

MULTICLASS PATTERN RECOGNITION OF THE GLEASON SCORE OF PROSTATIC CARCINOMAS USING METHODS OF SPATIAL STATISTICS

TORSTEN MATTFELDT[✉],¹, PAUL GRAHOVAC¹ AND SEBASTIAN LÜCK²

¹Institute of Pathology, University of Ulm, 89081 Ulm, Germany; ²Institute of Stochastics, University of Ulm, 89081 Ulm, Germany

e-mail: torsten.mattfeldt@uni-ulm.de, grahovac.pur@googlemail.com, seblueck@gmx.net

(Received October 10, 2013; revised November 25, 2013; accepted November 26, 2013)

ABSTRACT

The Gleason score of a prostatic carcinoma is generally considered as one of the most important prognostic parameters of this tumour type. In the present study, it was attempted to study the relation between the Gleason score and objective data of spatial statistics, and to predict this score from such data. For this purpose, 25 T1 incidental prostatic carcinomas, 50 pT2N0, and 28 pT3N0 prostatic adenocarcinomas were characterized by a histological texture analysis based on principles of spatial statistics. On sectional images, progression from low grade to high grade prostatic cancer in terms of the Gleason score is correlated with complex changes of the epithelial cells and their lumina with respect to their area, boundary length and Euler number per unit area. The central finding was a highly significant negative correlation between the Gleason score and the Euler number of the epithelial cell phase per unit area. The Gleason score of all individual cases was predicted from the spatial statistical variables by multivariate linear regression. This approach means to perform a multiclass pattern recognition, as opposed to the usual problem of binary pattern recognition. A prediction was considered as acceptable when its deviation from the human classification was no more than 1 point. This was achieved in 79 of these 103 cases when only the Euler number density was used as predictor variable. The accuracy could be risen slightly to 84 of the 103 cases, when 7 input variables were used for prediction of the Gleason score, which means an accuracy of 81.5%.

Keywords: classification, pattern recognition, prostate cancer, regression, spatial statistics.

INTRODUCTION

Staging and grading are of central importance for treatment decisions and prognostication of prostate cancer. For staging, the TNM classification of the UICC is the established standard (Sobin *et al.*, 2009). As grading procedure, the Gleason grading system is usually recommended (see, *e.g.*, Murphy *et al.*, 1994; Amin *et al.*, 2004; Eble *et al.*, 2004). According to this system, the histological textures within the tumour are evaluated at low magnification and graded in values from 1–5. The Gleason score is obtained by summing up the grades of the two most dominant textures (Gleason, 1966; 1992; Amin *et al.*, 2004; Eble *et al.*, 2004). Hence a Gleason score is an integer number in the interval [2,10]. While staging of prostatic carcinomas according to the TNM-scheme is usually considered as highly reproducible, there remains admittedly a subjective element in Gleason grading. This may lead to intra-observer and inter-observer variability, when the same cases are examined twice by the same observer or by two different observers, respectively.

In previous investigations, it has been shown that the texture of prostatic tissue, as seen at low magnification, may be characterized quantitatively in

terms of spatial statistics and stereology (Mattfeldt *et al.*, 1999; 2000; 2003; Mattfeldt, 2003). Basically, prostatic carcinoma tissue may be subdivided into three phases, namely the epithelial cells (the tumour cells), the lumina, and the stroma, which together account for 100% of the tumour tissue. Applying established methods of spatial statistics to digitized binary images, it is possible to characterize these phases quantitatively in terms of area, boundary length and Euler number per unit tissue area (see, *e.g.*, Ohser and Mücklich, 2000). In particular, the Euler number appears highly attractive for a quantitative characterization of prostatic carcinomas. The Euler number χ of a set of geometrical objects is the number of separate objects minus the number of holes in them (Fig. 1). Hence the Euler number should be directly linked to fundamental pathological tumor properties such as solid architecture where ideally $\chi > 0$ (epithelial blocks without holes), tubular architecture where $\chi \approx 0$ (approximately one hole per block), and cribriform architecture where ideally $\chi < 0$ (many holes inside a block; Mattfeldt *et al.*, 2007b).

A look at the well-known schematic images of the Gleason grades shows that grades 1–2 have a tubular differentiation, grade 3 consists of a mixture of tubular and cribriform (sieve-like) structures, grade

4 is predominantly cribriform, and grade 5 contains cribriform and solid patterns (Gleason, 1966; 1992; Amin *et al.*, 2004; Eble *et al.*, 2004). Hence, it seems plausible to characterize the texture of prostatic carcinomas in terms of the Euler number of the phase of the epithelial cells per unit area, in addition to other model parameters, in relation to the Gleason score. In this investigation, we wanted to find answers to the following questions: (i) Which objective quantitative textural changes occur in prostatic tissue with progression from low to high Gleason scores? (ii) To which extent is it possible to predict the Gleason score from such data in individual cases? In an earlier paper with less cases, we had tried to predict the stage of such tumours in a *binary* manner (pT2 vs. pT3) from similar data (Mattfeldt *et al.*, 2003). In the present paper, we deal with the prediction of the Gleason score, which means *multiclass* pattern recognition from spatial data, which is new as far as we can determine. Moreover, the number of cases has been increased, and tumours of stage T1 have been newly included.

