Cytometry
PART A
Journal of the International Society for Advancement of Cytometry

Entropy-Based Adaptive Nuclear Texture Features are Independent Prognostic Markers in a Total Population of Uterine Sarcomas

Birgitte Nielsen,[1,2] Tarjei Sveinsgjerd Hveem,[1,2,5] Wanja Kildal,[1,2] Vera M. Abeler,[3] Gunnar B. Kristensen,[1,4] Fritz Albregtsen,[1,5] Håvard E. Danielsen[1,2,5,6]*

[1]Institute for Cancer Genetics and Informatics, Oslo University Hospital, Oslo, Norway

[2]Centre for Cancer Biomedicine, University of Oslo, Oslo, Norway

[3]Department of Pathology, Oslo University Hospital, Oslo, Norway

[4]Department of Gynecologic Oncology, Oslo University Hospital, Oslo, Norway

[5]Department of Informatics, University of Oslo, Oslo, Norway

[6]Nuffield Division of Clinical Laboratory Sciences, University of Oxford, England

Received 9 May 2014; Revised 16 October 2014; Accepted 18 November 2014

Correspondence to: H.E. Danielsen; Institute for Cancer Genetics and Informatics, Oslo University Hospital, P.O. Box 4953 Nydalen, 0424 Oslo, Norway. E-mail: hed@ous-hf.no

Published online 00 Month 2014 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.22601

Abstract

Nuclear texture analysis measures the spatial arrangement of the pixel gray levels in a digitized microscopic nuclear image and is a promising quantitative tool for prognosis of cancer. The aim of this study was to evaluate the prognostic value of entropy-based adaptive nuclear texture features in a total population of 354 uterine sarcomas. 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, and two superior adaptive texture features were calculated from each matrix. The 5-year crude survival was significantly higher ($P < 0.001$) for patients with high texture feature values (72%) than for patients with low feature values (36%). When combining DNA ploidy classification (diploid/nondiploid) and texture (high/low feature value), the patients could be stratified into three risk groups with 5-year crude survival of 77, 57, and 34% (Hazard Ratios (HR) of 1, 2.3, and 4.1, $P < 0.001$). Entropy-based adaptive nuclear texture was an independent prognostic marker for crude survival in multivariate analysis including relevant clinicopathological features (HR = 2.1, $P = 0.001$), and should therefore be considered as a potential prognostic marker in uterine sarcomas. © 2014 International Society for Advancement of Cytometry

Key terms

uterine sarcomas; nuclear texture analysis; entropy; adaptive features; prognostic markers; leiomyosarcoma; endometrial stromal sarcoma; nucleotyping

Uterine sarcomas are rare tumors, comprising about 3% of uterine cancers (1–4) and 7% of all soft tissue sarcomas reported (5). Most uterine sarcomas are aggressive tumors, but there are marked differences in survival between the histological types (1,2). However, the histological diagnosis of uterine sarcomas is challenging (2) and because of the rarity of the disease, the experience in diagnosing these tumors is limited for most pathologists. Most studies on prognostic factors in uterine sarcomas include a small number of cases and results are conflicting. There is currently no consensus on the significance of various prognostic factors (1,6) and optimal treatment (6). Tumor stage and mitotic index (MI) are reported to be important prognostic factors in uterine sarcomas (1,2). However, information on the use of preoperative imaging for staging purposes is lacking, and therefore uterine sarcomas are still surgically staged (1). Calculation of MI is a labor-intensive method based on subjective manual classification of nuclei into cell cycle stages and has been shown to have a low reproducibility (7,8). Therefore, there is a need for new reliable prognostic factors in uterine sarcomas (1).

In digital pathology, the field of nuclear texture analysis (nucleotyping) gives information about the spatial arrangement of the pixel gray levels in a digitized

ISAC
International Society for Advancement of Cytometry

microscopic nuclear image, and it is well documented that such analysis is showing promising results as a novel diagnostic and/or prognostic marker (9). The aim of the current study was to evaluate the prognostic value of adaptive nuclear texture features based on local gray level entropy (10) in a series including all uterine sarcomas in Norway during the period 1970–2000 (2,11). To our knowledge there are no other publications on nuclear texture analysis in uterine sarcomas.

In the field of nuclear texture analysis, it is common to evaluate a large number of features on a limited training set of clinical cases, without testing the chosen classifier on an independent validation set. This easily leads to false or overoptimistic results (9,12). Identifying a few consistently valuable features is important as it improves classification reliability and enhances our understanding of what we are measuring. Most published work in this field is based on features developed from the pixel gray level co-occurrence matrix (GLCM) (13) and from the gray level run length matrix (GLRLM) (14), and on fractal texture methods (9,15). In the GLCM and GLRLM texture analysis methods, statistics on the relation between the pixel gray level values in digital images are stored in matrices (9). Predefined, nonadaptive texture features that directly describe the probability distribution within the matrix are extracted and they indirectly describe the image texture. Each feature may be seen as a weighted sum of the probability matrix elements, and 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 GLCM, indifferent to whether or not these parts actually contain useful information for discriminating between two clinical classes.

Walker et al. (16) proposed several methods for computation of adaptive features (AF) based on a multi-scale co-occurrence matrix and an associated discrimination matrix. Through a series of methodical articles (9,17,18), our group has established a unified approach to extracting only two superior adaptive texture features from matrices. This AF extraction is based on a Mahalanobis class distance matrix and a class difference matrix, which are computed from the training set images of two clinical classes and define the two adaptive weighting functions. The AF extract information from the parts of the matrix that actually contain information about the texture differences between the two clinical classes. In Ref. (17), we found that the new AF extraction scheme as applied to four relevant texture analysis methods (e.g., GLCM and GLRLM) outperformed the classical predefined texture features when applied to the most difficult set of 45 Brodatz texture pairs.

Yogesan et al. (19,20) introduced the use of predefined, nonadaptive texture features computed from gray level entropy matrices (GLEM) 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. Dunn et al. (21) applied adaptive nuclear textural features extracted from GLEMs for the assessment of dysplasia in Barrett's oesophagus, and found that these features differentiated dysplastic and nondysplastic cases. We also found that adaptive GLEM nuclear features were of independent prognostic significance for relapse-free survival in early stage ovarian cancer (10).

In a recent study on the same material as used in the current study, disease stage was found to be the most important prognostic factor for all tumor histology types (1,2). The prognosis of patients with leiomyosarcoma (LMS) at Stage I was also significantly related to tumor size and MI, and the prognosis of patients with endometrial stromal sarcoma (ESS) at Stage I was related to MI and tumor cell necrosis. Kildal et al. (11) analyzed DNA ploidy as a prognostic marker in the same series, and found that DNA ploidy was useful as a prognostic marker in patients with LMS and adenosarcoma (AS). In the current study, we evaluated the prognostic value of adaptive entropy-based features in the population based series of 354 cases of uterine sarcomas, and studied the relation between nuclear texture, DNA ploidy, β-catenin expression (22), and clinicopathological features of the sarcomas.

