64$, and a 2D entropy matrix was computed, using a window size $w = 9$. The choice of number of gray levels and window size is not important for the classification result (see Section 4 for a discussion).

The nuclear images from the *n*-th patient of class ω_c ($c = 1,2$; $\omega_1 =$ good prognosis, $\omega_2 =$ poor prognosis) give a set of average patient matrices $P_n(i, j|a, \omega_c)$, for $a = 1, \dots, 10$. Based on these patient matrices, we then calculate average matrices over all the $N(a, \omega_c)$ learning set patients in each area group of the two classes

$$\bar{P}(i, j|a, \omega_c)$$
$$= \frac{1}{N(a, \omega_c)} \sum_{n=1}^{N(a,\omega_c)} P_n(i, j|a, \omega_c),$$
$$a = 1, \dots, 10; c = 1, 2$$

Based on these matrices, we compute a class difference matrix in each area group (see Fig. 1)

$$\Delta_P(i, j|a, \omega_1, \omega_2)$$
$$= \bar{P}(i, j|a, \omega_1) - \bar{P}(i, j|a, \omega_2), a = 1, \dots, 10$$

as well as a class variance matrix in each area group

$$\sigma_P^2 (i, j|a, \omega_c)$$
$$= \frac{1}{N(a,\omega_c)} \sum_{n=1}^{N(a,\omega_c)} \{P_n(i, j|a, \omega_c) - \bar{P}(i, j|a, \omega_c)\}^2,$$
$$a = 1, \dots, 10; c = 1, 2$$

and finally the Mahalanobis class distance matrices

$$J_P(i, j|a, \omega_1, \omega_2)$$
$$= \left[2 \frac{\{\bar{P}(i, j|a, \omega_1) - \bar{P}(i, j|a, \omega_2)\}^2}{\sigma_P^2 (i, j|a, \omega_1) + \sigma_P^2 (i, j|a, \omega_2)} \right]^{1/2},$$
$$a = 1, \dots, 10$$

In the present study, the computation of class average, variance, difference and distance matrices were computed from 2D patient matrices, whereas in [16, 17, 19] the matrices were computed from the entropy matrices of all the cell nuclei of the learning cases in each class.

For each nuclear image, we extract two adaptive features from the 2D entropy matrix by using the 2D Mahalanobis class distance as weights, and the disjoint positive/negative parts of the 2D class difference matrix as the domains of the weighted summations. Class distance and difference matrices were selected according to the nuclear area (area group a) of each nucleus.

$$AF_+ = \sum_{\Delta_P(i,j|a,w,\omega_1,\omega_2)>0} P(i, j|a)[J_P(i, j|a, \omega_1, \omega_2)]$$

$$AF_- = \sum_{\Delta_P(i,j|a,w,\omega_1,\omega_2)<0} P(i, j|a)[J_P(i, j|a, \omega_1, \omega_2)]$$

Each case (patient) is represented by the (scalar) mean value of the adaptive texture feature values extracted from the 2D matrices of all nuclei within area groups $a = 2, 3, 4$ [19]. Matrices from $a = 1$ contained almost no class distance information, while matrices from $a = 5, \dots, 10$ were based on a decreasing number of nuclei, resulting in more noise, and a risk of very uncertain or even missing feature values for some patients.
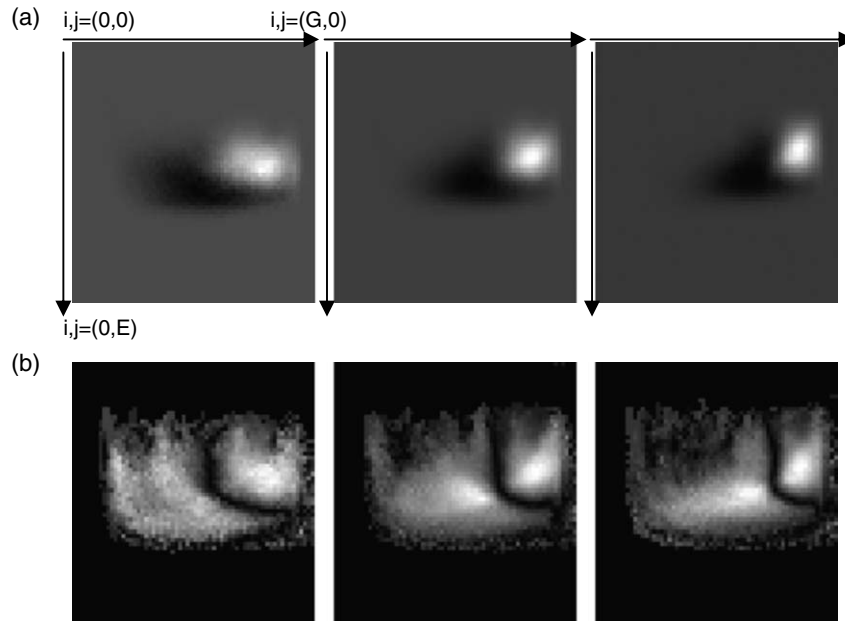
Fig. 1. (a) Gray level entropy matrix (GLEM) class difference matrices $\Delta_P(i, j|a, \omega_1, \omega_2)$ between the two prognostic classes of early ovarian cancer, computed from the learning set nuclear images within area groups a = 2–4. For each gray-scale plot, areas that are lighter (darker) than the background correspond to GLEM matrix elements that contain a higher (lower) probability for the good prognosis class compared to the poor prognosis class. (b) GLEM Mahalanobis class distance matrices $J_P(i, j|a, \omega_1, \omega_2)$.