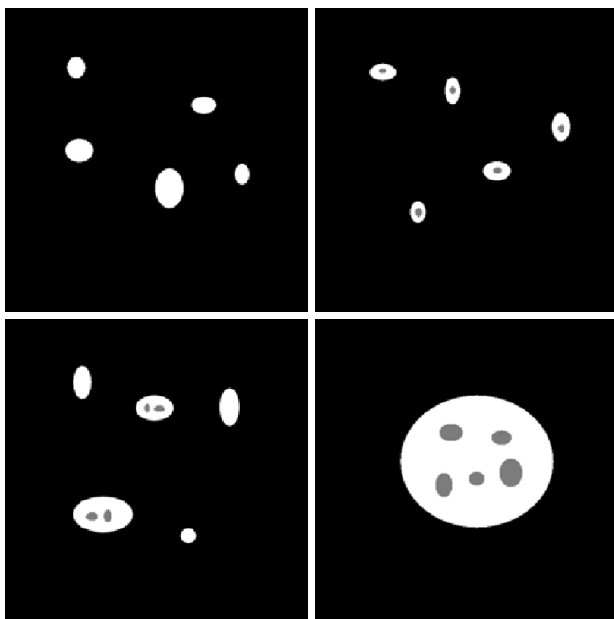


Fig. 1. Schematic illustration of the Euler number for a triphasic structure with the phases: epithelial cells (white), lumen (grey) and stroma (black). The Euler number χ is the number of individual structures minus the number of holes in them. For the white phase, we have in the upper left panel $\chi = 5$, in the upper right panel $\chi = 5 - 5 = 0$, in the lower left panel $\chi = 5 - 4 = 1$, and in the lower right panel $\chi = 1 - 5 = -4$.

MATERIAL AND METHODS

PATIENTS

The investigations were performed on prostatic carcinoma specimens of 103 patients from 3 groups. A first group (the T1-group) consisted of 25 cases of incidental prostatic carcinoma, diagnosed from transurethral resection material or from surgical resection specimens (adenomectomies) that had been removed because of a diagnosis of benign prostatic hyperplasia; these had been classified as T1a or T1b, respectively. A second group (the pT2-group) consisted of 50 primary prostatic adenocarcinomas with TNM classification pT2N0. A third group (the pT3-group) consisted of 28 prostatic adenocarcinomas classified as pT3N0. The primary tumor specimens in the pT2- and pT3-groups were radical prostatectomy specimens. In fact, the cases of the pT2- and pT3-groups had also been used in our previous study on the prediction of prostatic carcinoma stage on the basis of stereological data and CGH findings (Mattfeldt *et al.*, 2003). The tumor-bearing slides of all prostatectomy specimens had been evaluated by the first author with respect to the Gleason score prior to the investigation. The mean Gleason score in the T1 group was 4.16 (SD: 1.30), in the pT2 group it was 6.18 (SD: 0.92), and in the pT3 group it was 7.11 (SD: 0.99).

SPATIAL STATISTICS

In diagnostic histopathology, tissue sections are studied under light microscopical view. If it comes to quantitation, rectangular or quadratic windows are usually superimposed onto the sections. In the first instance, one is hence faced with planar textures. These may be interpreted as realizations of planar random closed sets (RACS) restricted to rectangular observation windows. As numerical descriptors for this type of data we will estimate

- A_A , the mean area of the interesting phase per unit reference area (area fraction),
- B_A , the mean boundary length of the interesting phase per unit reference area, and
- χ_A , the mean Euler number of the interesting phase per unit reference area.

These three model parameters are also denoted as the *specific intrinsic volumes* of the RACS in the plane. In order for the specific intrinsic volumes to be well-defined, the RACS needs to satisfy certain conditions. Hence, referring to Schneider and Weil (2000), Theorem 5.1.3, for our investigations we assume that our data sets may be viewed as realizations of RACSs with the following properties:

- The RACSs are stationary, *i.e.*, their distribution is invariant w.r.t. translations of the origin.
- With probability 1, the RACSs have realizations in the extended convex ring, *i.e.*, the restrictions of the RACSs to any compact and convex observation window K can be decomposed into finitely many compact and convex subsets.
- If $N([0, 1]^2)$ denotes the (random) minimum number of sets in such a decomposition w.r.t. the unit cube, $2^{N([0, 1]^2)}$ has finite expectation.

Since we are working with bounded pixel images, the degree to which these assumptions reflect the true nature of the tissue must remain unclear. However, the quite central stationarity assumption seems to be realistic for the inner parts of the tissues captured by the observation windows.

For the purpose of this investigation, paraffin sections of a nominal thickness of 5 μm stained with Haematoxylin-Eosin were used. From these sections, visual fields containing tumour tissue were selected according to technical quality criteria at an objective magnification of 10 \times . In practice, all sections were looked through, and the a fixed number of the first visual fields with clearly discernible tumour tissue and without artifacts were used for the study. In the T1 group 5 fields were studied, because the amount of tumour tissue was often rather sparse in these cases where only parts of the prostate glands had been removed, whereas 10 fields were examined in the pT2 and pT3 groups, where complete specimens of the complete prostate gland were available. The images had the size 510 \times 510 pixels (*i.e.*, 512 \times 512 pixels including a non-informative black border of 1 pixel width on each edge). They were acquired with a CCD camera connected to a Zeiss light microscope and transferred to the image analysis system Kontron IBAS 2000, where they were interactively segmented into the three phases: epithelial cells (tumour cells), lumina, and stroma (Figs. 2,3). Interactive segmentation consisted in tracing the profiles of the lumina and epithelial blocks with the electronic cursor on the digitizing tablet of the aforementioned Kontron system under visual control on a monitor; the remainder was considered as the stromal component. As the contrast between these components is very high (see Figs. 2,3), the risk of subjectivity is very low according to our experience from previous studies (Mattfeldt *et al.*, 1999, 2001, 2003).

The segmentation was performed by a technician, who was not provided with information on the Gleason score. The segmented images were transferred to a PC and converted to binary images containing i. only the luminal phase and its complement, ii. only the

epithelial cell phase and its complement, and iii. only the stromal phase and its complement. Finally, the resulting binary images were evaluated using routines of the software package Geostoch, a Java based open library system (Mayer *et al.*, 2004). In this package, the routine ‘Measure2D’ was used, which is based on established algorithms for the estimation of specific intrinsic volumes from digitized binary images in 2D (see Ohser and Mücklich, 2000, section 4.2, pp. 124–133). Thus, one obtains estimates of A_A , B_A and χ_A for the aforementioned phases i.–iii. of every image, *i.e.*, 9 specific intrinsic volumes are estimated per image in the first instance. To characterize the individual tumours, arithmetic mean values were computed between these estimates for the 5–10 images per case. The final magnification corresponded to a width of 0.4 mm of the quadratic visual field at the scale of the tissue.