Materials and Methods

Patients

This retrospective study was performed on tissue samples from 354 uterine sarcomas (11). A total of 587 uterine sarcomas were registered from 1970 to 2000 at the Norwegian Cancer Registry, which gathers information on all cancer events in Norway (2,11). Survival dates were provided by the Cancer Registry on 31st October 2007 for all patients. The tumors were reclassified by an experienced gynecological pathologist (VMA) according to the World Health Organization (WHO) recommendations (23), and the diagnosis of uterine sarcoma was confirmed in 419 of the 587 patients (2). The observation time for patients still alive was a minimum of 82 months (2). Of the 419 patients, 354 cases could be included in this study, 29 cases were not admitted to surgery, tissue blocks with tumor material could not be obtained in 15 cases, and DNA ploidy classification could not be obtained in 21 cases because of poor quality of the tumor material) (11). A more detailed description of the material has been given previously (2). The study was approved by the Regional Ethics Committee.

Cell Nuclei and Imaging

Paraffin-embedded tissue fixed in 4% buffered formalin was used for preparation of nuclei suspension. At histologic review, representative areas with (the most aggressive) tumor tissue and without necrosis, hemorrhage, or inflammation were selected for analysis. Monolayers (isolated nuclei) were prepared from one or more 50 μm sections using a modification of Hedley's method (24). The nuclei were stained with Feulgen-Schiff according to an established protocol (25).

A DNA ploidy system, 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, Japan) with 1,024 $\times$ 1,024 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. Each nucleus was

segmented from the background using a global threshold and the segmented nuclei were stored in galleries in each case. Trained personnel performed a screening of the nuclei and selected tumor nuclei for the analysis. Stromal nuclei, necrotic nuclei, doublets, and cut nuclei were discarded. The mean number of measured tumor nuclei per case was 1,229 (278-3,972), and the median number of nuclei was 1,295.

Training and Validation Data Sets

The data set was randomly divided into a training set ($n = 175$) and a validation set ($n = 179$), and then balanced for histology type by moving seven LMS cases from the validation set to the training set and five ESS and two AS cases from the training to the validation set. The training set cases were grouped into two different prognostic classes. The patients who survived for at least 5 years ($n = 92$, 52.6%) were defined as members of the good prognosis class, whereas the patients that died within 5 years ($n = 83$, 47.4%) were defined as members of the poor prognosis class. The training set, where the prognosis of each case was known, was used for designing classifiers based on single texture features, whereas the independent validation set was used for evaluation of the classifiers. The validation set included 92 (51.4%) good and 87 (48.6%) poor prognosis cases. The training set was also used in a search for useful combinations of texture with DNA ploidy (11), β-catenin expression (22), and relevant clinicopathological parameters (2), and only parameters that were significant for 5-year crude survival were applied on the complete data set.

Nuclear Texture Analysis

2D GLEM. 2D GLEMs (9,10,19) were computed from all nuclear images (Fig. 1a). The GLEM element *P(i,j)* contains the normalized frequency of a local first order gray level entropy value j within a window of size $w \times w$ centered around a pixel with gray level value i (19). The gray level entropy is defined as

$$j = -\sum_{i=1}^{G} P(i) \times \log P(i), 0 < P(i) \leq 1$$

where $P(i)$ is the normalized frequency of occurrence of gray level i within the window and G is the number of gray level re-quantization levels in the image. The entropy measures the gray level non-uniformity within the window. Homogeneous structures will give low entropy values whereas inhomogeneous structures will give high entropy values. The number of gray levels in the nuclear images was reduced by re-quantization to 64 before computation of the GLEMs (10).

In the current study, we extracted seven predefined, non-adaptive texture features (19) from GLEM matrices computed for $w = 15$. Each of these features is defined as a weighted sum of the GLEM element values, where the weighting is based on either the value of the matrix element [entropy homogeneity (EH), entropy non-normality (ENN)] or the position in the matrix [average entropy (AE), low-entropy emphasis (LEE), high-entropy emphasis (HEE), low gray-level entropy emphasis (LGEE), high gray-level entropy emphasis (HGEE)]. Each case (patient) was represented by the median feature value computed from all nuclei/case.

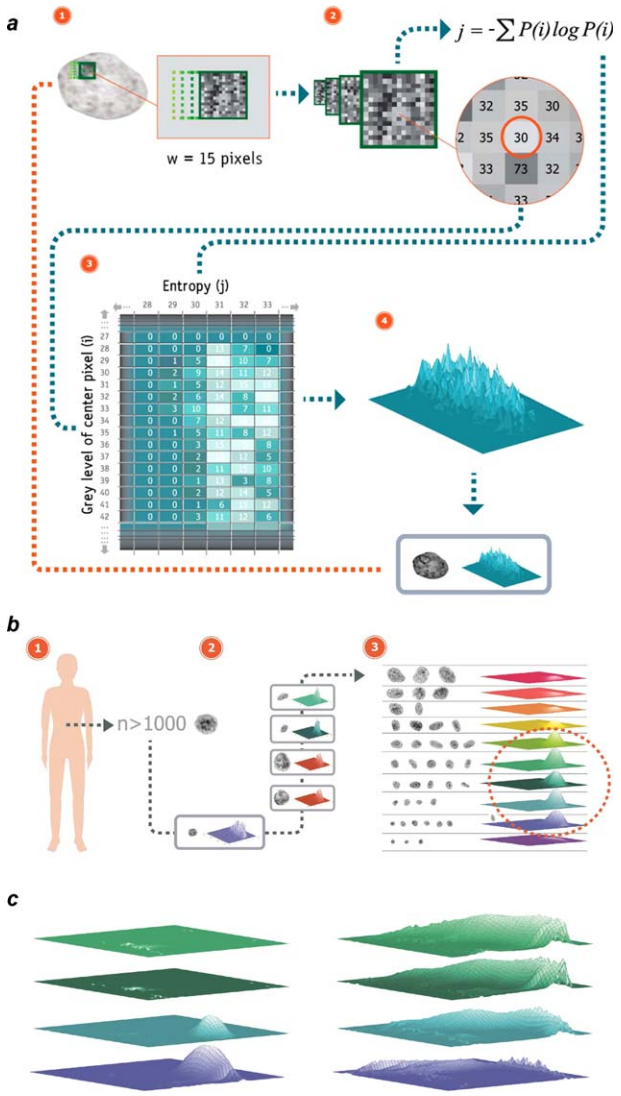

Figure 1. (**a**) The computation of a gray level entropy matrix (GLEM). 1: A moving window of size $w \times w$ pixels is centered around each pixel in a nuclear image, 2: For each position in the image, the gray value (i) of the center pixel and the gray level entropy value (j) of the pixels within the window are extracted, 3: i and (the scaled) j are used as indexes in the GLEM, and the frequency of obtaining different (i, j)-patterns (entropy patterns) is accumulated, and 4: The final GLEM is normalized by dividing each element in the matrix by the total number of pixels in the nuclear image. (**b**) Computation of a 3D patient matrix. 1,2: For each nucleus representing a given patient, a 2D GLEM is computed, and 3: a 3D patient matrix using the nuclear area group (a=1−10) as a third axis is accumulated. The 3D patient matrix is normalized by dividing each element by the number of nuclei representing the patient. (**c**) Entropy patterns that are emphasized in the computation of (left) *AF4Dpos* and (right) *AF4Dneg* using area groups a=2−5.

AF from 2D GLEM. As an alternative to the nine predefined, nonadaptive features defined in Ref. 19, we extract only

two AF from each GLEM (10,21). The nuclear images were grouped into 10 area groups $a=1,2,\ldots,10$, according to the number of pixels in the nucleus (where $a=1$ corresponds to a nuclear area of 1,000–1,999 pixels; $a=2$; nuclear area of 2,000–2,999 pixels, ..., $a=10$; nuclear area >10,000 pixels) (10,18). By using area groups we avoid problems caused by mixing data from cells having different nuclear area (18), and we can compute our features from the area groups that contain most of the prognostic information (10,18).