2.5. Adaptive features from 4D patient matrices

For each nucleus representing a given patient, a 2D entropy matrix is computed, and a 3D patient matrix $P(i,j,a|w)$ using the nuclear area group a as a third axis is accumulated. The 3D patient matrix is normalized by dividing each element by the number of nuclei representing the patient. Such 3D patient matrices are then computed for several different window sizes, w = 3, 5,.., 31, and a 4D patient matrix $P(i,j,a,w)$ is obtained by concatenating the 3D matrices computed for the 15 different window sizes.

Assuming that the *n*-th patient of class ω_c gives a 4D patient matrix $\mathbf{P_n}(\boldsymbol{i,j,a,w}|\omega_\mathbf{c})$ we then calculate an average matrix over all the learning set patients in each class ω_c, as well as two 4D class variance matrices. Based on these matrices, we compute a 4D class difference matrix and a 4D Mahalanobis class distance matrix in the same manner as described in the previous subsection. Finally, we extract two adaptive features from each 4D patient matrix by using the 4D Mahalanobis class distances as weights, and the disjoint positive/negative parts of the 4D class difference matrix as the domains of the weighted summation. Here the adaptive features are extracted from a single 4D matrix per patient, where nuclei from all area groups are included, and nuclear area is one of the axes. This contrasts the use of one 2D entropy matrix per area group for three selected area groups in the previous subsection.

2.6. Classification of each case

In order to evaluate the prognostic value of each adaptive feature, minimum Euclidean distance classifiers [5, page 39] based on single features were constructed to classify each case (patient) into the good or poor prognosis classes. The classifiers were designed on the learning data set, and the "best" classifier with the highest correct classification rate (CCR) was applied on the independent validation data set. The classification results were then used as input to survival analysis.

For comparison, we have extracted the nine texture features defined in [25] from GLEM matrices computed for G = 64 and w = 9, resulting in 36 possible pairs of predefined features. Linear discriminant analysis was performed for each feature combination and a minimum Euclidean distance classifier based on the resulting discriminant function was used to classify

each patient. The learning set was split into two parts for training (101 cases) and testing (33 cases).

2.7. Survival analysis

The SPSS statistical package (SPSS Statistics 18) was used for survival analysis. Survival of patients was estimated using univariate Kaplan-Meier analysis. Relapse-free survival was calculated from start of treatment to relapse or the end of the study period (31.12.1998). The log-rank test was used for test of equality of survival distributions for the different levels of each feature. The Cox proportional hazards regression model with backward stepwise variable selection, and a conditional algorithm for variable removal, was used for multivariable analysis.

3. Results

3.1. Results on the learning set

GLEM class difference matrices showed that the tumor nuclei from patients with good prognosis have a higher probability of higher gray level values and lower local entropy values compared to nuclei from patients with poor prognosis (Fig. 1a). Nuclei from patients with poor prognosis have lower gray level values, while the local entropy is higher, implying that the gray level variability within the local windows is higher in poor prognosis patients. The Mahalanobis class distance matrices $J_p(i,j|a)$ shown in Fig. 1b demonstrate that the weight functions used in the computation of the features adapt to the differences in gray level and entropy values described above.

