

2009

Automatic Diagnosis for Prostate Cancer Using Run-Length Matrix Method

Xiaoyan Sun
Old Dominion University

Shao-Hui Chuang
Old Dominion University

Jiang Li
Old Dominion University, jli@odu.edu

Frederic McKenzie
Old Dominion University, rdmckenz@odu.edu

Nico Karssemeijer (Ed.)

See next page for additional authors

Follow this and additional works at: https://digitalcommons.odu.edu/ece_fac_pubs



Part of the [Bioimaging and Biomedical Optics Commons](#), [Biomedical Commons](#), [Diseases Commons](#), [Investigative Techniques Commons](#), and the [Urology Commons](#)

Original Publication Citation

Sun, X., Chuang, S.-H., Li, J., & McKenzie, F. (2009). Automatic diagnosis for prostate cancer using run-length matrix method. In N. Karssemeijer & M. L. Giger (Eds.), *Medical Imaging 2009: Computer-Aided Diagnosis, Proceedings of SPIE Vol. 7260 (72603H)*. SPIE of Bellingham, WA. <https://doi.org/10.1117/12.811414>

This Conference Paper is brought to you for free and open access by the Electrical & Computer Engineering at ODU Digital Commons. It has been accepted for inclusion in Electrical & Computer Engineering Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Authors

Xiaoyan Sun, Shao-Hui Chuang, Jiang Li, Frederic McKenzie, Nico Karssemeijer (Ed.), and Maryellen L. Giger (Ed.)

Automatic Diagnosis for Prostate Cancer using Run-length Matrix Method

Xiaoyan Sun^a, Shao-Hui Chuang^a, Jiang Li^{a*}, Frederic McKenzie^a

^aDept. of Electrical and Computer Engineering
Old Dominion University, Norfolk VA, USA, 23529-0246

ABSTRACT

Prostate cancer is the most common type of cancer and the second leading cause of cancer death among men in US¹. Quantitative assessment of prostate histology provides potential automatic classification of prostate lesions and prediction of response to therapy. Traditionally, prostate cancer diagnosis is made by the analysis of prostate-specific antigen (PSA) levels and histopathological images of biopsy samples under microscopes. In this application, we utilize a texture analysis method based on the run-length matrix for identifying tissue abnormalities in prostate histology. A tissue sample was collected from a radical prostatectomy, H&E fixed, and assessed by a pathologist as normal tissue or prostatic carcinoma (PCa). The sample was then subsequently digitized at 50X magnification. We divided the digitized image into sub-regions of 20 X 20 pixels and classified each sub-region as normal or PCa by a texture analysis method. In the texture analysis, we computed texture features for each of the sub-regions based on the Gray-level Run-length Matrix (GL-RLM). Those features include LGRE, HGRE and RPC from the run-length matrix, mean and standard deviation of the pixel intensity. We utilized a feature selection algorithm to select a set of effective features and used a multi-layer perceptron (MLP) classifier to distinguish normal from PCa. In total, the whole histological image was divided into 42 PCa and 6280 normal regions. Three-fold cross validation results show that the proposed method achieves an average classification accuracy of 89.5% with a sensitivity and specificity of 90.48% and 89.49%, respectively.

Keywords: Prostate cancer, Histopathology, Texture analysis, CAD development

1. INTRODUCTION

Prostate cancer is the most common type of cancer and the second leading cause of cancer death among men in US¹. Out of every three men who are diagnosed with cancer each year, one is diagnosed with prostate cancer¹. It is known that prostate cancer can be cured by a number of therapies if it is detected early while still localized in the prostate. Prostate-specific antigen (PSA) in serum is currently the most widely used approach for early detection of PCa. A high PSA concentration implies a possible high risk of prostate cancer. Usually, pathologists can also check the histopathological images of biopsy samples from patients under microscopes, and make diagnoses based on their own experience. With the advancements of machine learning technology, automated cancer diagnosis becomes a possibility. In this paper, we present an automated prostate cancer diagnosis method using a Gray Level Run-Length Matrix (GL-RLM) based texture analysis technique.