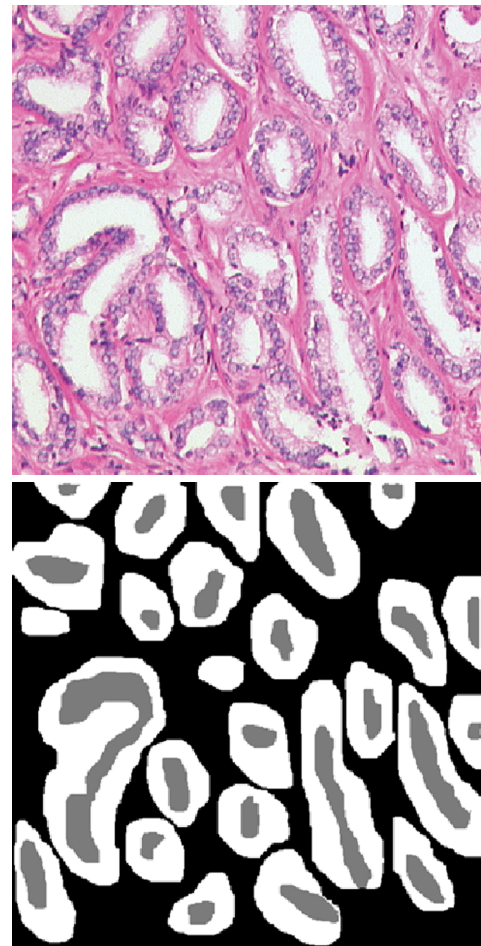


Fig. 2. Upper panel: visual field from a prostatic adenocarcinoma with tubular differentiation (primary and secondary Gleason grade 3). Haematoxylin-Eosin stain. Its segmentation leads to the lower panel, which contains the three phases: white – epithelial cells, gray – lumen, black – stroma. Gleason score 3+3=6.

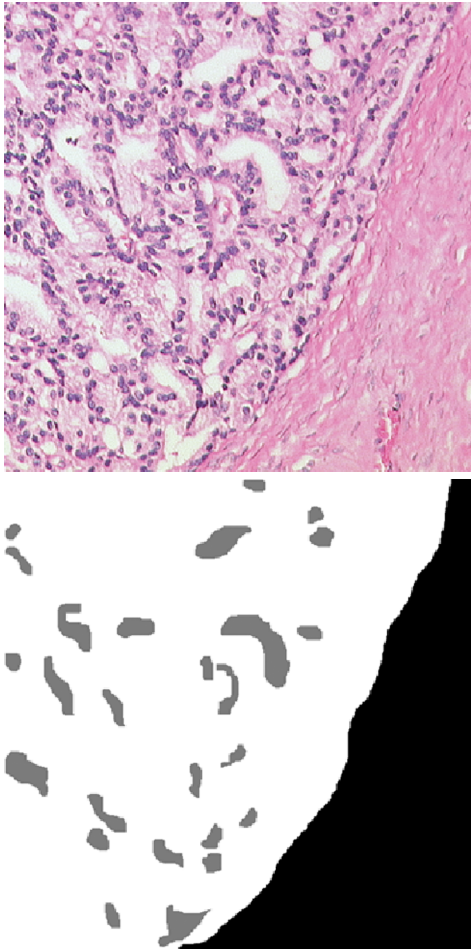


Fig. 3. Upper panel: visual field from a prostatic adenocarcinoma with cribriform differentiation (primary and secondary Gleason grade 4). Lower panel: the same image after segmentation. Haematoxylin-Eosin stain. Gleason score $4+4=8$.

STATISTICAL METHODS

VARIABLE SELECTION

An attempt was made to find the best combination of input variables for the classification (pattern recognition) of the cases with respect to the Gleason score (see below). To this aim, we applied *multivariate linear regression* of 9 influence variables (the 3 estimates \hat{A}_A , \hat{B}_A and $\hat{\chi}_A$ of the 3 main phases, *i.e.*, epithelial cells, lumen, and stroma) on the Gleason score as the dependent variable. In this case, we suppose that the output variable is a linear combination of the influence variables plus a random error term. Using this approach, the Gleason score is considered as a real variable, although in fact it takes only integer values, for the sake of simplification. To implement this approach, the ‘reg’ procedure of the program package SAS was used with the ‘maxr’ option. In

this mode the program performs multivariate linear regressions of all the 9 influence variables on the Gleason score and tries to find the best combination for each prescribed number of variables (1–9).

PREDICTION (MULTICLASS PATTERN RECOGNITION)

It was attempted to predict the Gleason score from quantitative image characteristics of individual cases by computer. This approach may be considered as an example of *statistical learning* with a supervised learning rule. The computer learns by training from preclassified cases, from which the set of input variables and the output variable (the Gleason score) are known. As input data, those combinations of the variables were selected that provided the best linear regression models for fixed numbers of input variables from 1 to 9. The output variable is the Gleason score, *i.e.*, an integer number in the interval [2,10]. Hence, we deal with a problem of *multiclass pattern recognition*. The values of the independent variables were inserted into the fitted regression equation. The value obtained by rounding this result to the next integer was then taken as the estimate of the Gleason score of the test case.

CROSS-VALIDATION

In our retrospective data set, the accuracy of prediction of the output data from the input data in the test phase was examined by *cross-validation*. This concept means that the total set of n cases is partitioned into a subgroup of $n - k$ cases (the training cases) and another subgroup which consists of the k remaining cases (the test cases). In the training phase, the algorithm ‘learns’ to estimate the output variable from the input variables within the training group. In the test phase thereafter, the output variable of the test cases is estimated from the input variables of the test cases making use of the information learnt previously from the training group. This strategy simulates a confrontation of the algorithm with a new case, and by this manner one tests its ability to generalize. When the number of cases is large, it is possible to use, *e.g.*, 25–33% of the cases as test cases. If the number of cases is relatively small (*e.g.*, $n \approx 100$), it is often recommended to choose $k = 1$, *i.e.*, to apply the *leave-one-out principle* (synonyms: jackknife, round-robin) (Tourassi and Floyd, 1997; Vapnik, 1998). The latter approach was also used in the present study. The prediction is repeated cyclically for every patient as test case with the complementary set of cases serving as its training group.