The CCR of the best adaptive feature extracted from either the ordinary 2D cell nuclear entropy matrices or from the 4D patient matrices are given in Table 1.

A boxplots of the adaptive feature, AF_-, computed from the 2D cell entropy matrices is shown in Fig. 2a, and Kaplan-Meier curves based on the same feature are shown Fig. 3a. The *P*-value estimated by the log-rank test was <0.001.

3.2. Evaluation on the independent validation set

The CCR of the adaptive features are given in Table 1, and a boxplot of the adaptive feature, AF_-, is shown in Fig. 2b.
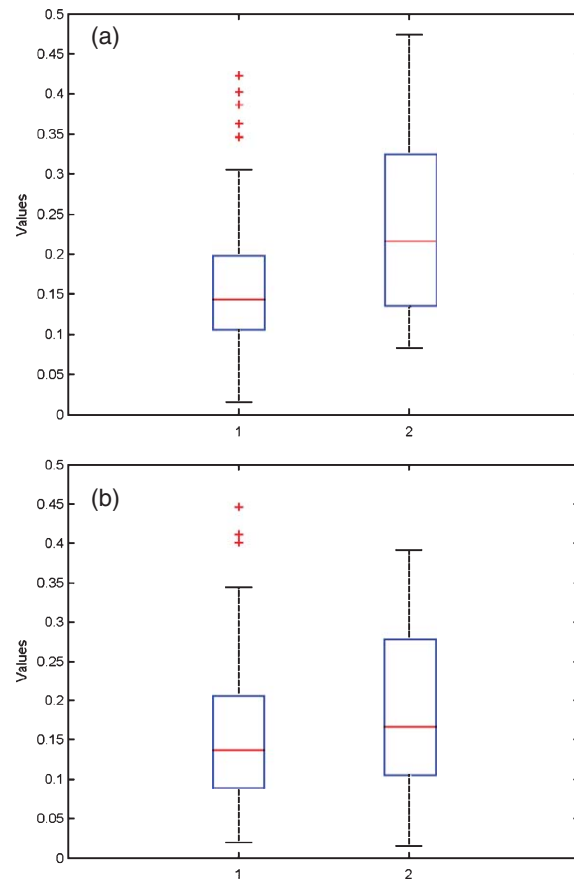

Fig. 2. Boxplot of the feature distribution of the adaptive feature AF_- for good (1) and poor (2) prognosis cases. Boxplots are shown for (a) the learning set and (b) the validation set. Outliers are plotted individually (+). (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/ACP-2012-0065)

Univariate Kaplan-Meier analysis showed significantly better relapse-free survival for patients with low adaptive feature values compared to patients with high adaptive feature values on the basis of a single adaptive entropy matrix feature value (Fig. 3). *P* values estimated by the log-rank test for survival curves estimated from the validation set and from the complete data set were 0.023 and <0.001, respectively. The 10-years relapse-free survival was about 78% for patients with low adaptive feature value and about 52% for patients with high adaptive feature value in our 2D entropy matrix approach, and did not change significantly when we used the 4D patient matrix (Table 2).

In an earlier study based on the same clinical material, DNA ploidy classification, histological grade (including clear cell tumors in the group of poorly differentiated tumors), and FIGO stage were found to

Table 1

The learning and validation set correct classification rate (CCR), sensitivity and specificity for the "best" single feature extracted from either 2D entropy matrices or 4D patient matrices

Feature	CCR (in %) learning/validation	Sensitivity (in %) learning/validation	Specificity (in %) learning/validation
2D:			
$AF_{-(A2A4)}$	70.2/62.9	57.5/41.7	75.5/73.9
4D:			
$AF_{-(4D)}$	69.4/65.7	60.0/47.2	73.4/75.4

Table 2

Relapse-free survival related to histological grade (including clear cell tumors in the group of poorly differentiated tumors), FIGO stage, and texture analysis by the adaptive features $AF_{-(A2A4)}$ and $AF_{-(4D)}$ in stage I ovarian cancer. The complete data set (246 cases) was included in the analysis

