

2010

Adjacent Slice Prostate Cancer Prediction to Inform MALDI Imaging Biomarker Analysis

Shao-Hui Chuang
Old Dominion University

Xiaoyan Sun
Old Dominion University

Lisa Cazares
Eastern Virginia Medical School

Julius Nyalwidhe
Eastern Virginia Medical School

Dean Troyer
Eastern Virginia Medical School

See next page for additional authors

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



Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Cancer Biology Commons](#), [Diagnosis Commons](#), and the [Diseases Commons](#)

Original Publication Citation

Chuang, S.-H., Sun, X., Cazares, L., Nyalwidhe, J., Troyer, D., Semmes, O. J., Li, J., & McKenzie, F. (2010) Adjacent slice prostate cancer prediction to inform MALDI imaging biomarker analysis. In Nico Karssemeijer & Ronald M. Summers (Eds.), *Medical Imaging 2010: Computer-Aided Diagnosis, Proceedings of SPIE Vol. 7624 (762433)*. SPIE. <https://doi.org/10.1117/12.844403>

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

Shao-Hui Chuang, Xiaoyan Sun, Lisa Cazares, Julius Nyalwidhe, Dean Troyer, O. John Semmes, Jiang Li, Frederic D. McKenzie, Nico Karssemeijer (Ed.), and Ronald M. Summers (Ed.)

Adjacent Slice Prostate Cancer Prediction to Inform MALDI Imaging Biomarker Analyses

¹Shao-Hui Chuang, ¹Xiaoyan Sun, ²Lisa Cazares, ²Julius Nyalwidhe,
²Dean Troyer, ²O. John Semmes, ¹Jiang Li, ¹Frederic D. McKenzie

¹Dept. Of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, USA,
23509-0246

²Dept. Of Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, Norfolk,
VA, USA, 23507

Abstract

Prostate cancer is the second most common type of cancer among men in US [1]. Traditionally, prostate cancer diagnosis is made by the analysis of prostate-specific antigen (PSA) levels and histopathological images of biopsy samples under microscopes. Proteomic biomarkers can improve upon these methods. MALDI molecular spectra imaging is used to visualize protein/peptide concentrations across biopsy samples to search for biomarker candidates. Unfortunately, traditional processing methods require histopathological examination on one slice of a biopsy sample while the adjacent slice is subjected to the tissue destroying desorption and ionization processes of MALDI. The highest confidence tumor regions gained from the histopathological analysis are then mapped to the MALDI spectra data to estimate the regions for biomarker identification from the MALDI imaging. This paper describes a process to provide a significantly better estimate of the cancer tumor to be mapped onto the MALDI imaging spectra coordinates using the high confidence region to predict the true area of the tumor on the adjacent MALDI imaged slice.

Keywords: Prostate Cancer, Texture Analysis, MLP Classification, Histopathology

1. Introduction

Prostate cancer is the second most common type of cancer among men in US [1]. Each year, more than 186,000 American men learn they have this disease [1]. Therefore, significant research continues to target this area. Prostate cancer diagnosis is made by the analysis of prostate-specific antigen (PSA) levels and histopathological images of biopsy samples under a microscope. A pathologist makes a classification based on her own experience, which may vary with training, experience and available time i.e. if cancer has already been found in one area of a biopsy slice there may be little motivation to examine other parts of the slice. Matrix-assisted Laser Desorption/Ionization (MALDI) molecular spectra imaging is used to visualize protein/peptide concentrations across biopsy samples to search for biomarker candidates. The highest confidence tumor regions gained from the histopathological analysis are then mapped to the MALDI spectra data to estimate the regions for biomarker identification from the MALDI imaging.