Our AF extraction consists of two steps: (i) Computation of a GLEM Mahalanobis class distance matrix and a GLEM class difference matrix between the two prognostic classes. Only the training set cases were included in the computation of these matrices. (ii) Extraction of two AF from the GLEM of each nuclear image. This feature extraction is based on the computed class distance and difference matrices from step (i).

Step (i): Computation of adaptive weighting functions. For each cell nucleus from a patient, a 2D GLEM was computed, using a window size of 15×15 pixels. The nuclear images from the n-th patient of class ω_c ($c=1,2$; ω_1 = good prognosis, ω_2 = poor prognosis) give a set of average patient matrices $P_n(i,j|a,\omega_c)$, for $a=1,\ldots,10$. On the basis of these patient matrices, we then calculate average matrices over all the $N(a,\omega_c)$ training set patients in each area group of the two classes

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

On the basis of these matrices, we compute a GLEM class difference matrix in each area group (see Fig. 2a for the matrix computed from $a=3$)

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

and a GLEM class variance matrix in each area group of the two classes

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

and finally the GLEM Mahalanobis class distance matrices

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

The class distance matrices will now have higher values in those parts of the GLEM domain where there are class-discriminating differences, making them more useful as weighting functions than longer lists of previously predefined static weighting functions.

Step (ii): Computation of AF. For each nuclear image, we extract two AF from the 2D GLEM by using the 2D squared Mahalanobis class distances 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

$$AF2Dpos=\sum_{\Delta_P(i,j|a,w,\omega,\omega_2)>0}P(i,j|a)[J_P^2(i,j|a,\omega_1,\omega_2)]$$
$$AF2Dneg=\sum_{\Delta_P(i,j|a,w,\omega_1,\omega_2)<0}P(i,j|a)[J_P^2(i,j|a,\omega_1,\omega_2)]$$

Figures 2b and 2c show the squared Mahalanobis distances computed from the training set of uterine sarcomas. These figures illustrate, which parts of the GLEM that were emphasized in the computation of each feature. Each case (patient) was represented by the (scalar) mean value of the texture feature values extracted from GLEM matrices of all nuclei within area groups $a=2-4$ (10,18). Because the two AF extract information from the parts of the matrix where the statistical differences between classes are highest, and take into account, which way the differences go, we have obtained a compact feature set. Thus, we avoid the computation of a long list of ad hoc features based on predefined, static weight functions, followed by some feature selection strategy.

Jensen-Shannon divergence between GLEMs. To quantify how similar the GLEM of a given nucleus was to each of the two class average GLEMs, we computed the Jensen-Shannon divergence (26) from the nuclear GLEM to these average matrices, thus getting two distance measures per nucleus, JS_{Good} and JS_{Poor}. We also computed the difference between these distances, $JS_{Diff}=JS_{Good}-JS_{Poor}$, where a negative JS_{Diff} corresponds to nuclei with GLEMs more similar to the average good prognosis matrix, whereas a positive JS_{Diff} corresponds to GLEMs more similar to the average poor prognosis matrix. We also computed the proportion of nuclei (within area groups $a=2-4$) with negative JS_{Diff} of each of the training set cases.

AF from 4D patient matrices. For each nucleus representing a given patient, a 2D GLEM was computed, and a 3D patient matrix $P(i,j,a|w)$ using the nuclear area group a as a third axis was accumulated. The 3D patient matrix was normalized by dividing each element by the number of nuclei representing the patient (Fig. 1b). Such 3D patient matrices were computed for several different window sizes $w=5,7,\ldots,31$, and a 4D patient matrix $P(i,j,a,w)$ was obtained by concatenating the 3D matrices computed for the 14 window sizes (10).

Assuming that the n-th patient of class ω_c, gives a 4D patient matrix $P_n(i,j,a,w|\omega_c)$ we then calculate an average matrix over all the training set patients in each class ω_c, and two 4D class variance matrices. On the basis of these matrices, we compute a 4D class difference matrix and a 4D Mahalanobis class distance matrix in the same manner as described for 2D matrices. Finally, we extract two AF (*AF4Dpos*, *AF4Dneg*)