Feature	n	10-year relapse-free survival	Hazard ratio (95% CI)	P
Degree of differentiation:				
Grade 1–2	142	86.5	1.0	
Grade 3/clear cell	104	47.9	5.1 (3.1–8.5)	<0.0001
FIGO stage:				
Ia	86	82.5	1.0	
Ib-c	160	62.9	2.6 (1.5–4.6)	0.001
Texture:				
Low $AF_{-(A2A4)}$	167	78.3	1.0	
High $AF_{-(A2A4)}$	79	51.9	2.7 (1.7–4.2)	<0.0001
Low $AF_{-(4D)}$	163	79.6	1.0	
High $AF_{-(4D)}$	103	50.6	2.9 (1.9–4.6)	<0.0001

be of independent prognostic significance for relapse-free survival in multivariate analysis [14]. Histological grade and FIGO stage were included in the present study for comparison (Table 2). In a multivariate analysis including histological grade, FIGO stage, and the adaptive feature AF_-, the adaptive feature was found to be of independent prognostic significance for relapse-free survival (Table 3). Histological grade was the strongest predictor of survival in the multivariate analysis followed by FIGO stage and the adaptive feature. A similar analysis based on the 4D patient matrix features gave almost exactly the same results.
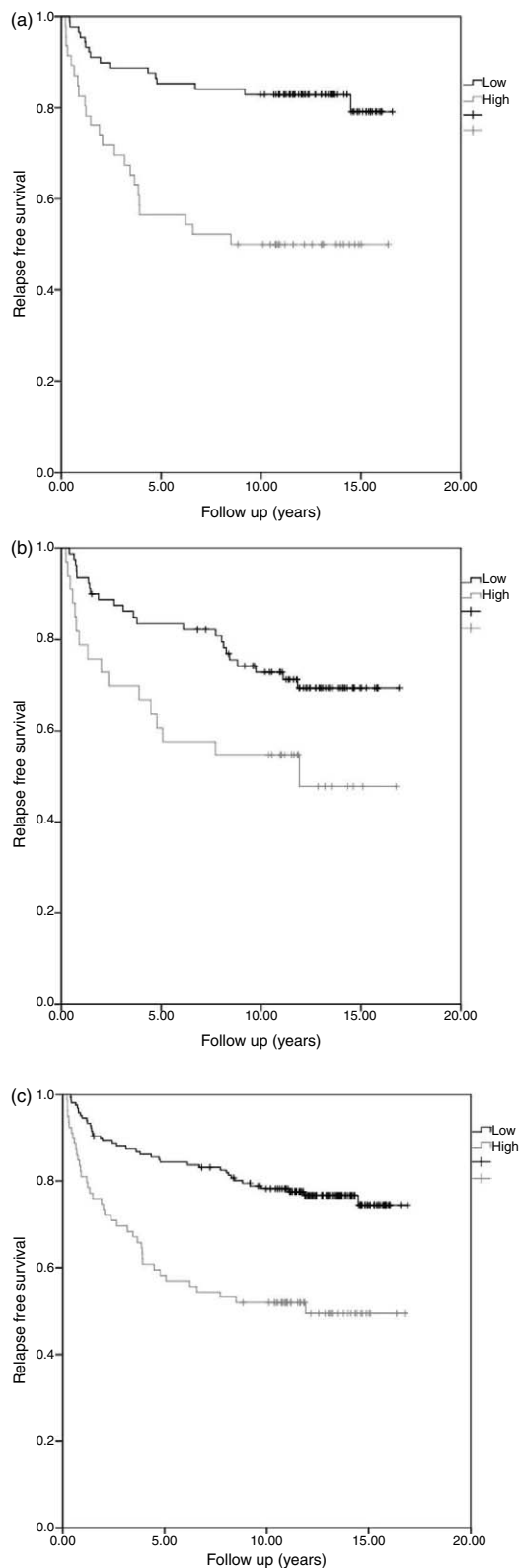
3.3. The non-adaptive GLEM features

For $G = 64$ and $w = 9$, there are 36 possible pairs of the 9 non-adaptive GLEM features defined by [25]. Four of these gave a training CCR >70%, but the test CCR dropped by an average of 10.5% for these four pairs (to an average of 60%), compared to an average drop of 2.1% for all 36 pairs. This may be linked to