Table 1. Group comparisons.

Group	1 (GS 2–4)		2 (GS 5–7)		3 (GS 8–10)	
Number of cases	13		79		11	
Variable	Mean	SD	Mean	SD	Mean	SD
$A_A(\text{epi})$	0.3452	0.0921	0.4282	0.0756	0.5405	0.1161
$A_A(\text{lumen})$	0.0876	0.0485	0.1132	0.0443	0.0917	0.0638
$A_A(\text{stroma})$	0.5670	0.1130	0.4584	0.0902	0.3676	0.1413
$B_A(\text{epi})$ [mm/mm ²]	36.9211	7.9281	48.8251	7.4582	39.9074	10.0248
$B_A(\text{lumen})$ [mm/mm ²]	11.4390	2.5549	17.4614	4.3937	18.1943	8.4128
$B_A(\text{stroma})$ [mm/mm ²]	25.4924	6.4175	31.3801	5.3632	21.7252	11.0343
$\chi_A(\text{epi})$ [mm ⁻²]	-13.7379	27.5399	-77.8718	61.9519	-186.0369	140.6322
$\chi_A(\text{lumen})$ [mm ⁻²]	109.0144	41.1674	167.5646	52.3744	221.7329	117.2564
$\chi_A(\text{stroma})$ [mm ⁻²]	-109.4831	42.0575	-128.9811	44.2810	-69.7869	94.8673

RESULTS

GROUP COMPARISONS

The cases were sorted into three groups with respect to the Gleason score: group I with low scores (2–4), group II with intermediate scores (5–7), and group III with high scores (8–10). The results are shown in Table I. The group mean values were tested for significant differences by pairwise *t*-tests between group I and II, and between group II and III.

RESULTS FOR THE EPITHELIAL CELL PHASE

The area fraction of epithelial cells rose highly significantly with increasing Gleason score ($p < 0.0001$) (Fig. 4, upper panel). The boundary length density of the epithelial cells rose significantly from group I to group II ($p < 0.0001$), but fell significantly from group II to group III ($p < 0.001$) (Fig. 4, middle panel). The Euler number density of the epithelial phase attained a slightly negative value already in group I, and declined to more and more negative values in groups II and III ($p < 0.0001$) (Fig. 4, lower panel).

RESULTS FOR THE LUMINAL PHASE

There was an increase of the luminal boundary length density in cases of Group II in comparison to group I ($p < 0.001$) (Fig. 5, upper panel). The Euler number density of the lumina remained positive throughout all groups. It rose significantly from group I to group II ($p < 0.001$) and from group II to group III ($p < 0.001$) (Fig. 5, lower panel).

RESULTS FOR THE STROMAL PHASE

The area fraction of the stroma declined highly significantly through all groups with increasing Gleason scores ($p < 0.01$). The boundary length density of the stromal phase increased highly significantly from group 1 to group 2 ($p < 0.001$) and thereafter decreased significantly from group 2 to group 3 ($p < 0.001$). The Euler number density of the stromal phase assumed strongly negative values in all 3 groups. It moved towards less negative values from group II to group III ($p < 0.001$), which reflects a decrease of the number of epithelial units (union of epithelial cells and lumen, *i.e.*, ‘holes’ from the viewpoint of the stroma) inside the stroma (Fig. 6).

CORRELATION ANALYSIS

Linear correlation analysis of the data revealed a highly significant positive correlation of the Gleason score with the area fraction of the epithelial phase ($r = 0.4380, p < 0.0001$). The Gleason score was also correlated positively with the boundary length density of the luminal phase ($r = 0.3798, p < 0.0001$). There were no significant correlations between the Gleason score and the area fraction of the luminal phase and with the boundary length densities of the epithelial cell phase, respectively.

A highly significant negative correlation was found between the Gleason score and the Euler number density of the epithelial cell phase per unit area ($r = -0.5284, p < 0.0001$). There was also a significant positive correlation of the Gleason score with the Euler number density of the lumina ($r = 0.4390, p < 0.0001$).