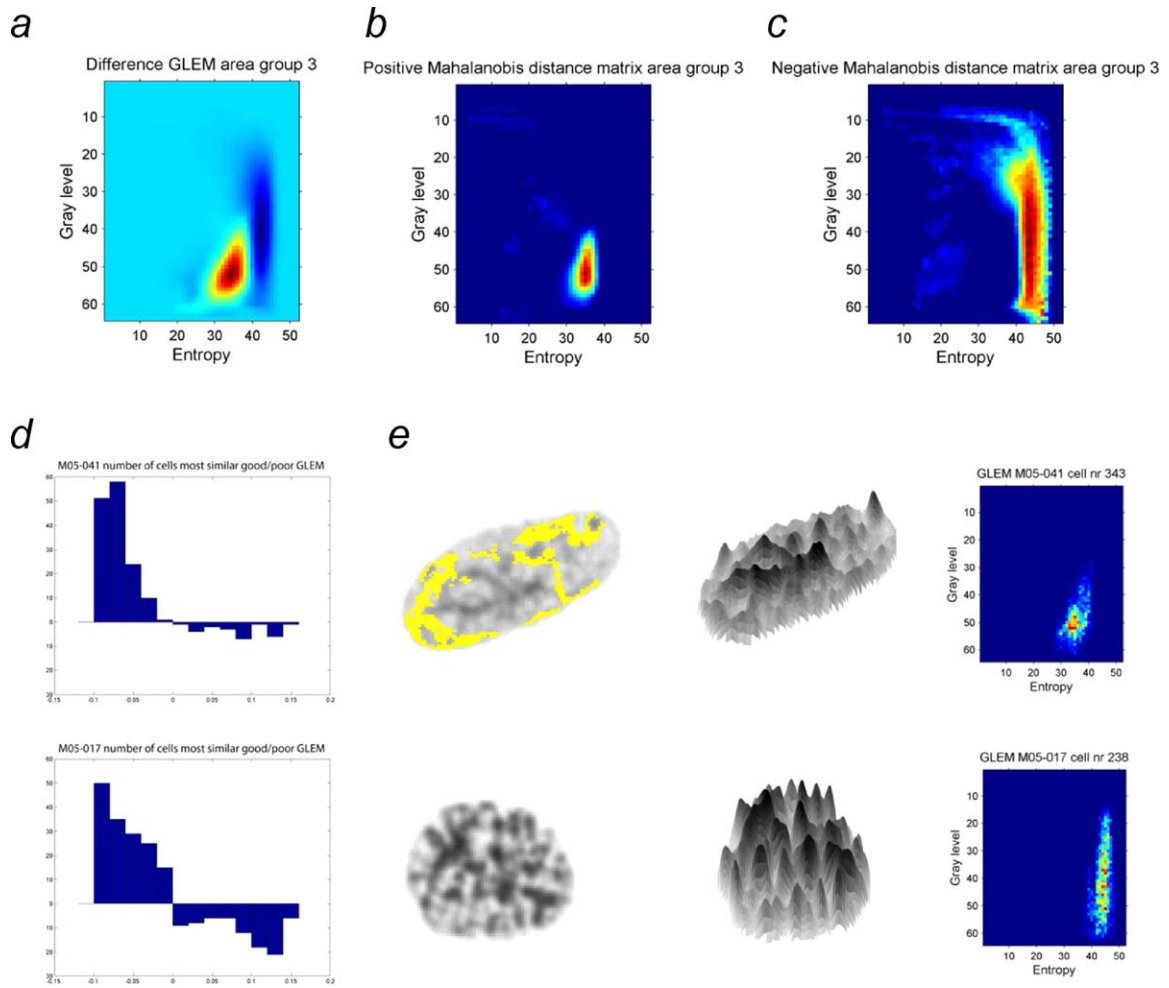

Figure 2. (**a**) The difference between average gray level entropy matrices (GLEMs) computed from the good and poor prognosis training cases. Positive (red, yellow)/negative (dark blue) values in the matrix correspond to entropy patterns that are more/less probable for good compared to poor prognosis cases. (**b, c**) Entropy patterns that are given the largest weights (squared Mahalanobis class distances) in the extraction of (b) *AF2Dpos* and (c) *AF2Dneg*. (**d**) Histogram of difference between Jensen-Shannon distances, $JS_{Diff} = JS_{Good} - JS_{Poor}$ based on nuclei from two example cases. The part of the histogram with positive *y*-values (negative JS_{Diff}) corresponds to nuclei with GLEMs more similar to the average good prognosis matrix, whereas the part of the histogram with negative *y*-values corresponds to nuclei with GLEMs more similar to the average poor prognosis matrix. (**e**) Two example nuclei with corresponding GLEMs. The entropy patterns that contributed most to the *AF2Dpos* feature value (i.e., positive differences in the class difference matrix and squared Mahalanobis distances > 0.3) are visualized as yellow pixels in the nuclei. The feature *AF2Dpos* gives a relatively "high" feature value for the example cell (number 343) from the good prognosis case M05-041 and a low feature value for the cell (number 238) from the poor prognosis cases M05-017. In the pseudo-3D representation of the nuclei, the inverse gray level (1,024 gray levels) of each pixel represents the height on the *z*-axis (i.e., black pixels correspond to maximum height). The illustrated matrices were computed from nuclei within area group $a=3$; i.e., nuclei with nuclear area of 3,000–3,999 pixels.

from each 4D patient matrix by using the squared 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. In the current study, the AF are extracted from a single 4D matrix per patient (10), using area groups $a=2-5$ in the computation of the features (Fig. 1c).

Implementation. The texture analysis method was implemented in Matlab (R2012b). A fast version of our code for computation of GLEMs was implemented in Java and called from the Matlab program.

Classification

To evaluate the prognostic value of each AF, minimum Euclidean distance classifiers (9) based on single features were constructed to classify each case (patient) into the prognosis classes. The classification results were then used as input to survival analysis. The classifiers were designed on the training data set. The correct classification rate (CCR), sensitivity, and specificity on the training set were estimated by resubstitution, and the classifiers that were significant in univariate analysis were evaluated on the separate independent validation set (holdout validation) (9,12).

For comparison, predefined features from GLEM matrices were extracted and minimum Euclidean distance classifiers based on single features and combinations of two and three features were constructed. Bootstrap estimates of CCR, sensitivity, and specificity (resampling with replacement, 10,000 iterations) were computed by dividing the training set into several smaller training (46 good and 42 poor prognosis cases) and test (46 good and 41 poor prognosis cases) sets.

Statistical Analysis

The SPSS statistical package (SPSS Statistics 20) was used for survival analysis. Survival of patients was estimated using univariate Kaplan-Meier analysis. Crude survival was calculated from date of diagnosis to death or end of (5-year) follow-up. 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 was used for both univariate and multivariate analyses. *P*-values <0.05 were considered statistically significant.

The variables were grouped as follows: histological subtype as LMS, ESS, AS, undifferentiated uterine sarcomas (UUS), or other sarcomas; DNA ploidy as diploid/nondiploid; MI as below ($\leq$)/above 10 per 10 high-power field (HPF); tumor extent as confined to uterus (Stage I)/not confined to uterus (Stages II–IV); tumor size as below ($\leq$)/above 10 cm; tumor margins as pushing or infiltrating; cellular atypia as mild, moderate or severe; tumor necrosis as present or absent, hyaline necrosis as present or absent, and vascular invasion as present or absent (11). The texture feature (*AF4Dpos*) was grouped as high/low feature value based on an optimal threshold identified in the training set. β-catenin expression was grouped as positive/negative membranous, cytoplasmic and nuclear (22).

Results

All adaptive texture features were significant for crude survival in Cox univariate analyses based on the training set, and were also found to be significant when evaluated on the independent validation set. The best texture feature (*AF4Dpos*) classified the training cases into good or poor prognosis with a CCR of 67%, and the validation cases were classified with a CCR of 68% (Table 1), and this feature was selected for further analyses. The 5-year crude survival, as computed from the complete data set, was significantly higher ($P<0.001$) for patients with high texture feature value (72%) than for patients with low feature value (36%, Fig. 3a). When combining DNA ploidy category (diploid/nondiploid) and texture (high/low feature value), the training set patients could be stratified into tree risk groups: a low risk group for cases classified as DNA diploid, a medium risk group for nondiploid cases with high texture feature value, and a high risk group for nondiploid cases with low texture feature value, and this result was verified on the validation set. The 5-year crude survival computed for the complete data set for the three risk groups were 77% ($n=110$), 57% ($n=69$), and 34% ($n=175$), with relative HR of 1, 2.3, and 4.1 ($P<0.001$), respectively (Fig. 3b).

Table 1. The correct classification rates (CCR), sensitivity, and specificity obtained by minimum Euclidean classifiers based on single texture features, and hazard ratios and *P*-values obtained by Cox proportional hazard regression model of the 175 training cases (92 good prognosis and 83 poor prognosis) and the 179 validation cases (92 good prognosis and 87 poor prognosis).

FEATURE:	CCR (%)	SENS. (%)	SPEC. (%)	HR (95% CI) (5-YEAR)	*P* (5-YEAR)
Training					
AF2Dpos	61	53	67	1.83 (1.19–2.81)	0.006
AF4Dpos	67	73	61	2.93 (1.80–4.78)	<0.001
Validation					
AF2Dpos	65	61	68	2.17 (1.41–3.34)	<0.001
AF4Dpos	68	74	63	2.96 (1.84–4.78)	<0.001

CI, confidence interval; HR, hazard ratio.

The CCR bootstrap estimates of the predefined, nonadaptive GLEM features were about 50% for AE, LEE, HEE, and ENN and about 55% for LGEE, HGEE, and EH. Combinations of two and three features did not increase the CCR values.

Univariate analyses were performed separately on patients with tumors of the two largest histology types (Figs. 3c–3f), whereas the other types were too small for similar analysis. The 5-year crude survival for patients with LMS were 65% for tumors with high texture feature value compared to 34% for patients with LMS with low feature value (HR = 2.4, $P<0.001$). The 5-year crude survival for patients with Stage I LMS tumors were 69% versus 42% for tumors with high versus low texture feature value (HR = 2.2, $P=0.002$). The 5-year crude survival was 82% for ESS patients with high feature value and 50% for patients with low feature value (HR = 3.5, $P=0.006$), and the 5-year crude survival for patients with Stage I ESS were 90% versus 67% for patients with high versus low texture feature value (HR = 3.7, $P=0.067$).