Table 3

Significant variables for survival in multivariate analysis including histological grade (including clear cell tumors in the group of poorly differentiated tumors), FIGO stage, and the adaptive texture feature $AF_{-(A2A4)}$

Characteristic	Hazard ratio (95% CI)	P
Degree of differentiation:		
Grade 1–2	1.0	
Grade 3/clear cell	4.2 (2.5–7.1)	<0.0001
FIGO stage:		
Ia	1.0	
Ib-c	2.3 (1.3–4.0)	0.002
Texture:		
Low $AF_{-(A2A4)}$	1.0	
High $AF_{-(A2A4)}$	1.8 (1.1–2.8)	0.015

the fact that all four pairs included the rather volatile feature MGE, which may be based on the gray level entropy within a single window around the minimum intensity in the image. Apart from the MGE feature, there are 6 feature pairs giving a training CCR above 65% and a test CCR barely above 60%.

Fig. 3. Kaplan-Meier survival curves based on the adaptive feature AF₋ computed from the 2D entropy matrices. Survival curves are shown for (a) the learning set, (b) the validation set and (c) the complete data set.

4. Discussion

Texture features used for classification in digital pathology are often selected from a large number of predefined and static features. Both the gray-level co-occurrence matrix method [9] and the gray-level run length method [8] are good examples of texture methods with a relatively large number of predefined features. Identifying a few consistently valuable features is important for many applications as it improves classification reliability, particularly when working with small data sets [12], but it also enhances our understanding of the phenomena that we are modeling. Therefore, through a series of methodical papers [1, 16, 17, 19], we have established a unified approach to extract a compact set of adaptive features.

Instead of extracting a relatively large set of features from a given probability matrix, using a set of predefined weighting functions (for example from the co-occurrence matrix) and subsequently perform a feature selection to obtain the best reduced set of features, we have suggested

1) To use weighting functions that are based on a class distance matrix and a class difference matrix. Thus, the weighting functions adapt to the image material under study, giving features that extract information from the parts of the matrix that actually contain information about differences between two different clinical classes.
2) To extract only a low dimensional feature set, thus avoiding feature selection, and thereby the risk of selecting the wrong features, based on a relatively small number of samples per feature candidate and class [21, 22].

In [17, 19] we applied this approach to four relevant texture analysis methods. For each of the texture analysis methods, we found that one adaptive feature contained most of the discriminatory power of the method [17]. We found that class difference and distance matrices clearly illustrated the difference in texture between the cell nucleus images from the two different prognostic classes of early stage ovarian cancer. We also found that tumor nuclei having an area

between 2000–4999 pixels contained most of the class distance information separating the prognostic classes [19].

In the present study, we have extracted adaptive features from patient gray level entropy matrices and found that one such feature classified the learning cases into good and poor prognosis with a correct classification rate of about 70%. This result is similar to the classification results obtained extracting one adaptive feature from other relevant texture matrices (e.g., co-occurrence matrices and run-length matrices) [19]. Both the correct classification rate of about 70% and the classifications of each case are similar using one adaptive feature from either co-occurrence, run length, or entropy matrices. So it seems that each of these methods capture the changes in chromatin structure that contains prognostic information.

In the present study, we have chosen to use $G = 64$ and a window size of 9×9. As seen in Fig. 1, the class difference and class distance matrices do not contain classification-relevant fine structures that would be significantly affected if the number of gray levels were further reduced. This supports the choices made in earlier studies [1, 16, 17, 19], where we reduced the number of gray levels to only 16 before extracting nuclear adaptive texture features. Experiments have shown that the number of gray levels (16, 32, . . . 1024) is not important for the prognostic classification of the ovarian cancer cases based on one adaptive feature from either co-occurrence, run-length, or entropy matrices.

As an additional experiment, we have computed class distance and difference matrices with several different window sizes, i.e. $w = 3, 5, . . . , 31$. The adaptive feature AF_- computed with different window sizes gave similar correct classification rates and classifications of each case in the learning set (e.g., $w = 3$ and $w = 31$ gave a correct classification rate of 69.4% and 70.2%, respectively).

We have applied the methodology of adaptive feature extraction using both 2D entropy matrices to extract features from each nuclear image to obtain an average for nuclei within certain area limits, and using 4D entropy matrices that include all nuclear areas and then extracting just the two features per patient. The results from the validation data set shows that the 4D patient matrix approach improves the single-feature CCR by 2.8% (to 65.7%), while the sensitivity and specificity are increased by 5.5% and 1.5% (to 47.2 and 75.4%), respectively.