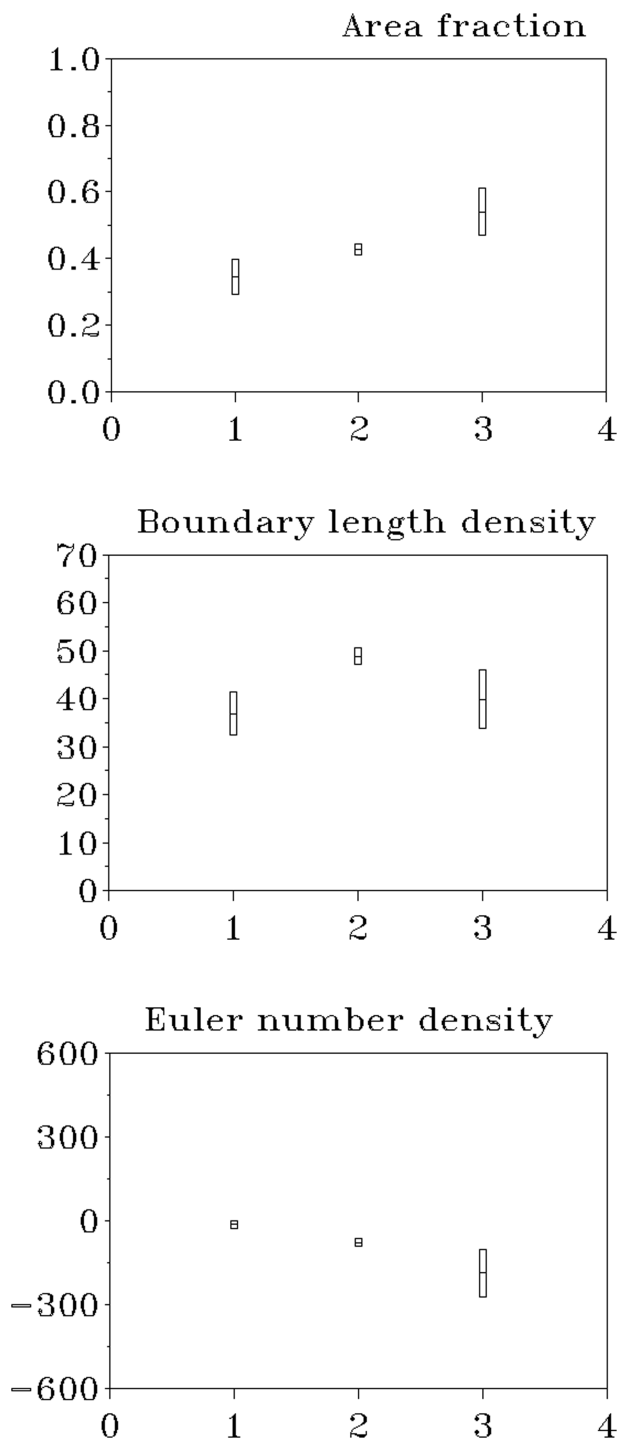


Fig. 4. Results for the epithelial cell phase. Abscissa value 1: Gleason score 2–4, 2: Gleason score 5–7, 3: Gleason score 8–10. Indicated are group mean values and bounds of 95% confidence intervals. Upper panel: Area fraction of epithelial cells in the three groups. Middle panel: Boundary length of epithelial cells per unit area in the same groups. Lower panel: Euler number of epithelial cells per unit area in the same groups. This parameter shows a highly significant decrease with increasing Gleason score.

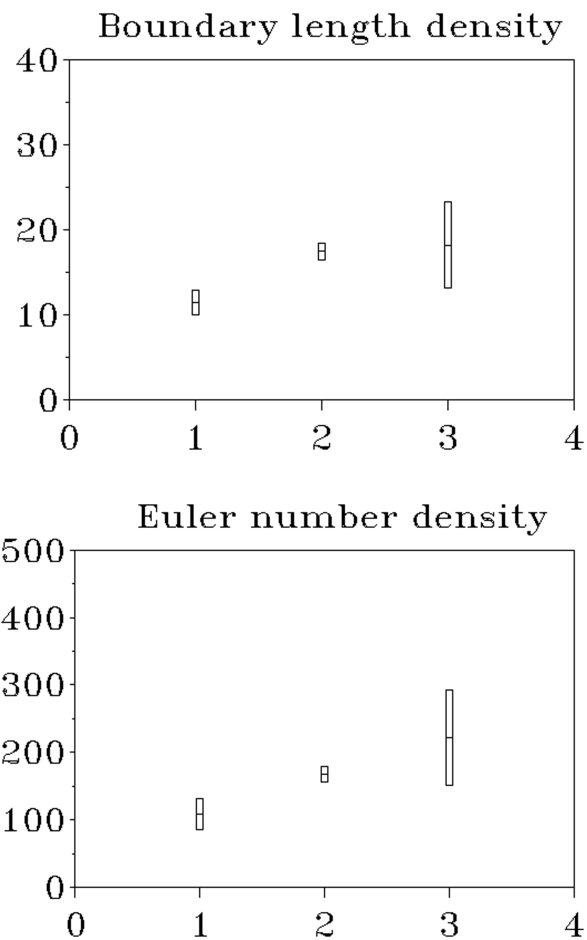


Fig. 5. Results for the luminal phase. Upper panel: Boundary length of lumina per unit area in the same groups. The increase from group 1 to group 2 is significant. Lower panel: Euler number of lumina per unit area in the same three groups. This parameter shows a significant increase with increasing Gleason scores.

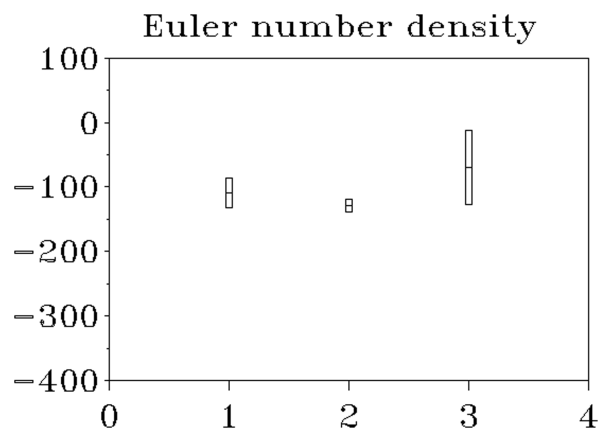


Fig. 6. Results for the stromal phase. Euler number of stroma per unit area in the same three groups.

PREDICTION

In our cross-validation study with the leave-one-out scheme, the Gleason score of each individual case was predicted by linear regression and rounding to the next integer, as outlined above, with cyclically training on 102 cases and testing on 1 case. Usually, in studies on observer variability of the Gleason score, a classification is considered as acceptable, when two judgments of a case differ by no more than 1 point (see, e.g., Bostwick, 1994). Using this criterium, we found that the Gleason score of 79 of the 103 cases (*i.e.*, 76.6%) could already be predicted sufficiently on the basis of only one parameter: the Euler number density χ_A of the epithelial cell phase. The highest accuracy of prediction was found when the following 7 variables were included into the regression model: Euler number density, area fraction and boundary length density of the epithelial cell phase; Euler number density and boundary length density of the luminal phase; Euler number density and boundary length density of the stroma. When the aforementioned criterium of sufficient accuracy at a discrepancy ≤ 1 was adopted, 84/103 predictions (81.5%) were considered as acceptable. The correlation coefficient between the predicted Gleason score and the preclassified Gleason score was $r = 0.6059$ ($p < 0.0001$) in the best model.