Texture was significant in multivariate analysis (Table 2). Figures 3g–3q show survival curves based on texture, stratified for the relevant clinicopathological features that were significant for 5-year crude survival on the training data set. β-catenin expression was not found to be significant for 5-year crude survival when including only the training cases.

Figure 4 shows survival curves for the three risk groups based on tumor size and MI that were defined for Stage I LMS cases in the study of Abeler et al. (2), and also survival curves based on texture, stratified for these risk groups.

Figure 2 illustrates the difference in nuclear texture between the two prognostic classes. The entropy-patterns were more homogeneous in nuclei from patients with a good prognosis (Fig. 2b), whereas entropy patterns in nuclei from poor prognosis patients showed more variation (Fig. 2c). The average proportion of nuclei with negative JS_{Diff} (i.e., with entropy matrices more similar to the average good prognosis matrix compared to nuclei with more deviating patterns) were 76.9% for the 92 good prognosis and 58.1% for 83 poor prognosis training set cases. Figures 2d and 2e

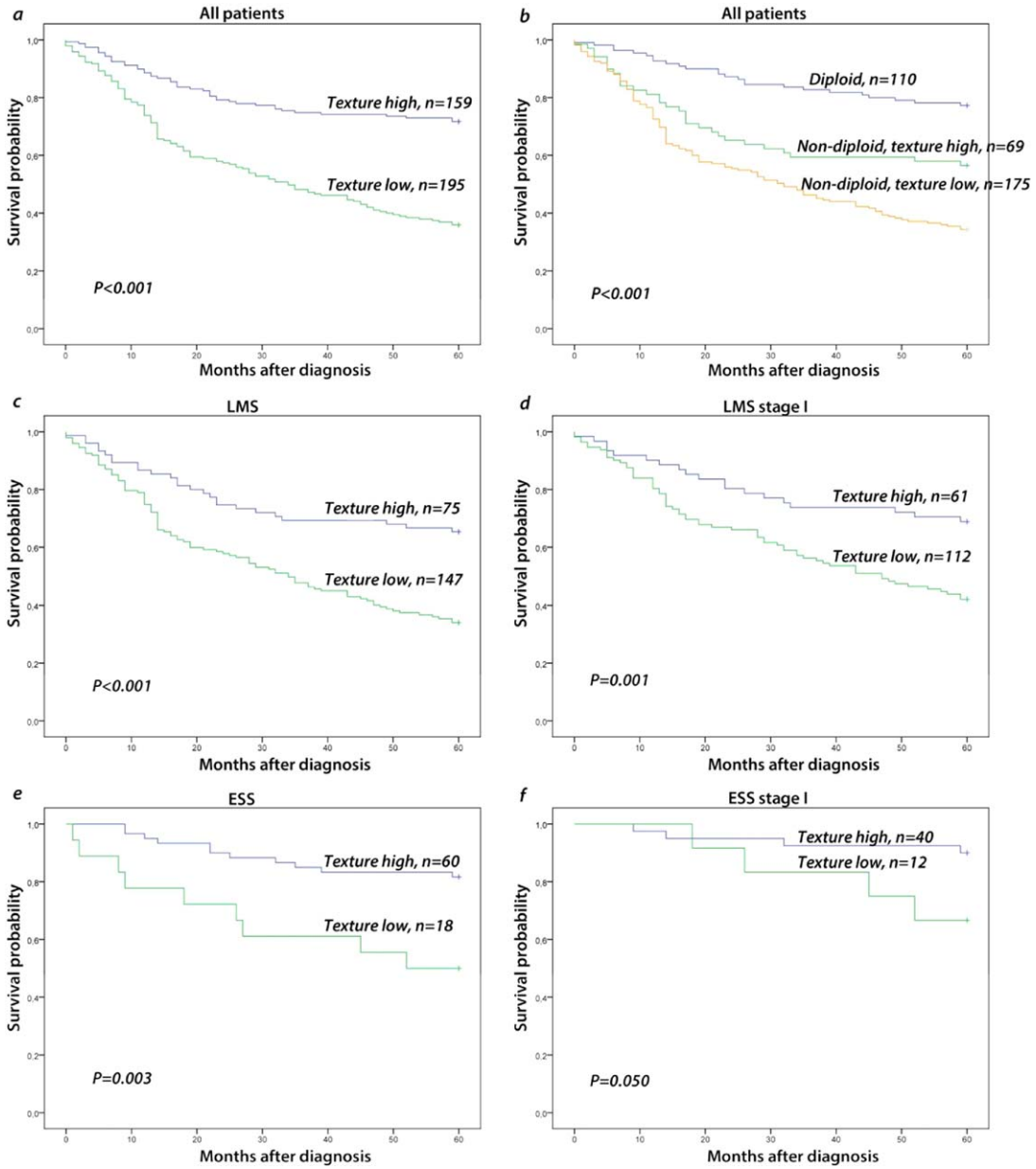

Figure 3. Kaplan-Meier 5-year crude survival curves based on texture. (**a**) Survival curves are based on the complete data set ($n = 354$), HR = 2.9 (2.1–4.1), (**b**) Survival curves based on a combination of texture (high/low value) and DNA ploidy category (diploid/nondiploid) on the complete data set. Survival curves based on (**c**) all LMS cases ($n = 222$), HR = 2.41 (1.56–3.72) (**d**) LMS Stage I ($n = 173$), HR = 2.25 (1.35–3.75) (**e**) all ESS cases ($n = 78$), HR = 3.47 (1.43–8.38), and (**f**) ESS Stage I ($n = 52$), HR = 3.65 (0.91–14.64). Five-year crude survival curves based on texture stratified for (**g**) tumor extent; tumor confined to the uterus ($n = 267$ cases), (**h**) tumor spread outside the uterus ($n = 87$), (**i**) MI; 0–10 per 10 high power field ($n = 207$), (**j**) >10 per 10 high power field ($n = 143$), (**k**) tumor size; ≤10 cm ($n = 260$) (l) above 10 cm ($n = 75$), (**m**) tumor necrosis; absent ($n = 86$), (**n**) present ($n = 264$), (**o**) cellular atypia; mild ($n = 106$), (**p**) moderate ($n = 130$), and (**q**) severe ($n = 112$). *P*-values were estimated by the log-rank test and hazard ratios were estimated by the Cox model.

illustrate the difference in texture between two example training set LMS cases with different prognosis. The proportion of nuclei with negative JS_{Diff} were higher for the good prognosis case M05-041 (85% within area Group 3) than for the poor prognosis case M05-017 (64%, Fig. 2d). The figure also illustrates an example of mapping windows contributing to the entropy matrices back to cell nuclei. The entropy patterns that were given the largest weights in the extraction of *AF2Dpos* (as shown in Fig. 2b) were identified, and the pixels in two example nuclei that contributed to these patterns were identified and visualized (Fig. 2e). These patterns occurred 811 times (for 25% of the nuclear pixels) in the nucleus from the good prognosis case and zero times in the nucleus from the poor prognosis case.