MALDI mass spectrometry has become a useful tool for identifying protein/peptide concentrations in biopsy samples since the mid 1980s [2-4]. The biopsy sample preparation for MALDI required use one fresh frozen tissue, which by itself is not conducive to histopathological review. Unfortunately, traditional processing methods require histopathological examination on one slice of a biopsy sample while the adjacent slice is subjected to the tissue destroying desorption and ionization processes of MALDI. Therefore, the adjacent biopsy sample was required for image analysis using a dual staining procedure using haematoxylin and eosin (H&E) preferred after formalin fixation. This staining procedure is used to increase the visual contrasts among different cytological components [5]. Thus, after this procedure, the sample tissue is no longer a fresh frozen tissue. Formalin breaks down protein structures in tissue; therefore, MALDI mass spectrometry process and image analysis are not able to work on the same biopsy sample tissue. This is the reason for using an adjacent slice for the MALDI mass spectrometry process.

This research describes a process to provide a significantly better estimate of the cancer tumor to be mapped onto the MALDI imaging spectra coordinates using the high confidence region to predict the true area of the tumor on the adjacent MALDI imaged slice.

2. Background

2.1 Texture Analysis

In image processing, texture research has been a popular and well discussed topic. Texture analysis methods have important roles in medical, industrial and remote sensing areas. We base our texture-recognition on a high-optical magnification (420x) image analysis of traditionally formalin fixed tissue stained with hematoxylin and eosin (H&E) on microscope slides.

The advantage of using high resolution images is that we can increase the accuracy of results. The disadvantage is that it takes significant preparation time including digitizing the tissue samples with a high optical magnification lens and combining sub-images into a larger image. However, our recognition algorithms based on gray-scale intensities can appropriately analyze the histological features.

2.2 MALDI

MALDI (Matrix-Assisted Laser Desorption/Ionization) is a soft-ionization technique used in mass spectrometry for desorption/ionization of large molecules such as peptides and proteins. Particular protein structures are related to diseased tumor tissue. The MALDI laser treats the target tissue at single spots of about 10 to 50 micrometers at contiguous coordinates across the whole tissue yielding spectra for every spot. The spots are identified and mapped to the matrix which in effect produces a visualization of the tissue itself but one created based on concentrations of a particular targeted protein found in each spot spectra where they are illustrated by intensity and color (heat map). This essentially provides a contour map of protein expressions over the landscape of the piece of tissue being analyzed. Now, the problem is to find one or a combination of molecular biomarkers that adequately represents the normal and diseased tumor tissue within the biopsy section. Based on MALDI imaging, researchers can compare the tumor area with a tumor-free area on the same section for the different proteins/peptides selected.

2.3 Gray Level Run Length Matrix (GL-RLM)

The Gray Level Run Length Matrix (GL-RLM) method was proposed by Galloway in 1975 [6] and by Chu et al. in 1990 [7]. 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 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 image. For an image with n_i 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. 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 degrees [8]. 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 [9]. The features for GL-RLM are Short Run Emphasis (SRE), Long Run Emphasis (LRE), Low Gray-Level Run Emphasis (LGRE), High Gray-Level Run Emphasis (HGRE), Short Run Low Gray-Level Emphasis (SRLGE), Short Run High Gray-Level Emphasis (SRHGE), Long Run Low Gray-Level Emphasis (LRLGE), Long Run High Gray-Level Emphasis (LRHGE), Gray-Level Non-uniformity (GLNU), Run Length Non-uniformity (RLNU) and Run Percentage (RPC)[6].

2.4 Gray Level Co-occurrence Matrix (GLCM)

The gray level co-occurrence matrix (GLCM) proposed by Haralick et al. in 1973 [9] corresponds to the second order statistics. When an image has a gray scale level in the range of $[0, N_g]$, a matrix can be generated by the size of $(N_g+1) * (N_g+1)$ in which rows and columns represent the possible image value. N_g is the number of distinct gray levels in the quantized image. There are two important parameters used in the GLCM method: distance (d) and angle (a). The

distance (d) is the distance, and the angle (a) is the direction between the pixels where we would like to compute the comparison of gray-tone value. When the distance is smaller, the integration of the texture pattern is higher. Therefore, the distance d used in this research is set as 1.