Dividing the 134 patient learning set into a training (101 cases) and a test set (33 cases), we have concluded that the best pair of non-adaptive entropy matrix features obtained a test CCR of about 60%, and was outperformed by the best single adaptive feature. The non-adaptive features were therefore not applied on the final validation data set.

As discussed above, using our adaptive feature extraction, neither the choice of texture method (co-occurrence, run-length, entropy matrix), the number of gray-level quantization levels G, or the window size w in the computation of the entropy matrices, were important for the classification of each case of early stage ovarian cancer. In the present study, we therefore have chosen to evaluate one adaptive entropy matrix feature computed with $G = 64$ and $w = 9$ on an independent validation set of 112 cases of early stage ovarian cancer. Univariate Kaplan-Meier analysis showed significant better relapse-free survival for patients classified as "good" prognosis (with low adaptive feature values) compared to patients classified as "poor" prognosis (with high adaptive feature values) on the basis of a single adaptive entropy matrix feature.

In a multivariate analysis including DNA ploidy classification, histological grade, FIGO stage, and the adaptive feature AF_-, the adaptive feature was removed. However, in a multivariate analysis including histological grade, FIGO stage, and the adaptive feature AF_-, the adaptive feature was found to be of independent prognostic significance for relapse-free survival. In the present study, DNA ploidy classification and nuclear texture analysis were based on the same high-resolution digital images. However, DNA ploidy classification is not a routine method that is implemented at every lab. Furthermore, our main goal is to perform nuclear texture analysis on histological sections, where DNA ploidy classification is not an option. In our department, we are now performing high-throughput nuclear texture analysis in order to perform a prognostic classification of cancer patients, based on up to 50.000 measured nuclei/case.

In conclusion, we have demonstrated that adaptive nuclear texture features contain prognostic information and are of independent prognostic significance for relapse-free survival in early stage ovarian cancer. Univariate Kaplan-Meier analysis showed significant better relapse-free survival for patients with low adaptive feature values compared to patients with high adaptive feature values on the basis of a single adaptive entropy matrix feature. The 10-years relapse-free

survival was about 78% for patients with low adaptive feature values and about 52% for patients with high adaptive feature values. Adaptive nuclear texture analysis is a promising high-resolution, high-throughput method for assessing large scale genomic instability that may aid pathologists in prognosis of cancer.

Acknowledgments

References

[1] F. Albregtsen and B. Nielsen, Texture classification based on cooccurrence of gray level run length matrices, *Australian Journal of Intelligent Information Processing Systems* **6** (2000), 38–45.

[2] M.A. Brewer, J. Ranger-Moore, M.H. Greene, D.S. Alberts, Y. Liu, H.G. Bartels, A.C. Baruch and P.H. Bartels, Preneoplastic changes in ovarian tissues, *Anal Quant Cytol Histol* **26** (2004), 207–216.

[3] M.A. Brewer, J. Ranger-Moore, A. Baruche, D.S. Alberts, M. Greene, D. Thompson, Y. Liu, J. Davis and P.H. Bartels, Exploratory study of ovarian intraepithelial neoplasia, *Cancer Epidemiol Biomarkers Prev* **14** (2005), 299–305.

[4] L. Deligdisch, J. Gil, H. Kerner, H.S. Wu, D. Beck and R. Gershoni-Baruch, Ovarian dysplasia in prophylactic oophorectomy specimen – Cytogenetic and morphometric correlations, *Cancer* **86** (1999), 1544–1550.

[5] R.O. Duda, P.E. Hart and D. Stork, Pattern Classification, 2nd edition, Wiley-Interscience: New York, 2001, p. 39.

[6] J.M. Dunn, T. Hveem, M. Pretorius, D. Oukrif, B. Nielsen, F. Albregtsen, L.B. Lovat, M.R. Novelli and H.E. Danielsen, Comparison of nuclear texture and image cytometric DNA analysis for the assessment of dysplasia in Barrett's oesophagus, *British Journal of Cancer* **105** (2011), 1218–1223.