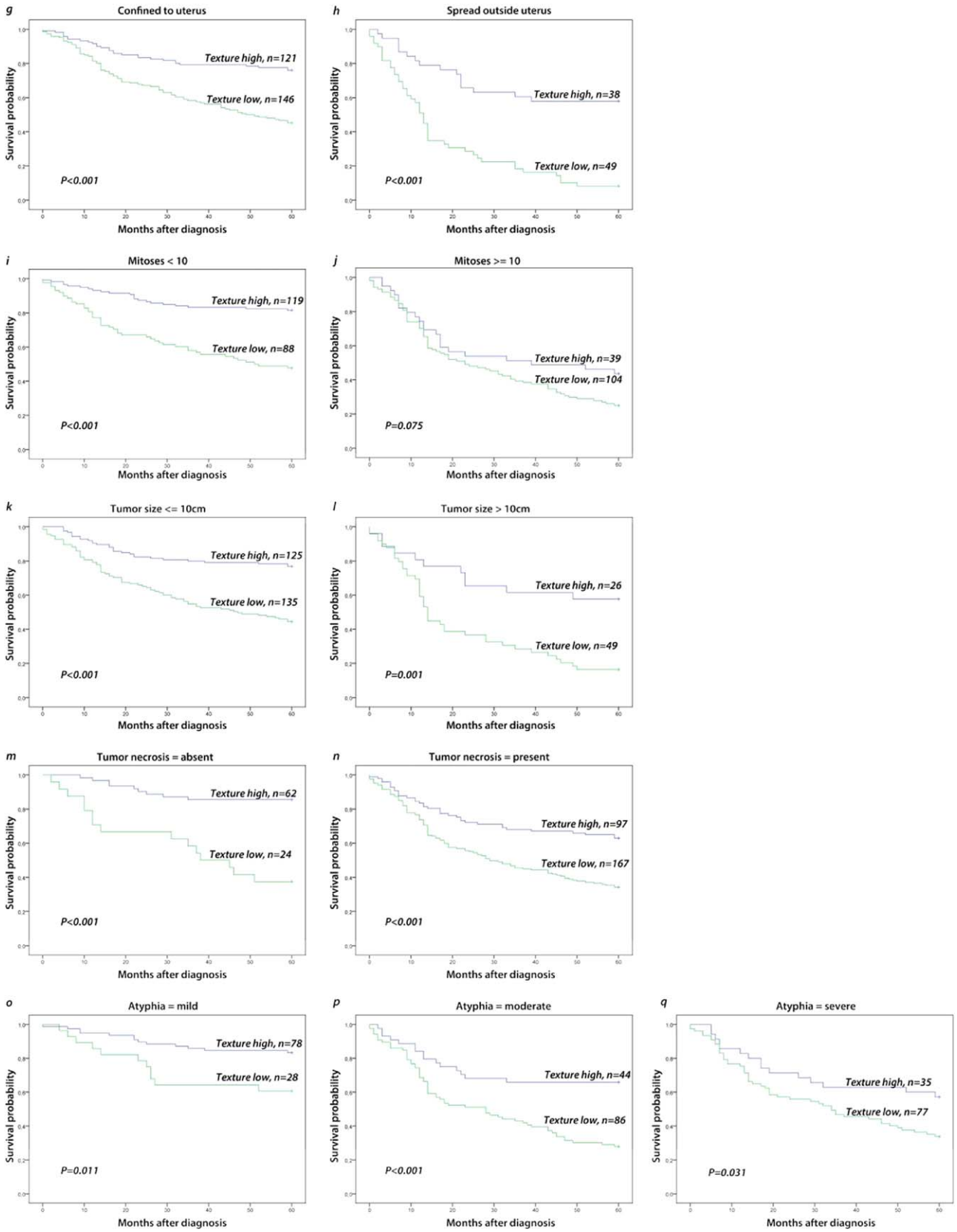

Figure 3. (Continued)

Table 2. Five-year crude survival related to nuclear texture, tumor extent, MI, tumor size, tumor necrosis, cellular atypia, hyaline necrosis, vascular invasion, tumor margins, and tumor type.

FEATURE:	UNIVARIATE ANALYSIS: *P*	MULTIVARIATE ANALYSIS HR (95% CI)	*P*
Texture:			
High value	<0.001	1.0	0.001
Low value		2.1 (1.4–3.2)	
Tumor extent:			
Confined to the uterus	<0.001	1.0	<0.001
Spread outside the uterus		2.7 (1.8–4.0)	
MI:			
0–10 high-power field	<0.001	1.0	<0.001
>10 high-power field		2.3 (1.6–3.4)	
Tumor size:			
0–10 cm	<0.001	1.0	0.003
>10 cm		1.8 (1.2–2.6)	
Tumor necrosis:			
Present	<0.001	1.0	0.122
Absent		1.5 (0.9–2.6)	
Cellular atypia:			
Mild	<0.001	1.0	0.478
Moderate		1.4 (0.7–2.6)	
Severe		1.2 (0.6–2.3)	
Hyaline necrosis:			
Present	0.045	1.0	0.770
Absent		1.1 (0.7–1.5)	
Vascular invasion:			
Present	0.023	1.0	0.116
Absent		1.3 (0.9–1.9)	
Tumor margins:			
Pushing	0.040	1.0	0.300
Infiltrating		1.3 (0.8–2.0)	
Tumor type:			
LMS	<0.001	1.0	0.066
ESS		0.7 (0.3–1.5)	
AS		1.4 (0.5–3.7)	
UUS		0.6 (0.3–1.5)	
Other sarcomas		2.3 (1.2–4.4)	

P value: Univariate analysis; crude survival analysis (log-rank), Multivariate analysis; Cox proportional regression model; HR, hazard ratio; CI, confidence interval.

Missing values; tumor size 19, cellular atypia 6, MI 4, tumor necrosis 4, hyaline necrosis 12, vascular invasion 21 and tumor margins 16.

LMS, leiomyosarcoma; ESS, endometrial stromal sarcoma; AS, adenosarcoma; USS, undifferentiated uterine sarcoma.

Other sarcomas include 10 sarcoma not otherwise specified, four rhabdomyosarcoma, two giant cell tumors with/without LMS and one PEComa.

Discussion

In the field of cancer medicine, much current effort goes into the development of predictive genetic profiling tests for common cancers. The rationale here is that because all cancers are believed to originate in genomic alterations, it could be possible to identify clusters of genes whose activity is enhanced or depressed in patients having a good prognosis, or in patients whose cancer will fail to respond to a specific therapy regime. A similar rationale supports the use of quantitative image-based DNA ploidy measures because it appears that a tumor's ability to metastasize is connected to the presence within it of cell lines with aneuploidy DNA content. Such large-scale genomic instability must correlate with large-scale rearrangement of interphase nuclear chromatin. Nuclear texture analysis reflects such large-scale chromatin rearrangement and is therefore interesting as a potential method of choice for the prediction of significant clinical outcomes (9,15,27). In general, the literature on nuclear texture analysis supports this idea, though there are methodological challenges related to the large number of possible features that can be measured, just as there are difficulties with approaches to the same problem that are based on genomic profiling (9,28). In the current study, we computed a compact set of superior adaptive nuclear texture features based on local nuclear gray level entropy. The training set of 175 patients was used for designing classifiers based on single texture features. When evaluating the classifiers on the validation set of 179 patients, we obtained similar classification results as obtained on the training set (Table 1). The strength of using separate training and validation data sets is that an eventual overfitting during training will be showed by testing (9). The only true reliable estimate of the performance of a classifier is obtained from separate training and validation sets (12,28,29). When comparing the adaptive nuclear texture feature extraction approach with the more conventional predefined features extracted from the same entropy matrices, we found that both single nonadaptive features and feature combinations were outperformed by our low dimensional AF extraction approach.