Accordingly, 14 features are extracted from GLCM, which include f1: Angular Second Moment, f2: Contrast, f3: Correlation, f4: Sum of Squares: Variance, f5: Inverse Difference Moment, f6: Sum Average, f7: Sum Variance, f8: Sum Entropy, f9: Entropy, f10: Difference Variance, f11: Difference Entropy, f12: Information Measures of Correlation 1, f13: Information Measures of Correlation 2 and f14: Maximal Correlation Coefficient. The Angular Second Moment (f1) is also called as Energy, Uniformity and Uniformity of Energy. The Contrast (f2) is also called as Inertia. The Inverse Difference Moment (f5) is also called Homogeneity. The maximal correlation coefficient (f14) was not calculated due to computational instability; therefore, there were 13 texture features from GLCM for each sub image.

3. Methods

3.1 Data collection and Image Pre-processing

The prostate tissue sample was obtained from a prostate biopsy, formalin fixed and paraffin embedded, then two adjacent slices approximately six microns thick were mounted on microscope slides. One slice was prepared for MALDI processing and the other H&E stained was assessed by a pathologist as normal tissue or prostatic carcinoma (PCa). The pathologist mapped a high confidence PCa region to the MALDI processed tissue and the associated coordinates were identified.

The H&E sample was captured with 184 color images by Hirox HI-SCOPE KH-1300 at the magnification of 420x. Each image was overlapped by at least 20% between images to facilitate reconstruction of the whole histopathology image. The sample had been rotated and images had been captured as vertical strip in order to minimize the image size. The image was saved as TIFF format so no information would be lost by file compression. The histopathology tissue image was reconstructed by using E-tiling and Adobe Photoshop CS3. The resolution of the histopathology tissue (Fig. 1a) image is 7428*43333*3 pixels. Due to very high resolution of the biopsy histopathology tissue image, a section of the biopsy histopathology tissue image that contains all PCa regions and most of normal regions had been cropped for subsequent processing in this paper. The cropped section (Fig. 1b) has 7150*13650 color pixels.

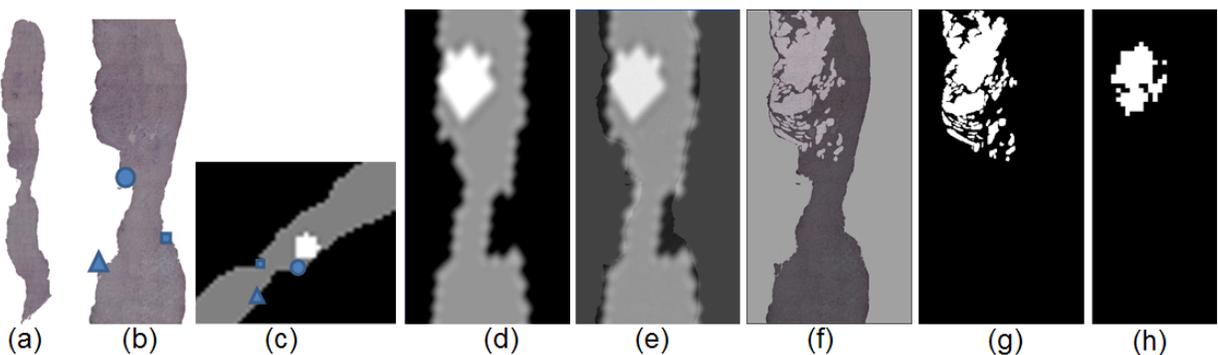


Figure 1. (a) Histopathology tissue image (b) Histopathology tissue section image with characteristic landmarks (c) MCV with characteristic landmarks (d) Low magnification ground truth (e) Overlaid LMGT image after registration (f) Overlay high magnification images (g) High magnification ground truth (h) Combined ground truth intersection

MALDI Coordinates Visualization (MCV) (Fig. 1c) with identified high confidence PCa coordinates (in white) was used to generate what we will call the low magnification ground truth (LMGT) (Fig. 1d) by registering the MCV to the histopathology tissue section image. A landmark-based registration method is used by finding three characteristic locations that match in both images. Since the histopathology tissue section image and MCV are adjacent tissue slices, the overlay image after registration (fig. 1e) shows some offsets in the shape. The LMGT white part of the MCV is mapped and now becomes the LMGT of the histopathology image which is the highest probability PCa area determined under low magnification.