2. BACKGROUND

The Gray Level Run Length Matrix (GL-RLM) method was proposed by Galloway in 1975² and by Chu et al. in 1990³. GL-RLM extracts higher order statistical texture features and provides a quantitative description of the texture features in a given image. By definition, consecutive pixels with the same gray level, in a given direction, constitute a gray level run and the number of pixels in the run is called run length. For a given image, we can compute its GL-RLM along a given

*JLi@odu.edu; phone 1 757-683 6748; fax 1 757 683-3220; odu.edu

direction. Each element $p(i, j)$ in the GL-RLM represents the total number of runs with pixel gray level of i and run length of j in the whole images. For an image with n_r pixels, the size of the run-length matrix will be M by N , where M and N are the maximum gray level and the possible maximum run length in the corresponding image, respectively.

As indicated by its definition, relatively long gray-level runs would occur more often in a coarse texture while primarily short runs would describe fine textures. From the run-length matrix, various texture features can be derived. The following are eleven descriptors typically extracted from the run-length matrices⁴: short run emphasis (SRE), long run emphasis (LRE), high gray-level run emphasis (HGRE), low gray-level run emphasis (LGRE), pair-wise combinations of the length and gray level emphasis (SRLGE, SRHGE, LRLGE, LRHGE), run-length non-uniformity (RLNU), grey-level non-uniformity (GLNU), and run percentage (RPC). Formulas for calculating these feature descriptors are listed in the appendix.

For a given image, GL-RLMs with any given direction can be computed. Typical orientations used in run length method are four principle directions: 0° , 45° , 90° , and 135° .⁵ Once the run-length matrices are calculated along each direction, those feature descriptors can be used either with respect to each direction or a combination of all directions to get a global view of texture information.⁶

3. MATERIALS

3.1 Tissue image

A tissue sample was collected from a radical prostatectomy, H&E fixed, and assessed by a pathologist as normal tissue or prostatic carcinoma (PCa). Nine sub-images of the histopathological prostate tissue sample were taken by Hirox HI-SCOPE KH-3000 at $50\times$ magnification. All sub-images were RGB color images and saved as TIFF format. A whole image was manually reconstructed from these 9 sub-images by using Adobe Photoshop CS3. The reconstructed image has a resolution of 3416×4213 pixels (Fig.2(a)).

3.2 Cancerous region diagnosis

The regions containing cancerous cells were histopathologically identified. In addition, MALDI Mass Spectrometric data of the same tissue area as the histopathological image was analyzed by BioMap, which provides powerful processing and displaying abilities for MSI data⁷, to get another diagnostic result. These two results were combined to get the most probable region where cancer occurred.

4. METHODS

4.1 System overview

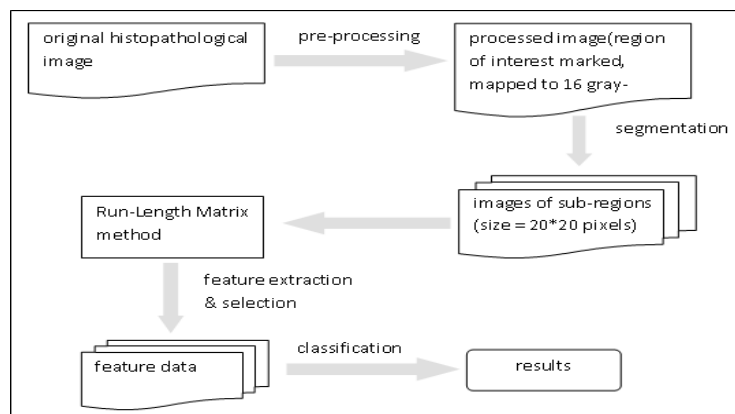


Fig. 1. Overview of the procedure for automatic prostate cancer diagnosis.

Figure 1 shows the system diagram where each step will be described in this section.

4.2 Pre-processing

4.2.1 Region of interest segmentation

In the histopathological image, many pixels act as background that has no cell type information. To accelerate processing speed, the region of interest (tissue regions) was extracted from its background. Because the background in the image is all dark, sub-images whose pixel values were lower than a preset threshold can be identified as background. Otherwise, pixels were marked as within the region of interest for later processing.