DISCUSSION

TISSUE CHANGES WITH INCREASING GLEASON SCORE

Let us first consider the changes of the elementary parameters: area fraction and boundary length density. The area fraction of the epithelial cells rose with increasing Gleason score. This is very plausible when one considers the large epithelial areas devoid of stroma in high grade prostatic carcinomas. The area fraction of the luminal phase remained nearly constant with increasing Gleason scores. This finding was, however, accompanied by a rise of the boundary length of the luminal phase per unit tissue area. If the boundary length of the lumina rises despite an unchanged area fraction, this means that a geometrical change must have occurred by which boundary length has been gained. It means that the boundary-to-area ratio of the luminal phase must have increased. This may occur if either the luminal units, maintaining their shape, become smaller and increase in number, or if the luminal units change to more elongated or wrinkled shapes without increasing their number. Clearly, also a combination of mechanisms can account for the phenomenon. All three mechanisms

augment boundary length but keep the area fraction constant. That the first effect is stronger is suggested by the finding that the Euler number of the luminal phase per unit area increased with the Gleason score.

The geometrical meaning of A_A and B_A is intuitively obvious. The Euler number density χ_A is more complex. It is the number of units minus the number of holes in them per unit area. The main finding of this study was a rather strong negative correlation ($r = -0.5284$, $p < 0.0001$) between the Euler number density of the epithelial cell phase and the Gleason score. The epithelial cells form a complex phase, which essentially has two disjoint boundaries: an outer boundary, directed towards the stroma, and an inner boundary, directed towards the lumina. The Euler number density of the epithelial cell phase is influenced by both components. First, the data show that an increase to large Gleason scores led to a decrease of the Euler number density of the epithelial blocks (including cells and lumina). This change is also reflected in the transition of the Euler number density of the stroma to less negative values in the step from intermediate to very high Gleason scores. Hence, we have a reduction of the Euler number density 'from outwards'. Second, there was a strong positive correlation of the Gleason score with the Euler number density of the lumina. Each newly formed lumen reduces the Euler number of the epithelial cell phase by 1. Hence, the increase of the Euler number density of the lumina led to a further decrease of the Euler number density of the epithelial phase 'from inwards'. To sum up, the strong negative correlation of χ_A of the epithelial phase to the Gleason score is due to a decrease of the number of the epithelial tumour tissue units per unit area as a whole, and to the formation of an increasing number of epithelial lumina inside them.

The reader will have noted that a highly significant negative correlation between the Gleason score and the Euler number of the epithelial phase per area could be shown, but nevertheless the absolute value of the correlation coefficient was not very high ($r = -0.5284$). Various causes can be discussed to account for this finding. On one hand, a sampling error must be considered: for statistical learning, 5–10 visual fields were selected, whereas the human decision on a Gleason score is based on the evaluation of the whole section. Furthermore, one has to account for the inherently subjective nature of Gleason grading. There is another aspect which becomes apparent when one considers the well-known schematic diagrams of the Gleason grades 1–5. In grade 1–4 tumours, a gradual transition occurs from purely tubular to more and more cribriform structures, which leads to an expected transition from positive to negative Euler

numbers of the epithelial phase per area, as outlined above. Gleason grade 5 patterns are characterized by the development of solid parts, *i.e.*, unstructured epithelial blocks without lumina at all, in addition to the cribriform component which further persists. Such solid structures will tend to increase the Euler number per area a little. This means that, even under ideal conditions, if no sampling error occurs and subjective errors are kept to a minimum, there will be no absolutely linear relation between the Euler number of the epithelial phase per area and the Gleason score any more, when grade 5 patterns are present. Fortunately, this effect is probably not very relevant in reality due to the following considerations. A grade 5 pattern usually occurs in a case with Gleason score 9 or 10. If the Gleason score is 9, this means that grades 4 and 5 dominate, hence a strong cribriform component is still present due to the contribution of grade 4. Even in cases with Gleason score 10, the cribriform variety is rarely totally lost, as the original drawing of Gleason shows (Gleason, 1966; 1992). Moreover, cases with Gleason scores ≥ 9 are generally rare; in our unselected case series, there were only 4 cases with Gleason score 9 (3.9%), and there was no case with Gleason score 10. These data are in accordance with larger series of prostatectomy specimens, where only 2.8% and 0.05% of cases were found with Gleason scores 9 and 10, respectively (Amin *et al.*, 2004). These facts imply that in general, the negative correlation of the Gleason score to the Euler number of the epithelial cell phase per area should hold in good approximation for the whole spectrum of cases.

PREDICTION

A major result of the present study was the finding that in 84/103 cases (81.5%), the Gleason score of a prostatic carcinoma could be successfully predicted from a set of 7 variables of spatial statistics. The criterium for success was that machine prediction and human classification differed by no more than 1 point. Using the aforementioned 7 variables for prediction, only 19/103 cases were insufficiently predicted. It was looked up whether they were systematically undergraded or overgraded. Overgrading was found in 9, and undergrading was found in 10 of these 19 cases. We conclude that the statistical learning method leads neither to a systematic overgrading nor to a systematic undergrading.

Overfitting is a well-known trap in pattern recognition studies. It is likely to occur if too many input variables are used, which makes the model too complex. In this case, one may obtain good results in the training phase, but the system is characteristically

unable to generalize to new cases. Here a cross-validation step was performed to avoid this pitfall. Slight overfitting was seen to emerge when the number of input variables was increased from 1 to 2, where the accuracy of prediction sank slightly despite using additional input information: with 2 input variables, the number of sufficiently classified variables decreased from 79 to 76; a similar decrease was found when the number of input variables was increased from 7 to 8, *i.e.*, a decrease from 84 to 81 acceptably classified cases.