Next, we used the high magnification tissue images that would not normally be viewed at such a resolution and submitted them for histopathological re-review and manual classification. Our pathologist-identified cancer areas are shown (Fig 1f) under 420x magnification which generated the high magnification ground truth (HMGT) (Fig. 1g). A combined ground truth (Fig. 1h) was generated by intersection of the LMGT and the HMGT. Thus, the white area of the combined ground truth indicated a more reasonable high confidence PCa area of the LMGT.

Note that identifying the high magnification ground truth is very time consuming for pathologists to identify. Therefore, predicting this area and mapping back to the MCV will not only give biomarker searching better ground truth information but will also save time and money. In general, HMGT does not exist for the tissue samples. The main purpose of this paper is to describe a step towards an automated process to classify PCa in the high magnification tissue and to map this information back to the adjacent slice coordinates for a more accurate prediction of HMGT PCa in the MALDI imagery.

3.2 System Overview

The system can be separated by some main processes, including image preprocessing, feature extraction and diagnosis. In the image pre-processing step, it contains data collection, image reconstruction and image registration. In the feature extraction step, it contains feature selection. In the diagnosis step, it contains k-fold cross validation, multi-layer perceptron classification and system evaluation. Detailed descriptions of these processes are in the sections below. Figure 2 is the overview of system procedures for first experiment.

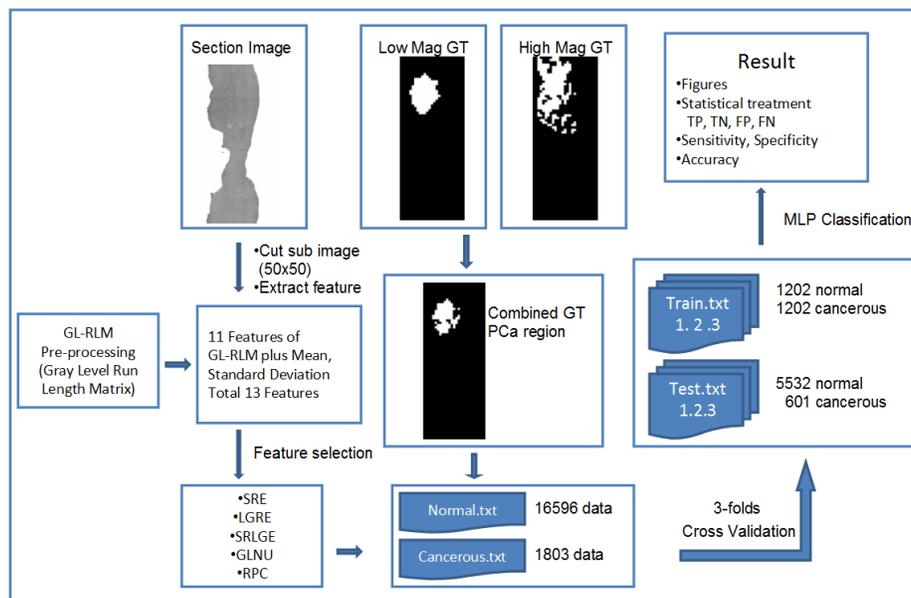


Figure 2. Overview of the system procedures

3.2.1 Feature Extraction

In this paper, a RLM feature matrix is generated by summing the four angles of the directional RLM feature matrices together for a global view. In order to correctly classify each tissue area, the size of the sub-images should be chosen carefully. Too large area may contain both cancerous and normal cells while too small areas may contain insufficient information. The size of sub-images used in this paper is 200*200 pixels. Based on each sub-images, 13 features are extracted including 11 features of the GL-RLM, mean and standard deviation. Each sub-image is eventually classified as either normal or PCa. The code of this feature extraction section was programmed in Matlab 8.0 (MathWork. Inc.)