4.2.2 Gray-level mapping

Using too many gray-level intensities may cause large amounts of small run-length values in the run-length matrix. Therefore, it is necessary to group the gray levels into smaller bins. For 8-bit gray level images, research indicates that usually 16 gray levels are sufficient for texture discrimination or segmentation⁸. Here, we use a linear mapping to quantize the gray levels into values of 0~15.

4.2.3 Sub-regions separation

Feature extraction for automated cancer diagnosis can be done at the cellular-level or the tissue-level.⁹ The latter does not necessitate cell segmentation therefore simplifies the pre-processing. In this paper, we extract texture features at the tissue-level. By dividing the whole image into several sub-images, features are extracted from the small tissue area within each sub-image. Each sub-images is then classified as PCa or normal tissue based on the extracted features. In order to correctly classify each tissue area, the size of the sub-images should be chosen carefully. Too large areas may contain both cancerous and normal cells while too small areas contain insufficient information. By observing the object image, we choose the size of sub-images to be 20*20 pixels.

4.3 Feature extraction

After gray level mapping, four run-length matrices according to the four principle directions are calculated for each sub-image. Because we only want to get a global view of the texture information in the image, a new matrix is generated by summing the four matrices together,

$$P_{com} = P_0 + P_{45} + P_{90} + P_{135}.$$

where P_0 , P_{45} , P_{90} and P_{135} is the run-length matrix along directions of 0° , 45° , 90° , and 135° , respectively. Then, for each sub-region image, the 11 texture features are extracted from matrix P_{com} according to the formulas listed in appendix. Along with two intensity-based features (mean and standard deviation), 13 features in total are extracted for each sub-region. The work in preprocessing and feature extraction parts was done by using Matlab 7.0 (MathWork. Inc.)

4.4 Feature selection

Irrelevant features may cause problems like increased computation, curse of dimensionality and convergence difficulties.¹⁰ Feature selection, which is used to generate a compact subset of features leading to an accurate model based on the limited amount of data, is necessary before the classification stage to avoid these problems. A piecewise linear network method was used here to perform feature selection.¹⁰ The algorithm performs the following steps:

- 1) Generate an appropriate piecewise linear network (PLN) model for the given data set using a cross-validation (CV) method;
- 2) Apply the orthonormal least squares (OLS) procedure to the PLN model;
- 3) Select useful feature subsets through a floating search algorithm.

The feature selection algorithm described above is used in our experiment to obtain useful feature subsets. Among the total 13 features, the following 5 features are selected for the classification: Low Gray-Level Run Emphasis(LGRE), High Gray-Level Run Emphasis(HGRE), Run Percentage(RPC), mean and standard deviation.

4.5 Classification

4.5.1 Classification method

A multi-layer perceptron (MLP), one of the commonly used neural network classifiers, is used for classification. MLP is a feedforward artificial neural network model, which is designed through iterative regression to minimize the mean

squared error (MSE) between the target class and a network mapping function¹¹ The three layer MLP (input-hidden-output) is one of the universal approximators. In our experiment, we chose sigmoid and linear function as activation functions in hidden layer and output layer, respectively.

An output Weights Optimization – Backpropagation (OWO-BP) algorithm with Output Reset (OR) was used to solve the MLP training. In the OWO-BP algorithm, linear equations are solved to find output weights and backpropagation is used to find hidden weights.¹² Each instance of residual error contains at least two types of bias: an additive constant inherent to each output vector and a component due to difference between desired output and a correctly classified individual output value. The OR method we applied here adjusted the desired output to find the values of these two types of biases which minimize the training error.^{11,13}

Program for OWO-BP based MLP classifier is written in C.

4.5.2 Cross validation

One of the evaluation methods for classification systems is the *k*-fold cross validation. This approach randomly partitions the data set into *k* groups. Then, *k*-1 groups are combined as training data and the remaining group is used as testing data. This process is repeated *k* times such that each group is used for testing once. Finally, the *k*-group of testing results are used to estimate the error rate.

