RESEARCH ARTICLE

USING MACHINE LEARNING AND TEXTURE ANALYSIS TO PREDICT SIGNIFICANT AND NONSIGNIFICANT PROSTATE LESIONS

Abdullah S. Mirza[1*], Yazeed Alsulaiman[2], Metab A. Alkubeyyer[3]

[1]Medical Imaging Department, King Saud Medical City, Saudi Arabia.
[2]Radiology and Nuclear Medicine Department, Security Forces Hospital, Saudi Arabia
[3]Diagnostic Radiology Department, King Saud University Medical City, Saudi Arabia.

*Corresponding Author: mirza.as@hotmail.com

ABSTRACT

Background: Prostate cancer is a well-recognized medical problem accounting for the most diagnosed type of cancer in men. The importance of early detection and its improved survival rate have motivated research on the best cancer detection method. Consequently, computer-aided diagnosis was introduced; however, more datasets are needed, and more testing and trials are required to reach a feasible and reliable diagnostic method. In this study, we use MRI T2 WI and ADC-map sequences to build a classifier to differentiate between clinically significant and insignificant prostate lesions.

Methods: Haralick's first and second order statistical features were extracted from pathologically proven prostate lesions found in The Cancer Imaging Archive open data source. We used the WEKA platform for data analysis, including 152 lesions divided into 70% training set and 30% testing set.

Results: The proposed classifier showed sensitivity, specificity, F-measure, and AUROC of 82.6%, 87%, 84.4%, and 92.6%, respectively.

Conclusions: The proposed classifier does not require a high-end computer, outperforms many previous classifiers, and has the potential to discriminate clinically significant from insignificant prostate lesions.

Keywords: Prostate, machine learning, AI, cancer, prediction.

INTRODUCTION

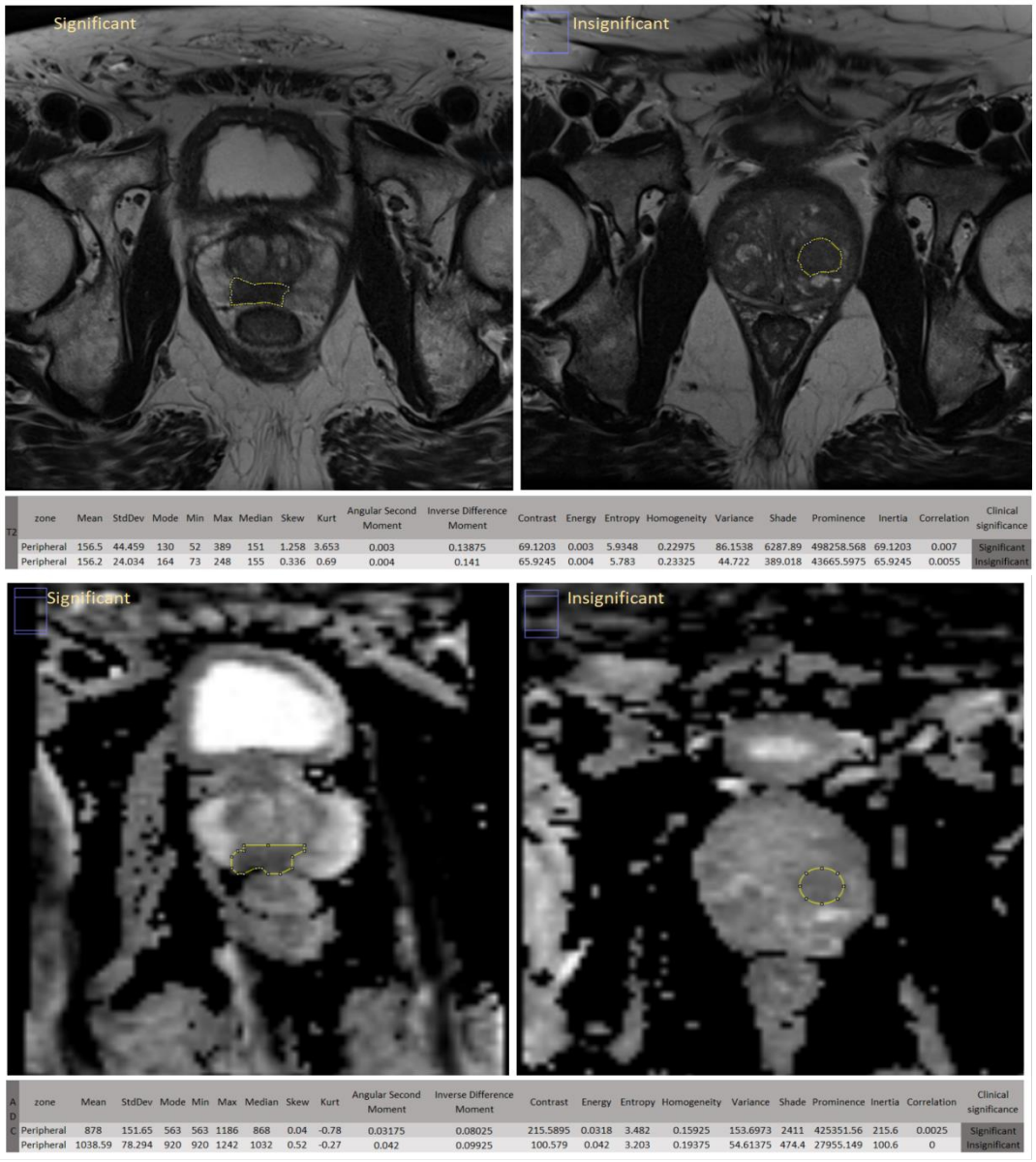
Prostate cancer is a well-recognized healthcare problem, accounting for the most diagnosed form of cancer in men.[1] Although it is considered the third leading cause of death among men, fortunately, it has a good prognosis when detected early, with a survival rate of 100% in the early stages and 31% in the late stages. In comparison, it has a 10-year survival rate of 98% in all cancer stages.[2] Therefore, prostate cancer detection has become a priority in recent decades. With computer-aided diagnosis (CAD), which mainly improves radiologists' diagnostic process,[3] many studies have focused on prostate cancer localization and detection, proposing promising methods for CAD in prostate cancer detection.[4,5] These studies were conducted using various textural analysis methods, MRI sequences, MRI magnetic powers, and classifiers.[4,5] While putting into consideration the limitless number of options when selecting imaging parameters available in MRI machines, let aside the differences in machine type, model, and available features; This numerous variables in fact can make it inapplicable in practice.[6] Therefore, to overcome these technical factors, Haralick features were successfully introduced.[6,7] Haralick et al. presented a mathematically originated textural features extraction method in which the first and second statistic order features can be extracted, as they rely on the occurrence of a specific spatial gray tone about its neighboring tones.[8] First-order features (mean, variance, standard deviation, and skewness) are values drawn from histograms based on a gray level image. The second-order features are derived from the gray-level co-occurrence matric (GLCM), which is a quantifying method of the spatial relation in neighboring gray tones in an image,[8] as seen in Fig.1. Previously, the availability of proper online datasets was challenging.[5] This motivated many to upload more datasets than that are currently available online.[9,10,11] Many studies were conducted on prostate cancer detection using CAD. However, the insufficient number of studies using Haralick textural analysis motivated us to further validate and enrich the technique through our work, expecting that using an online dataset will further aid in the growth of CAD. We built a classifier based on textural analysis to enable the classification of prostate lesions into clinically significant (malignant) or insignificant (benign) lesions using the MRI