We would like to stress that the type of cross-validation that we used in this investigation – the leave-one-out approach – seems particularly suitable for applications in histopathology. It simulates the situation that an observer has gained experience in a certain number of cases, and on the basis of this learning process he is confronted with a single new case he has to classify. This way of learning is quite analogous to histopathological diagnostics, where the pathologist learns to generalize from multiple similar cases to a new case. In fact, it has been advocated to train Gleason grading by studying preclassified cases, published, *e.g.*, in textbooks (Amin *et al.*, 2004) and on the internet (see <http://217.8.156.155/norcyt/prostata/PROST.htm>).

With regards to a potential application in practice, our method using spatial statistics applied to interactively segmented images is clearly too laborious for everyday use. However, it becomes increasingly common to work with *virtual microscopic slides*. These are large image files generated by a computer linked to a conventional light microscope which scans the physical microscopic slide completely and fully automatically. If the epithelial phase could be reproducibly segmented, *e.g.*, by an immunohistochemical stain with an antibody that detects specifically the tumour cells, the whole texture analysis as described here could be performed fully automatically.

METHODOLOGICAL ASPECTS

PREDICTION METHODS

Linear regression methods are clearly not the only way to perform a multiclass pattern recognition of the Gleason score from spatial data. Alternatively, one could try to work with robust nonparametric methods of prediction, which do not presuppose a linear model assumption in the relation between the influence variables and the dependent variable. In fact, such a nonlinear behaviour could be observed, *e.g.*, for the

variable B_A of lumen per tissue, which rose from group I to group II and declined in group III as compared to group II. In this context, *artificial neural networks* appear as an attractive alternative, *e.g.*, multilayer feedforward networks with backpropagation, learning vector quantization (LVQ) or support vector machines (SVM; Kohonen *et al.*, 1996; Burges, 1998; Saunders *et al.*, 1998). Such neural paradigms have been used by our group to predict various properties of prostatic carcinomas from input data sets (Mattfeldt *et al.*, 1999; 2001; 2003). For example, it was tried to predict from a set of input variables whether a prostatic carcinoma was still confined to the prostate (stage pT2), or had already extended beyond the organ (stage \geq pT3; Mattfeldt *et al.*, 2001; 2003). In this case, we are faced with *binary* pattern recognition. The same is true when it is tried to predict a relapse from primary tumour data, see Mattfeldt *et al.* (1999). In another study, it was attempted to predict the Gleason score from spatial data in a binary manner, *e.g.*, Gleason score < 7 versus Gleason score ≥ 7 (Wittke *et al.*, 2007). As far as we could determine, however, *multiclass* prediction of the Gleason score on the basis of spatial statistical data has not been performed before.

It is easily possible to adapt the aforementioned algorithms LVQ and SVM in such a manner that they learn to classify items with an ordinal dependent variable (Kohonen *et al.*, 1996; Burges, 1998; Saunders *et al.*, 1998). For example, this option is provided for support vector machines as the ‘multiclass pattern recognition mode’. In further research work, it will be examined whether these paradigms may lead to an increase of the predictive accuracy.

SPATIAL STATISTICS

In the present study, the Gleason scores were characterized in terms of planar spatial statistics by estimating specific intrinsic volumes of tissue phases, which were basically considered as *random closed sets with positive area*. While this is a well established field of spatial statistics, much more work has been done in the field of the statistical analysis and modelling of *spatial point patterns* (see, *e.g.*, Illian *et al.*, 2008). Such an approach is also feasible in the case of prostatic cancer, *e.g.*, by studying the point patterns of the tumour cell nucleus profiles. Such patterns may be characterized nonparametrically in terms of first and second order properties. It is also possible to fit parametric point process models to such patterns, *e.g.*, Gibbs processes (Mattfeldt *et al.*, 2007a; Illian *et al.*, 2008). It must however be kept in mind that the present approach based on volume processes is more natural when it is intended to predict

the Gleason score from image data, as this method of grading means to focus entirely on the texture and to disregard all nuclear changes deliberately. Pure point process statistics would imply that an important property — the topology of the epithelial cell phase — is neglected. Nevertheless, point process statistics could bring potentially valuable additional information, which is not reflected by the Gleason score. In this context we mention grading systems for prostate cancer which have been suggested as alternatives to the Gleason grading system, see, *e.g.*, (Mostofi, 1975; Böcking *et al.*, 1982; Bostwick, 1994). In contrast to Gleason grading, findings on changes of the tumour cell nuclei are considered in these systems. Here we have concentrated only on Gleason grading, because it is the standard procedure favoured by most urologists and recommended by a WHO consensus conference since 1993 (Murphy *et al.*, 1994).

STEREOLOGY

Based on the following established estimators from stereology, the measured morphological changes in 2D sections of prostatic tissue may be related to alterations of 3D tumour morphology. One has

$$\hat{V}_V = A_A \quad (1a)$$

$$\hat{S}_V = \frac{4}{\pi} B_A \quad (1b)$$

$$\hat{M}_V = 2\pi\chi_A \quad (1c)$$

where V_V is the volume fraction, S_V is the mean surface area per unit reference volume, and M_V is the curvature density (integral of mean curvature per unit volume of the corresponding 3D RACS, see Weibel, 1980, pp. 84–101, Fig. 3.16); by \hat{V}_V , \hat{S}_V and \hat{M}_V we denote the estimators of these quantities. In addition to the assumptions made above, which ensure the intrinsic volumes of RACS to be well-defined, these stereological estimators are only unbiased if the RACS is isotropic (see, *e.g.*, Kiderlen, 2010, p. 37). This appears to be a reasonable assumption for our data. Especially the stereological link (1c) between the mean Euler number χ_A and the curvature density M_V is quite instructive: It allows to interpret decreasing values of χ_A (and thus M_V) as a transformation of the 3D tissue surface, where locally convex parts are reduced in favor of a locally concave geometry, which is, *e.g.*, represented by infoldings and holes.

In terms of automated Gleason grading, the use of stereological formulas as above can obviously not be expected to improve the quality of prediction, since the 3D stereological estimates are related to the measured 2D characteristics by multiplication of constant factors.