In our experiment, the number of cancerous data is especially small compared with the number of normal data(42 vs 6280). To handle the unbalanced data set, we use a modified *k*-fold cross-validation method to evaluate the system. For the training data groups in the cross-validation, we only keep part of the patterns from the normal class to insure that the same number of patterns are included into training process from the two classes (normal and cancerous).

From the total of 42 cancerous patterns and 6280 normal patterns, 3 groups of training data sets and 3 groups of testing data sets are generated correspondingly. Each training data has 28 patterns from cancerous class and 28 patterns from normal class, while each testing data has 14 and 2093 patterns from cancerous and normal class, respectively.

4.6 System evaluation

Accuracy is calculated for the classification results, which is defined as:

$$Accuracy = 1 - \frac{n_{mis-classification}}{n_{total\ patterns}}$$

In addition, the two most widely used statistics for describing diagnostic tests, sensitivity and specificity, are used to assess the success of the diagnostic results.¹⁴

$$Sensitivity = \frac{n_{true\ positive}}{n_{true\ positive} + n_{false\ negative}}, \quad Specificity = \frac{n_{true\ negative}}{n_{true\ negative} + n_{false\ positive}}$$

where definitions of the four parameters above are listed below:

	Samples with disease (Cancerous)	Samples without disease (Normal)
Test is positive	$n_{true\ positive}$	$n_{false\ positive}$
Test is negative	$n_{false\ negative}$	$n_{true\ negative}$

Therefore, sensitivity and specificity give the probability of getting accurate diagnoses results among samples with disease and without disease, respectively.

5. RESULTS.

5.1 Feature statistics

For each sub-region, we selected five texture and intensity features as inputs to the training system. Table 1 gives statistic analyses of the five features for normal and cancerous regions, respectively. One may observe that the values of HGRE, RPC, and STD have obvious differences between the two groups.

Table 1. statistical result for normal and cancerous regions (mean±standard deviation)

	Number	LGRE	HGRE	RPC	Mean	STD
Normal	6280	0.01 ± 0.00	108.24 ± 30.52	1.29 ± 0.31	9.40 ± 1.59	0.87 ± 0.63
Cancerous	42	0.01 ± 0.01	85.39 ± 20.54	2.22 ± 0.41	8.01 ± 1.34	1.63 ± 0.63

5.2 Diagnosis results

In the evaluation stage, we tested 3 groups of data by using the modified 3-fold cross-validation described in the system evaluation section. Each training data set included 28 data for cancerous regions and 28 data for normal regions, while each testing data set is composed with 14 and 2093 data from cancerous and normal regions, respectively. Table 2 shows the diagnosis results for these 3 groups. The average accuracy is 89.5%. Sensitivity and specificity are 90.48% and 89.49%, respectively.

Table 2. Diagnosis results

	n _{TP}	n _{FN}	n _{TN}	n _{FP}	Accuracy	Sensitivity	Specificity
Group 1	12	2	1910	183	91.22%	85.71%	91.26%
Group 2	12	2	1947	146	92.98%	85.71%	93.02%
Group 3	14	0	1762	331	84.29%	100.00%	84.19%
Overall	38	4	5619	660	89.50%	90.48%	89.49%

Figure 2 shows the comparison of the original histopathological image and its corresponding diagnosis result provided by our system, where the gray area was diagnosed as normal and black area was diagnosed as cancerous. In figure 2(c), the result produced by BioMap software is also given, where pixels with red color (or light gray in grayscale) indicates high possibility of cancer. Note that not all the red areas are cancer. Compare these three pictures, one can intuitively conclude that our diagnostic result is acceptable.