3.2.2 Feature Selection

Irrelevant features may cause problems like increased computation, curse of dimensionality and convergence difficulties [11]. The idea of feature selection is to choose a subset of input variables by eliminating features with little or no predictive information [12] or features that are highly correlated with an already chosen more predictive feature. A piecewise linear network method was used to perform feature selection [11]. The algorithm performs by generating an appropriate piecewise linear network (PLN) model for the given data set using cross-validation method, applying the orthonormal least squares (OLS) procedure to the PLN model, then selecting useful features subsets through a floating search algorithm[13]. The program code for this procedure is written in C language.

3.2.3 k-fold Cross Validation

K-fold cross validation is one of the useful evaluation methods for classification. The data set is divided into k subsets, and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other $k-1$ subsets are put together to form a training set. Performance is then computed based on the testing results. A 3-fold cross validation was used in first experiment of this paper. The program code for this procedure is written in C language.

3.2.4 Multi-layer Perceptron (MLP) Classification

A MLP, one of the commonly used neural network classifiers, is used for classification. The MLP is an artificial neural network that learns nonlinear function mappings. The MLP is capable of learning a rich variety of nonlinear decision surfaces [14]. The three-layer MLP (input-hidden-output) is the most frequently used structure. In this paper, sigmoid and linear function was used as activation functions in the hidden layer and the output layer, respectively. The program code for this procedure is written in C language.

3.2.5 Confusion Matrix Table

The two most widely used statistics for describing diagnostic tests, sensitivity and specificity, are used to assess the success of the diagnostic results. A confusion Matrix table (Table 1) can easily calculate and describe sensitivity, specificity and accuracy.

Table 1. Confusion Matrix Table.

		Actual Cancer Condition	
		TRUE (Cancer)	FALSE (Normal)
Predicted Result	Positive (Cancer)	True Positive	False Positive
	Negative (Normal)	False Negative	True Negative
		Sensitivity TP/(TP+FN)	Specificity TN/(TN+FP)
Accuracy (TP+TN)/(TP+TN+FP+FN)			

A true positive indicated cancerous area correctly is diagnosed as cancerous, a false positive indicated normal area is wrongly identified as cancerous, a false negative indicated cancerous area is wrongly identified as normal and a true negative indicated normal area is correctly identified as normal. Sensitivity is the ratio of the number of true positives over the total number of actual true conditions. In conditional probability notation, sensitivity is expressed as the probability of a predicted positive test given that the actual condition is cancerous. Specificity is the ratio of the number of true negatives over the total number of actual false conditions, and accuracy is the sum of true positives and true negatives over the sum of total patterns. In conditional probability notation, specificity is expressed as the probability of a predicted negative test given that normal is the actual condition.

4. Experiments and Results

4.1 Experiments

The purpose of experiment A is to validate the performance of our classification algorithm by training and testing on known PCa ground truth and known normal ground truth areas. In the first three experiments of this paper, five features are selected from 13 features for the classification including SRE, LGRE, SRLGE, GLNU and RPC. The PCa ground truth used in this experiment is the combined ground truth intersection (Fig 1h). This PCa region contains 102 points. Each point refers to a 200x200 pixel block, which is the size of the sub-image. The normal cells ground truth (102 points) is selected from the bottom half section of the histopathology tissue section image where it is known to contain a low probability of any cancerous regions. Figure 3 a & b clearly show the dividing line for the top and bottom half separation. The total normal region used in this experiment, which is located below the separation, contains 534 points. In each run of the 3-fold cross validation, the number of training data is 136 points that contains two sets of PCa cells (68 points) and two sets of a matching amount of normal cells (68 points). The number of testing data is 212 points that contains the remaining sets of PCa cells (34 points) and the full set of normal cells (178 points).

In experiment B, no cross validation process is involved. The training data contains PCa ground truth (102 points) that is the same as that in experiment A, and a matched number of negative points from the lower part of the tissue. After training, we test the trained model on the whole data set using the pathologist classification results as our ground truth (255 PCa and 944 normal points).

In experiment C, the training PCa region ground truth data is based on the low magnification ground truth (Fig. 1e). The training data in this experiment contains all PCa regions in the LMGT (138 points) and a matching amount of low probability PCa regions (138 points). The testing data is the same as in experiment B.