When comparing crude survival for diploid ($n = 28$) and nondiploid ($n = 194$) LMS tumors, Kildal et al. (11) found that the difference did not reach statistical significance ($P = 0.051$), whereas texture gave significantly better 5-year crude survival ($P < 0.001$) for LMS with tumors with high feature values (65% survival, $n = 75$) compared to low feature values (34%, $n = 147$, Figs. 3c and 3d). Kildal et al. (11) found that 5-year crude survival for diploid ESS cases was 83% and nondiploid ESS cases was 40% ($P < 0.001$). Texture gave similar results for these tumors, 82% 5-year crude survival for the 60 cases with high feature values and 50% survival for the 18 cases with low feature values ($P = 0.003$, Figs. 3e and 3f).

Abeler et al. (2) performed a histopathological review of all cases (uterine sarcomas in Norway between 1970 and 2000) and examined the possibility of overdiagnosis of leiomyomatous tumors. The diagnosis of uterine sarcomas was confirmed in 419 (71%) cases, whereas most of the 168 excluded cases got a revised diagnosis (28 benign leiomyoma, 29 cellular leiomyoma, 15 atypical leiomyoma, 20 leiomyomatous tumor of uncertain malignant potential, two inflammatory myofibroblastic tumor, three LMS outside uterus + benign leiomyoma in the uterus, 14 carcinoma, 16 carcinoma + leiomyomatous tumor, one lymphoma). Because of this diagnostic uncertainty, we have chosen to develop our

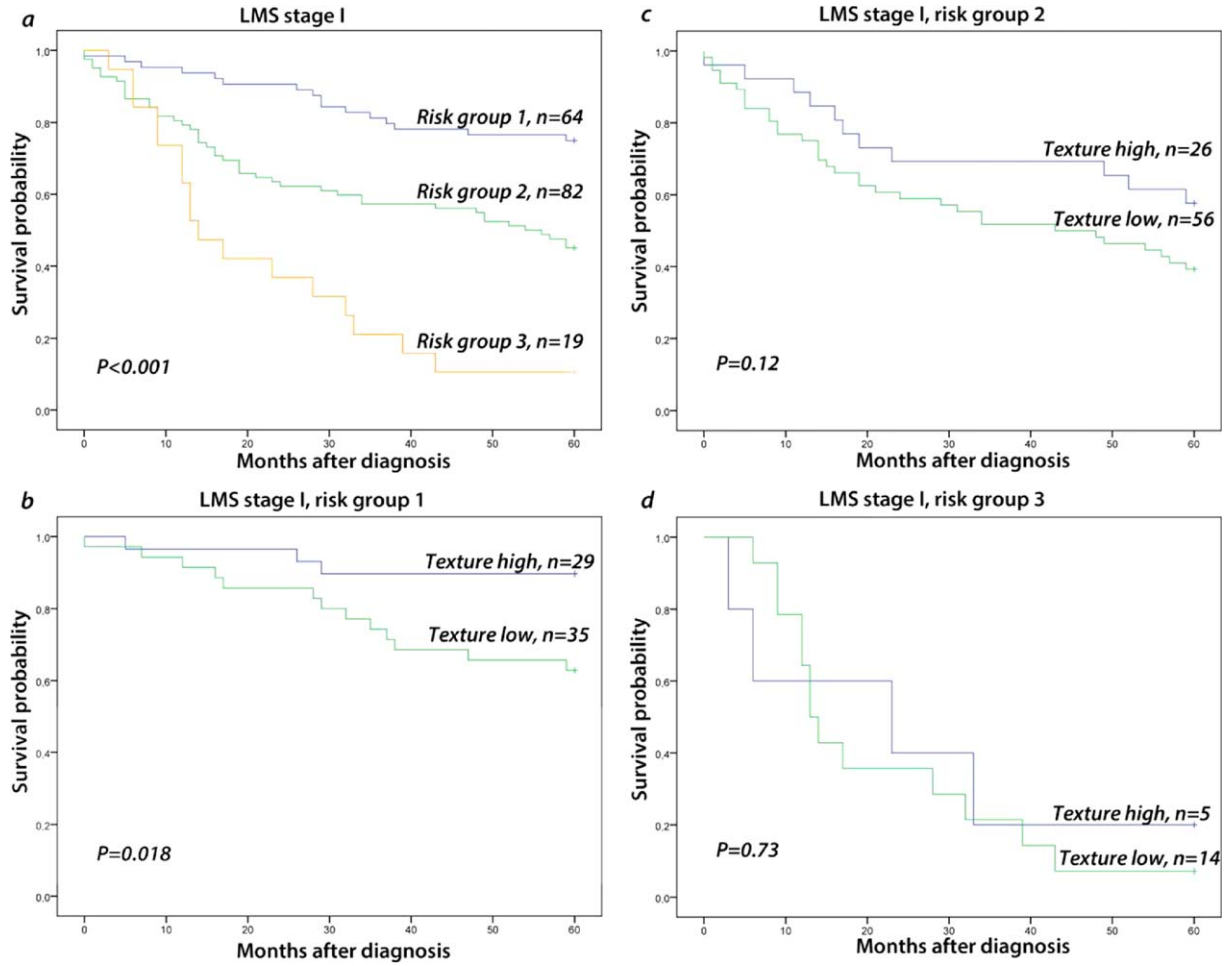

Figure 4. Kaplan-Meier 5-year crude survival curves for risk groups based on MI and tumor size for (**a**) LMS Stage I ($n = 165$). Crude survival based on the texture stratified for (**b**) risk Group 1; LMS Stage I ($n = 69$), (**c**) risk Group 2; LMS Stage I ($n = 95$) and (**d**) risk Group 3; LMS Stage I ($n = 19$). The risk groups were defined as in Ref. (2): low risk: tumor size ≤ 10 cm and MI ≤ 10 per high power field (HPF), medium risk: either tumor size > 10 cm or MI > 10 per 10 HPF, high risk: tumor size > 10 cm and MI > 10 per HPF. *P*-values were estimated by the log-rank test.

prognostic marker based on nuclear texture independently of histological type. However, we have shown that nuclear texture is of prognostic value both independently of histological types and also within the two types tested.

In a multivariate analysis, including nuclear texture, tumor extent, MI, tumor size, tumor necrosis, cellular atypia, hyaline necrosis, vascular invasion, tumor margins, and tumor type, only nuclear texture, tumor extent, MI, and tumor size were found to be of independent prognostic significance for crude survival (Table 2). Texture contained additional prognostic information in combination with each of the clinicopathological features that were significant in univariate analysis of the training data set (Figs. 3g–3q). Abeler et al. (2) found that tumor size and MI were significant prognostic markers ($P < 0.001$) in leiomyosarcomas confined to the uterus (Stage I) and that these parameters allowed for a separation into three risk groups with marked differences in prognosis. In the current study, we have shown that nuclear texture contained additional prognostic information when stratified for the two largest risk groups (Figs. 4b and 4c).

In several studies (9,10,17,18), we have found a difference in chromatin structure between cell nuclei from different clinical classes. In the current study, we have found that the entropy patterns were more homogeneous in nuclei from patients with a good prognosis, whereas entropy patterns in nuclei from poor prognosis patients showed more variation. 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. 2a). These findings were also observed in a study of tumors from patients with early ovarian cancer (10), and in a study on dysplasia in Barrett's oesophagus, we found that entropy patterns in nuclei from dysplastic cases showed more variation than in nuclei from nondysplastic cases (21).