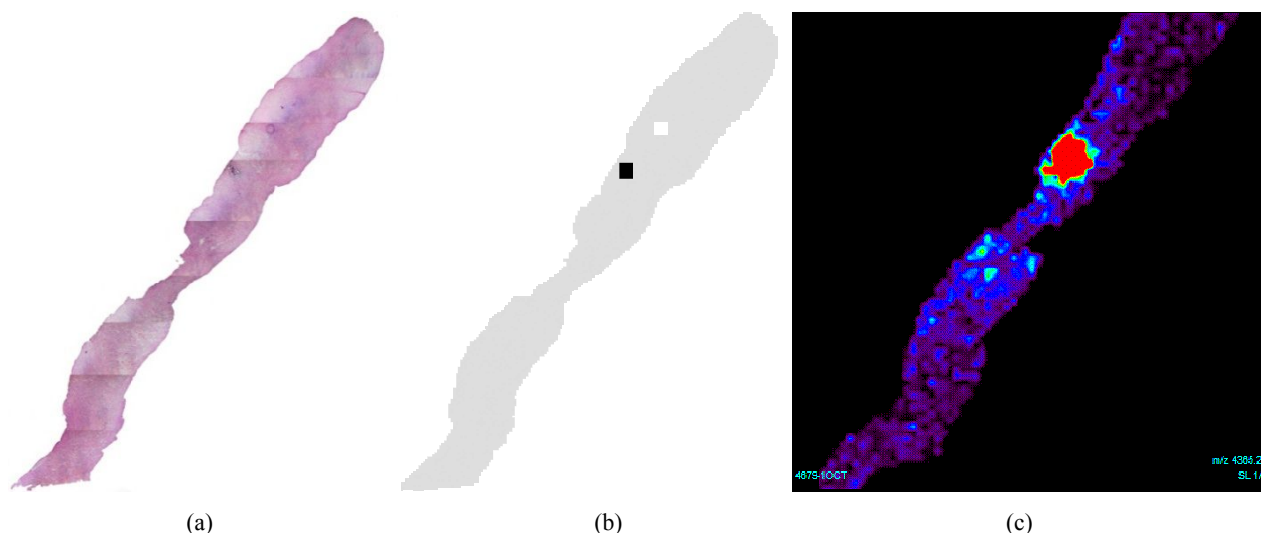


Fig. 2. Texture analysis result compared with the result produced by the BioMap software: (a) original histopathological image; (b) result produced by our automatic diagnosis system (black indicates cancerous tissue area); (c) result produced by BioMap software (red color or light gray in grayscale indicates high probability of cancer)

5.3 Discussion for the wrong diagnosis

Figure 3 is an illustration for the regions that were incorrectly diagnosed by this system. In this figure, gray points show the normal regions that were decided as cancerous, which means the 660 false-positive sub-regions; the four white points indicate the false-negative situations. Note that the 4 false-negative sub-regions are on the edge of the big cancerous area, where diagnosis mistakes are prone to happen.



Fig. 3. Distribution of diagnostic error occurrence (gray points present data of false positive and write points present data of true negative)

From table 3, it is observed that, compared with actual cancerous(true-positive) regions, the false-negative regions have much higher HGRE values and lower RPC and STD values. Also, RPC values are higher in false-positive regions than the average values of actual normal(true-negative) regions. These differences may cause the incorrect diagnosis results in these regions.

Table 3. feature data of incorrect diagnosis regions

Category	LGRE	HGRE	RPC	Mean	STD
True-positives	0.01 ± 0.01	85.39 ± 20.54	2.22 ± 0.41	8.01 ± 1.34	1.63 ± 0.63
No.1 false-negative	0.008	122.260	1.225	10.045	0.577
No.2 false-negative	0.009	110.160	1.210	9.577	0.565
No.3 false-negative	0.010	106.490	1.585	9.395	1.141
No.4 false-negative	0.009	106.860	1.337	9.370	0.741
True-negatives	0.01 ± 0.00	108.24 ± 30.52	1.29 ± 0.31	9.40 ± 1.59	0.87 ± 0.63
False-positives	0.01 ± 0.002	103.97 ± 21.28	1.541 ± 0.29	9.21 ± 1.21	0.99 ± 0.56

6. CONCLUSION

In this paper, a method for automatic prostate cancer diagnoses using machine learning was presented. Texture and intensity features were extracted from the histopathological image of prostate tissues which contain a cancerous area. Texture features of the tissue image were calculated by using the gray-level run-length matrix(GL-RLM) method. Among the popular eleven texture features, along with mean and standard deviation of the intensity information, five of them were selected for classification, including LGRE, HGRE, RPC, mean and standard deviation. These five features were inputted into a multi-layer perceptron(MLP) classifier to determine if a sub-region is normal or cancerous.