In experiment D, the feature selection algorithm was rerun because the majority of training data is different from that in previous experiments. The used features are LRE, HGRE, SRLGE, LRLGE, and mean. The training PCa region ground truth image is based on HMGT, high magnification ground truth (Fig 1g). The training data in this experiment contains all PCa regions (255 points) and a matching amount of low probability PCa regions (255 points). The testing data is the same as experiment B.

A new feature selection is also required for the last three experiments because they use the combination features of GLCM and GLRLM. In experiment E, the feature methods used in this part are GLRLM, GLCM, mean and standard deviation. Selected from these features, the feature subsets used in this part are SRE, SRLGE, LGRE, RPC, GLNU, LRLGE from GLRLM, Sum Entropy, Contrast, Angular Second Moment from GLCM, and mean.

In experiment F, the feature methods used in this part of the experiment are GLRLM, GLCM, mean and standard deviation. The useful feature subsets are the same as in part 2 of experiment E.

In experiment G, the feature methods used in this experiment are GLCM, GLRLM, mean and standard deviation. The useful subsets are HGRE, SRE, SRLGE, GLNU from GLRLM features, Inverse Difference Moment, Sum Variance, Contrast, Sum Entropy, Angular Second Moment, Difference Entropy, Information Measures of Correlation I and Correlation from GLCM, mean and standard deviation.

4.2 Results

The basic algorithm validation result image (Fig. 3a) shows the classification and error locations identified by the dark spots in the bottom half of the image. Using an assumption that isolated spots are highly unlikely to be PCa we obtain the final result image (Fig. 3b) by removing isolated error points in an 8-connected neighborhood. The two most widely used statistics for describing diagnostic tests, sensitivity and specificity, are used to assess the success of the diagnostic results [15]. The confusion matrix (Table 2) indicates the number of correct or erroneous classification points. From Table 1, in the original prediction, sensitivity and specificity are 99.02% and 91.39%, respectively, and the average accuracy is 92.61%. In the final prediction (after removing isolated points), sensitivity and specificity are 100% and 96.25%, respectively, and the average accuracy is 96.86%.

In the final prediction (after removing isolated points) of experiment B, sensitivity and specificity obtained 83.92% and 81.25%, respectively, and the accuracy was 81.82%. The final prediction of experiment C obtained 87.45% sensitivity, 75.0% specificity, and 77.65% accuracy. For experiment D, the final results show 95.29% sensitivity, 72.03% specificity, and 76.98% accuracy. The overall result figures (Fig. 3c-3e) and table (Table 3) are shown below. The final prediction of experiment E obtained 83.14% sensitivity, 82.52% specificity, and 86.65% accuracy. The final prediction of experiment F obtained 85.10% sensitivity, 74.15% specificity, and 76.48% accuracy. For experiment G, the final results show 94.12% sensitivity, 70.34% specificity, and 75.40% accuracy. Table 4 shows the result using both GLCM and GLRLM.

4.3 Results Discussion

The result of experiment A is as expected. The training data and testing data only contained the highest probability of PCa region and highest probability of normal cells region. Therefore, it provides the high sensitivity, specificity and accuracy.

In general, HMGT does not exist for all tissue samples. We therefore try to predict the HMGT area from the LMGT area as performed by experiment C. Experiment B used a refined LMGT area by obtaining its intersection with the HMGT thereby removing identified normal regions from the LMGT. Even so, experiment C outperforms experiment B in sensitivity. However, it over-predicts the cancer area. It is reasonable that the experiment D obtained a high sensitivity because the pathologist identified all cancer regions thereby reducing its uncertainties. Interestingly, experiment C outperforms experiment D in terms of specificity. One might take from this result that some of the normal regions' appearances are indeed quite close to those of cancer spots which may justify the lower specificity values and thus the larger predicted areas of PCa (Fig. 3d).