Stereology also provides methods for the estimation of 2D intrinsic volumes from micrographs, thus presenting alternatives to the quantitation techniques used in this study (Weibel, 1980, pp. 97–101; Stoyan *et al.*, 1995, eqs. 7.3.8, 7.3.9 and 11.2.5, Fig. 7.2). The stereological methods have been implemented in various software packages for the analysis of digitized images. The non-stereological estimation technique applied in our study is however quite appropriate for a fully automatic machine learning approach to Gleason grading that does not require any user interaction. Thus, once image segmentation of the different tissue phases can be done in an automatic way, the proposed method of algorithmic Gleason grading can be conducted in a highly efficient way.

ACKNOWLEDGMENTS

Thanks are due to Michael Held and Rolf Kunft for technical assistance.

REFERENCES

- Amin MB, Grignon DJ, Humphrey PA, Srigley JR (2004). Gleason grading of prostatic cancer. A contemporary approach. Philadelphia: Willincott.
- Böcking A, Kiehn J, Heinzl-Wach M (1982). Combined histologic grading of prostatic carcinoma. *Cancer* 50:288–94.
- Bostwick DG (1994). Grading prostate cancer. *Am J Clin Pathol* 102(Suppl 1):38–56.
- Burges JC (1998). A tutorial on support vector machines for pattern recognition. *Data Mining Knowl Discov* 2:121–67.
- Eble JN, Sauter G, Epstein JI, Sesterhenn IA (2004). Pathology and genetics of tumours of the urinary system and male genital organs. Lyon: IARC Press.
- Gleason DF (1966). Classification of prostatic carcinoma. *Cancer Chemother Rep* 50:125–8.
- Gleason DF (1992). Histologic grading of prostate cancer: a perspective. *Hum Pathol* 23:273–9.
- Berner A, Busch C, Halvorsen OJ, Haugen OA, Scott H, Sund S, Svindland A: Web-training set for Gleason grading. The Norwegian study group for Prostate Cancer (NUCG).
- Illian J, Penttinen A, Stoyan H, Stoyan D (2008). Statistical analysis and modelling of spatial point patterns. Chichester: Wiley.
- Kiderlen M (2010). Introduction into integral geometry and stereology. In: Spodarev E, Ed. Stochastic geometry, spatial statistics and random fields. *Lect Notes Math* 2068:21–48.
- Kohonen T, Hynninen J, Kangas J, Laaksonen J, Torkkola K (1996). LVQ_PAK: The learning vector quantization program package. Technical Report A30, Helsinki University of Technology, Laboratory of Computer and Information Science, Otaniemi, Finland.
- Mattfeldt T (2003). Classification of binary spatial textures using stochastic geometry, nonlinear deterministic analysis and artificial neural networks. *Int J Pattern Recogn* 17:275–300.
- Mattfeldt T, Kestler HA, Hautmann R, Gottfried HW (1999). Prediction of prostatic cancer progression after radical prostatectomy using artificial neural networks: a feasibility study. *BJU Int* 84:316–23.
- Mattfeldt T, Gottfried H-W, Schmidt V, Kestler HA (2000). Classification of spatial textures in benign and cancerous glandular tissues by stereology and stochastic geometry using artificial neural networks. *J Microsc* 198:143–58.
- Mattfeldt T, Kestler HA, Hautmann R, Gottfried H-W (2001). Systematic biopsy-based staging of prostatic carcinoma using artificial neural networks. *Eur Urol* 39:530–7.
- Mattfeldt T, Gottfried H-W, Wolter H, Schmidt V, Kestler HA, Mayer J (2003). Classification of prostatic carcinoma with artificial neural networks using comparative genomic hybridization and quantitative stereological data. *Pathol Res Pract* 199:773–84.
- Mattfeldt T, Eckel S, Fleischer F, Schmidt V (2007a). Statistical modelling of the geometry of planar sections of prostatic capillaries on the basis of stationary Strauss hard-core processes. *J Microsc* 228:272–81.
- Mattfeldt T, Meschenmoser D, Pantle U, Schmidt V (2007b). Characterization of mammary gland tissue using joint estimators of Minkowski functionals. *Image Anal Stereol* 26:13–22.
- Mayer J, Schmidt V, Schweiggert F (2004). A unified simulation framework for spatial stochastic models. *Simul Model Pract Th* 12:307–26.
- Mostofi FK (1975). Grading of prostatic carcinoma. *Cancer Chemother Rep* 59(I):111–7.
- Murphy GP, Busch C, Abrahamsson PA, Epstein JI, McNeal JE, Miller GJ, Mostofi FK, Nagle RB, Nordling S, Parkinson C (1994). Histopathology of localized prostate cancer. Consensus conference on diagnosis and prognostic parameters in localized prostate cancer. Stockholm, Sweden, May 12–13, 1993. *Scand J Urol Nephrol Suppl* 162:7–42.
- Ohser J, Mücklich F (2000). Statistical Analysis of Microstructures in Materials Science. Chichester: Wiley.
- Saunders R, Stitson MO, Weston J, Bottou L, Schölkopf B, Smola A (1998). Support vector machine reference

- manual. Technical Report. Royal Holloway, University of London.
- Schneider R, Weil W (2000). *Stochastische Geometrie*. Stuttgart: Teubner.
- Sobin LH, Gospodarowicz MK, Wittekind C, Eds. (2009). *TNM classification of malignant tumours*. Wiley: New York.
- Stoyan D, Kendall WS, Mecke J (1995). *Stochastic Geometry and Its Applications*, 2nd Ed. Wiley: Chichester.
- Tourassi GD, Floyd CE (1997). The effect of data sampling on the performance evaluation of artificial neural networks in medical diagnosis. *Med Decis Making* 17:186–92. .
- Vapnik VN (1998). *Statistical Learning Theory*. Wiley: New York.
- Weibel ER (1980). *Stereological Methods. II. Theoretical Foundations*. London: Academic Press.
- Wittke C, Mayer J, Schweiggert F (2007) On the classification of prostate carcinoma with methods from spatial statistics. *IEEE Trans Inf Technol Biomed* 11:406–14. .