The average accuracy of the classification result was 89.5%. Overall sensitivity and specificity of this diagnosis system were 90.48% and 89.49%, respectively.

The proposed system provided a feasible method for automatic diagnosis of prostate cancer. Tissue texture features were proved to give good results for the classification from the histopathological image.

7. ACKNOWLEDGEMENT

We thank Dr. John Semmes at the Eastern Virginia Medical School (EVMS) for providing the prostate tissue sample.

APPENDIX

Feature	Formula
Short Run Emphasis	$SRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j)}{j^2}$
Long Run Emphasis	$LRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N p(i, j) * j^2$
Low Gray-Level Run Emphasis	$LGRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j)}{i^2}$
High Gray-Level Run Emphasis	$HGRE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N p(i, j) * i^2$
Short Run Low Gray-Level Emphasis	$SRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j)}{i^2 * j^2}$
Short Run High Gray-Level Emphasis	$SRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j) * i^2}{j^2}$
Long Run Low Gray-Level Emphasis	$LRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j) * j^2}{i^2}$
Long Run High Gray-Level Emphasis	$LRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N p(i, j) * i^2 * j^2$
Gray-Level Non-uniformity	$GLNU = \frac{1}{n_r} \sum_{i=1}^M \left(\sum_{j=1}^N p(i, j) \right)^2$
Run Length Non-uniformity	$RLNU = \frac{1}{n_r} \sum_{j=1}^N \left(\sum_{i=1}^M p(i, j) \right)^2$
Run Percentage	$RPC = \frac{n_r}{M * N}$

REFERENCES

- [1] <http://www.cancer.gov/>
- [2] Galloway M. M., "Texture Analysis Using Gray Level Run Lengths", Computer Graphics and Image Processing, Papers 4, 172-179 (1975).
- [3] Chu A., Sehgal C. M. and Greenleaf J. F., "Use of Gray Value Distribution of Run Lengths for Texture Analysis", Pattern Recognition Letters, Papers 11(6), 415-420 (1990).
- [4] Tang X., "Texture Information in Run-Length Matrices", IEEE Transactions on image processing, Papers 7(11) 1602-1609 (1998).
- [5] Loh Horng-Hai, Leu Jia-Guu, "The analysis of natural textures using run length features", IEEE Transactions on industrial electronics, Papers 35(2) 323-328 (1998).
- [6] Harlick R. M., Shanmugam K., "Textural features for image classification", IEEE Transactions on systems, man, and cybernetics, Papers 3(6) 610-621 (1973).
- [7] http://www.maldi-msi.org/index.php?Itemid=32&id=14&option=com_content&task=view
- [8] Xu Dong-Hui, Kurani A. S., "Run-length encoding for volumetric texture", Proc. (452) Visualization, Imaging, and Image Processing, 6-8 (2004).
- [9] Demir C., Yener B., "Automated cancer diagnosis based on histopathological images - a systematic survey", Technical report, Rensselaer Polytechnic Institute, Department of Computer Science, TR-05-09 (2005).
- [10] Li J., Manry M. T., "Feature selection using a piecewise linear network", IEEE Transactions on networks, Papers 17(5) 1101-1115 (2006).
- [11] Gore R. G., J. Li, "Iterative design of neural network classifiers through regression", International Journal on Artificial Intelligence Tools, Papers 14(1&2), 281-301 (2005).
- [12] Manry M.T., et al, "Fast Training of Neural Networks for Remote Sensing", Remote Sensing Reviews, Papers 9, 77-96 (1994).
- [13] Li J., Manry M. T., "Iterative improvement of neural classifiers", FLAIRS 700-705 (2004).
- [14] Altman DG, Bland JM, "Diagnostic tests. 1: Sensitivity and specificity". BMJ 308 (6943): 1552 (1994).