[7] J.P. Geisler, H.E. Geisler, G.A. Miller, M.C. Wiemann, Z. Zhou and W. Crabtree, Markov optical texture parameters as prognostic indicators in ovarian carcinoma, *Int J Gynecol Cancer* **9** (1999), 317–321.

[8] M.M. Galloway, Texture analysis using gray level run lengths, *CVGIP* **4** (1975), 172–179.

[9] R.M. Haralick, K. Shanmugam and I. Dinstein, Textural features for image classification, *IEEE Trans. on Systems Man and Cybernetics* **3** (1973), 610–621.

[10] D.W. Hedley, DNA analysis from paraffin-embedded blocks, *Methods Cell Biol* **41** (1994), 231–240.

[11] A. Jain and D. Zongker, Feature selection: Evaluation, application, and small sample performance, *IEEE Trans. on Pattern Analysis and Machine Intell* **19** (1997), 153–158.

[12] A.K. Jain and B. Chandrasekaran, Dimensionality and sample size considerations in pattern recognition practice, In: Handbook of statistics, vol. 2, P.R. Krishnaiah, L.N. Kanal, eds., North-Holland Publishing: Amsterdam 1982, pp. 835–855.

[13] T. Jørgensen, K. Yogesan, K.J. Tveter, F. Skjørten and H.E. Danielsen, Nuclear texture analysis: A new prognostic tool in metastatic prostate cancer, *Cytometry* **24** (1996), 277–283.

[14] G.B. Kristensen, W. Kildal, V.M. Abeler, J. Kaern, C.G. Trope and H.E. Danielsen, Large-scale genomic instability predicts long-term outcome for women with invasive stage I ovarian cancer, *Annals of Oncology* **14** (2003), 1494–1500.

[15] B. Nielsen, F. Albregtsen and H.E. Danielsen, Statistical nuclear texture analysis in cancer research: A review of methods and applications, *Critical Reviews in Oncogenesis* **14** (2008), 89–164.

[16] B. Nielsen, F. Albregtsen, W. Kildal and H.E. Danielsen, Prognostic classification of early ovarian cancer based on very low dimensionality adaptive texture feature vectors from cell nuclei from monolayers and histological sections, *Analytical Cellular Pathology* **23** (2001), 75–88.

[17] B. Nielsen, F. Albregtsen and H.E. Danielsen, Low dimensional adaptive texture feature vectors from class distance and class difference matrices, *IEEE Trans. on Medical Imaging* **23**(1) (2004), 73–84.

[18] B. Nielsen, F. Albregtsen and H.E. Danielsen, Nuclear texture analysis of two different prognostic classes of early ovarian cancer based on gray level entropy matrices, *Analytical Cellular Pathology* **25** (2003), D.17.

[19] B. Nielsen and H.E. Danielsen, Prognostic value of adaptive textural features – The effect of standardizing nuclear first-order statistics and mixing information from nuclei having different area, *Cellular Oncology* **28** (2006), 85–95.

[20] E. Protopapa, G. Delides, G. Miaoulis, D. Thompson and P.H. Bartels, Image analysis of mesoyhelioma. II. Discrimination of mesothelioma from metastatic serous ovarian adenocarcinoma, *Anal Quant Cytol Histol* **22** (2000), 338–345.

[21] H. Schulerud and F. Albregtsen, Effects of many feature candidates in feature selection and classification, *Lecture Notes Comput Sci* **2396** (2002), 480–487.

[22] H. Schulerud and F. Albregtsen, Many are called but few are chosen. Feature selection and error estimation in high dimensional spaces, *Comput Methods Programs Biomed* **73**(2) (2004), 91–99.

[23] H.J. Tanke and E.M. van Ingen, A reliable Feulgen-acriflavine-SO_2 staining procedure for quantitative DNA measurements, *J Histochem Cytochem* **28** (1980), 1007–1013.

[24] B.A. Werness, A.M. Afify, K.L. Bielat, G.H. Eltabbakh, M.S. Piver and J.M. Paterson, Altered surface and cyst epithelium of ovaries removed prophylactically from women with a family history of ovarian cancer, *Human Pathol* **30** (1999), 151–157.

[25] K. Yogesan, T. Jørgensen, F. Albregtsen, K.J. Tveter and H.E. Danielsen, Entropy-based texture analysis of chromatin structure in advanced prostate cancer, *Cytometry* **24** (1996), 268–276.