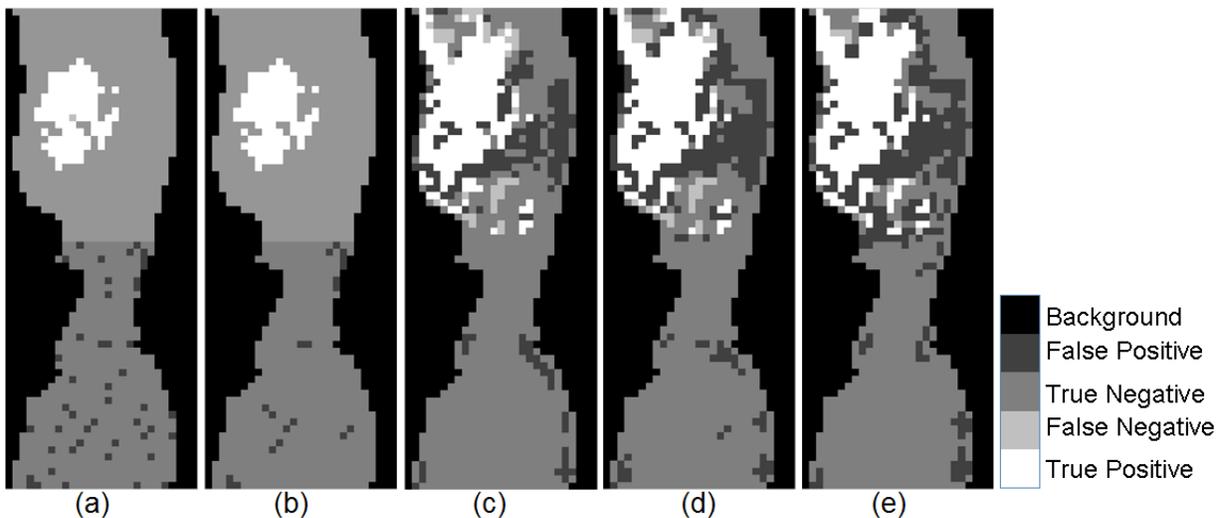


Figure. 3. (a) original algorithm result (b) final algorithm result (after removing isolated points) for experiment A (c) final prediction result for experiment B (d) final prediction result for experiment C (e) final prediction result for experiment D

Table 2. Confusion matrix for experiment A

Experiment A		Original Prediction		Final Prediction (After removing isolates points)	
		Actual Cancer Condition		Actual Cancer Condition	
Predicted Result	Positive (Cancer)	TRUE (Cancer) True Positive 101	FALSE (Normal) False Positive 46	TRUE (Cancer) True Positive 102	FALSE (Normal) False Positive 20
	Negative (Normal)	False Negative 1	True Negative 488	False Negative 0	True Negative 514
		Sensitivity 99.02%	Specificity 91.39%	Sensitivity 100%	Specificity 96.25%
		Accuracy 92.61%		Accuracy 96.86%	

Table 3. The overall result of Experiments B, C & D

	Overall result	Sensitivity	Specificity	Accuracy
Exp B	Train: LMGT & HMGT Test: HMGT	83.92%	81.25%	81.82%
Exp C	Train: LMGT Test: HMGT	87.45%	75.00%	77.65%
Exp D	Train: HMGT Test: HMGT	95.29%	72.03%	76.98%

Table 4. The overall result of Experiments E, F & G

	Overall result	Sensitivity	Specificity	Accuracy
Exp E	Train: LMGT & HMGT Test: HMGT	83.14%	82.52%	86.65%
Exp F	Train: LMGT Test: HMGT	85.10%	74.15%	76.48%
Exp G	Train: HMGT Test: HMGT	94.12%	70.34%	75.40%

5. Conclusion

In this paper, Texture features using the GL-RLM method were extracted from the histopathological image of a prostate biopsy tissue that contains normal cells and PCa cells. Among the 11 RLM texture features, SRE, LGRE, SRLGE, GLNU and RPC are the most important features to distinguish PCa cells and normal cells because they were selected in all of our experiments. After using the combination of GLRLM and GLCM features (experiments E, F and G), the result does not show significant difference.

In this research, the system generates a high accuracy result when testing on a high probability cancer region and the bottom half normal region (experiments A and B). When the size of the sub-image is 200x 200 pixels, the system provides a better result (experiment B). When the system is tested on the high probability cancer region and all normal regions, there are many misdiagnosed locations on the top half normal region (experiment C).