In digital pathology, the concept of entropy is useful for describing and quantifying the heterogeneity in digital images of cancer tissue and nuclei (27,30). It is not unreasonable to believe that the mutations of the genome and epigenetic

alterations that occur during carcinogenesis will increase the randomness and spatial entropy inside the nucleus. Therefore, with increasing malignant potency of a tumor an increase in entropy is both expected and observed in cytological and histological preparations in diagnostic and prognostic studies (27). Several studies have also shown that the fractal dimension of chromatin in routinely stained histological or cytological preparations increases in various neoplasias during carcinogenesis and tumor progression, and is higher in patients with poor prognosis compared to patients with good prognosis (15). Both fractal dimension and gray level entropy quantify the complexity of the chromatin texture in digital images (31). However, the relation between fractal dimension computed from the complete nucleus and local entropy computed within small windows needs to be investigated further.

This study showed that adaptive nuclear texture features based on local entropy are independent prognostic markers for crude survival in uterine sarcomas. Given these results, these features should be considered as a potential prognostic marker in uterine sarcoma. On the basis of the result in the current study, together with results from earlier studies on ovarian cancer (9,10,18) and Barrett's oesophagus (21), we believe that nuclear texture is a more general prognostic marker that is promising because of its ability to measure large scale genomic instability in cancer cell nuclei. Further studies are required to investigate the underlying mechanisms behind these results, and how pixel-based texture features link to large-scale rearrangement of interphase nuclear chromatin.

Acknowledgments

The authors thank Andreas Kleppe for the implementation of a fast Java-version of our Matlab-code for computation of gray level entropy matrices, and the code for computing the bootstrap estimates of the correct classification rates. The authors also thank Marian Seiergren for the design and production of Figures 1 and 2.

Literature Cited

1. Tropé CG, Abeler VM, Kristensen GB. Diagnosis and treatment of sarcoma of the uterus. A review. Acta Oncol 2012;51:694–705.
2. Abeler VM, Røyne O, Thoresen S, Danielsen HE, Nesland JM, Kristensen GB. Uterine sarcomas in Norway. A histopathological and prognostic survey of a total population from 1970 to 2000 including 419 patients. Histopathology 2009;54:355–364.
3. Nordal RR, Thoresen SØ. Uterine sarcomas in Norway 1956–1992: Incidence, survival and mortality. Eur J Cancer 1997;33:907–911.
4. Amant F, Coosemans A, Debiec-Rychter M, Timmerman D, Vergote I. Clinical management of uterine sarcomas. Lancet Oncol 2009;10:1188–1198.
5. Toro JR, Travis LB, Wu HJ, Zhu K, Fletcher CD, Devesa SS. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978–2001: An analysis of 26,758 cases. Int J Cancer 2006;119: 2922–2930.
6. D'Angelo E, Prat J. Uterine sarcomas: A review. Gynecol Oncol 2010;116:131–139.
7. Gal R, Rath-Wolfson L, Rosenblatt Y, Halpern M, Schwartz A, Koren R. An improved technique for mitosis counting. Int J Surg Pathol 2005;13:161–165.
8. Silverberg SG. Reproducibility of the mitosis count in the histologic diagnosis of smooth muscle tumors of the uterus. Hum Pathol 1976;7:451–454.
9. Nielsen B, Albregtsen F, Danielsen HE. Statistical nuclear texture analysis in cancer research: A review of methods and applications. Crit Rev Oncog 2008;14:89–164.
10. Nielsen B, Albregtsen F, Kildal W, Abeler VM, Kristensen GB, Danielsen HE. The prognostic value of adaptive nuclear texture features from patient gray level entropy matrices in early stage ovarian cancer. Anal Cell Pathol 2012;35:305–314.
11. Kildal W, Abeler VM, Kristensen GB, Jenstad M, Thoresen SØ, Danielsen HE. The prognostic value of DNA ploidy in a total population of uterine sarcomas. Ann Oncol 2009;20:1037–1041.
12. Schulerud H, Kristensen GB, Liestøl K, Vlatkovic L, Reith A, Albregtsen F, Danielsen HE. A review of caveats in statistical nuclear image analysis. Anal Cell Pathol 1998; 16:63–82.
13. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE T Syst Man Cyb 1973;3:610–621.
14. Galloway MM. Texture analysis using gray level run lengths. CVGIP 1975;4:172–179.
15. Metze K. Fractal dimension of chromatin: Potential molecular diagnostic applications for cancer prognosis. Expert Rev Mol Diagn 2013;13:719–735.
16. Walker RF, Jackway PT, Longstaff ID. Recent developments in the use of co-occurrence matrix for texture recognition. In: Proceedings of the 13th International Conference on Digital Signal Processing, Vol. 1. Santorini: Greece; 1997. pp 63–65.
17. Nielsen B, Albregtsen F, Danielsen HE. Low dimensional adaptive texture feature vectors from class distance and class difference matrices. IEEE T Med Imaging 2004; 23:73–84.
18. Nielsen B, Danielsen HE. Prognostic value of adaptive textural features—The effect of standardizing nuclear first-order gray level statistics and mixing information from nuclei having different area. Cell Oncol 2006;28:85–95.
19. Yogesan K, Jørgensen T, Albregtsen F, Tveter KJ, Danielsen HE. Entropy-based texture analysis of chromatin structure in advanced prostate cancer. Cytometry 1996;24: 268–276.
20. Jørgensen T, Yogesan K, Tveter KJ, Skjørten F, Danielsen HE. Nuclear texture analysis: A new prognostic tool in metastatic prostate cancer. Cytometry 1996;24:277–283.
21. Dunn JM, Hveem T, Pretorius M, Oukrif D, Nielsen B, Albregtsen F, Lovat LB, Novelli MR, Danielsen HE. Comparison of nuclear texture analysis and image cytometric DNA analysis for the assessment of dysplasia in Barrett's oesophagus. Brit J Cancer 2011;105:1218–1223.
22. Kildal W, Pradhan M, Abeler VM, Kristensen GB, Danielsen HE. Beta-catenin expression in uterine sarcomas and its relation to clinicopathological parameters. Eur J Cancer 2009;45:2412–2417.
23. Tavassoéli FA, Devilee P. Pathology and Genetics Tumours of the Breast and Female Genital Organs. IARC Press: Lyon; 2003.
24. Hedley DW. DNA analysis from paraffin-embedded blocks. Method Cell Biol 1994; 41:231–240.
25. Tanke HJ, van Ingen EM. A reliable Feulgen-acriflavine-SO2 staining procedure for quantitative DNA measurements. J Histochem Cytochem 1980;28:1007–1013.
26. Lin J. Divergence measures based on the Shannon entropy. IEEE T Inform Theory 1991;37:145–151.
27. Kayser K, Kayser G, Metze K. The concept of structural entropy in tissue-based diagnosis. Anal Quant Cytol Histol 2007;29:296–308.
28. Shulerud H, Albregtsen F. Many are called, but few are chosen. Feature selection and error estimation in high dimensional spaces. Comput Meth Prog Biomed 2004;73: 91–99.
29. Isaksson A, Wallman M, Göransson H, Gustafsson MG. Cross-validation and bootstrapping are unreliable in a small sample classification. Pattern Recogn Lett 2008;29: 1960–1965.
30. Kayser G, Kayser K. Quantitative pathology in virtual microscopy: History, applications, perspectives. Acta Histochem 2013;115:527–532.
31. Cattani C, Pierro G. On the fractal geometry of DNA by the binary image analysis. Bull Math Biol 2013;75:1544–1570.