We successfully developed a system to predict the high magnification ground truth from LMGT (experiment C). Comparing experiments B, C and D, the training data in experiment C does not contain HMGT. However, it shows a higher sensitivity result than that in experiment B and a higher specificity and accuracy result than that in experiment D.

6. Future Works

In this research, only a section of the histological image was used. The system should be tested on different histological images to ensure the quality of the system. Also, all images are gray-scale. Currently, there is a trend to use color images, and some researchers have started investigating the color image [11, 16-17].

In the preprocessing step, filtering images may remove the unwanted parts of the image and present a better quality image before feature extraction. Some filtering methods such as wavelet features may be helpful. In the diagnosis step, it

may also be possible to classify the malignancy level of the tissue using the Gleason staging system, which is used to help evaluate the prognosis of men with prostate cancer.

Finally, the 3D topology including height difference between cancer and normal cells is an interesting topic for enhancing this work and none of the published papers mentioned this yet.

REFERENCES

- [1] National Cancer Institute, <http://www.cancer.org/>
- [2] F. Hillenkamp, M. Karas, R. C. Beavis, B. T. Chait, Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.*, Vol. 63, Issue 24, pp. 1193A-1203A. (1991).
- [3] M. Karas, F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.*, Vol. 60, Issue 20, pp. 2299-2301. (1988).
- [4] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yohida, Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry, *Rapid Communications in Mass Spectrometry*, Vol 2. Issue 8, pp. 151-153. (2005).
- [5] J. Kong, O. Sertel, H. Shimada, K. L. Boyer, J. H. Saltz, M. N. Gurcan, Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation, *Pattern Recognition*, Vol. 42, pp. 1080-1092. (2009).
- [6] Galloway M.: Texture Analysis Using Gray Level Run Length. In: *Computer Graphics and Image processing*, Paper 4, 172-179 (1975).
- [7] Chu A., Sehgal C., Greenleaf J.: Use of Gray Value Distribution of Run Lengths for Texture Analysis. In: *Pattern Recognition Letters*, Papers 11(6), 415-420 (1990)
- [8] Loh H., Leu J.: The analysis of natural textures using run length features. In: *IEEE Transactions on industrial electronics*, Papers 35(2) 323-328 (1998)
- [9] Haralick R., Shanmugam K.: Textural features for image classification. In: *IEEE Transactions on systems, man, and cybernetics*, Papers 3(6) 610-621 (1973)
- [10] Demir C., Yener B.: Automated cancer diagnosis based on histopathological images – a systematic survey. In: Technical report, Rensselaer Polytechnic Institute, Department of Computer Science, TR-05-09 (2005)
- [11] Li J., Manry M. T.: Feature Selection using a piecewise linear network. In: *IEEE Transaction on networks*, Papers 17(5) 1101-1115 (2006).
- [12] Kim Y., Street N., Menczer F.: Feature Selection in Data Mining. University of Iowa, USA. <http://dollar.biz.uiowa.edu/~street/research/dmcc.pdf>
- [13] Sun X., Chuang S., Li J., McKenzie F.: Automatic Diagnosis for Prostate Cancer using Run-length Matrix Method. In: *Proc. SPIE*, Vol. 7260, 72603H (2009).
- [14] CIS311: Neural Networks, <http://homepages.gold.ac.uk/nikolaev/311multi.htm>
- [15] Altman D., Bland J.: Diagnostic test 1: Sensitivity and Specificity. In: *BMJ* 308(6943):1552 (1994)
- [16] S. Doyle, A. Madabhushi, M. Feldman, J. Tomaszewski, A Boosting Cascade for Automated Detection of Prostate Cancer from Digitized Histology, *MICCA, LNCS* 4191, pp. 504-511. (2006).
- [17] A. Porebski, N. Vandenbroucke, L. Macaire, Haralick Feature Extraction from LBP Images for Color Texture Classification, *Image Processing Theory, Tool & Applications, IPTA*. (2008).