T2 WI and diffusion-weighted image (DWI) derivatives' apparent diffusion coefficients (ADC).

METHODS

We extracted first- and second-order statistical Haralick features[12] with 38 attributes from publicly accessible data and retrospectively collected MRI prostate studies pathologically proven and found in The Cancer Imaging Archive (TCIA).[9,10] For confirmation, all patients underwent MRI-guided biopsy, and the lesions' histopathological Gleason scores determined clinical significance: seven and higher labeled as clinically significant and six and below as clinically insignificant.[10]

All lesions were manually traced by a senior radiology registrar and validated by a 10-year-radiology-consultant using Image-J, which is an open-source software,[13,14] as shown in Fig. 1. First, the dataset and information files were downloaded from the TCIA. Second, using Image-J for segmentation, image sequences in axial T2 WI and ADC were uploaded for each lesion and converted into an 8-bit gray scale format, where the 2D segmentation method was used to evaluate every selected area. The localization process was performed using the dataset source files containing each lesion's corresponding coordinates information in addition to the image thumbnail for each lesion. Third, each lesion was manually traced and duplicated to eliminate extra-lesional features. Fourth, for each lesion, texture extraction was performed using the GLCM textural analysis method at four angles (0°, 45°, 90°, and 135°).



	zone	Mean	StdDev	Mode	Min	Max	Median	Skew	Kurt	Angular Second Moment	Inverse Difference Moment	Contrast	Energy	Entropy	Homogeneity	Variance	Shade	Prominence	Inertia	Correlation	Clinical significance
T2	Peripheral	156.5	44.459	130	52	389	151	1.258	3.653	0.003	0.13875	69.1203	0.003	5.9348	0.22975	86.1538	6287.89	498258.568	69.1203	0.007	Significant
	Peripheral	156.2	24.034	164	73	248	155	0.336	0.69	0.004	0.141	65.9245	0.004	5.783	0.23325	44.722	389.018	43665.5975	65.9245	0.0055	Insignificant

	zone	Mean	StdDev	Mode	Min	Max	Median	Skew	Kurt	Angular Second Moment	Inverse Difference Moment	Contrast	Energy	Entropy	Homogeneity	Variance	Shade	Prominence	Inertia	Correlation	Clinical significance
ADC	Peripheral	878	151.65	563	563	1186	868	0.04	-0.78	0.03175	0.08025	215.5895	0.0318	3.482	0.15925	153.6973	2411	425351.56	215.6	0.0025	Significant
	Peripheral	1038.59	78.294	920	920	1242	1032	0.52	-0.27	0.042	0.09925	100.579	0.042	3.203	0.19375	54.61375	474.4	27955.149	100.6	0	Insignificant

Figure 1 : The upper half shows an example of lesion tracing and data extraction schedule in the T2 WI sequence for significant and insignificant lesions. The lower half shows an example of lesion tracing and data extraction schedule in ADC sequence for clinically significant and insignificant lesions.

Fig. 2 illustrates the GLCM method. Finally, the extracted data were stored in an Excel spreadsheet for analysis using the WEKA platform. Our dataset comprised 330 prostate lesions from 204 subjects. We excluded a total of 178 lesions, of which ten were excluded due to incomplete image sequence acquisition, while 168 were excluded after a size-based filtration was made where small lesional size areas (less than 64 mm^2) with insufficient textural features to be extracted were filtered. The end dataset was composed of 152 instances (76 clinically significant malignant lesions versus 76 benign), which was then split into two datasets: 70% of the data was used as the training set (106 lesions: 54 clinically significant versus 52 insignificant) and 30% as the testing set (46 lesions: 22 clinically significant versus 24 insignificant). After applying correlation-based feature selection[15] using the WEKA machine learning platform for classification.[16] Classification process: The data were firstly randomized using Microsoft Excel and then uploaded to the WEKA platform where several classifiers and options were tested based on class (Significant/Insignificant lesion).

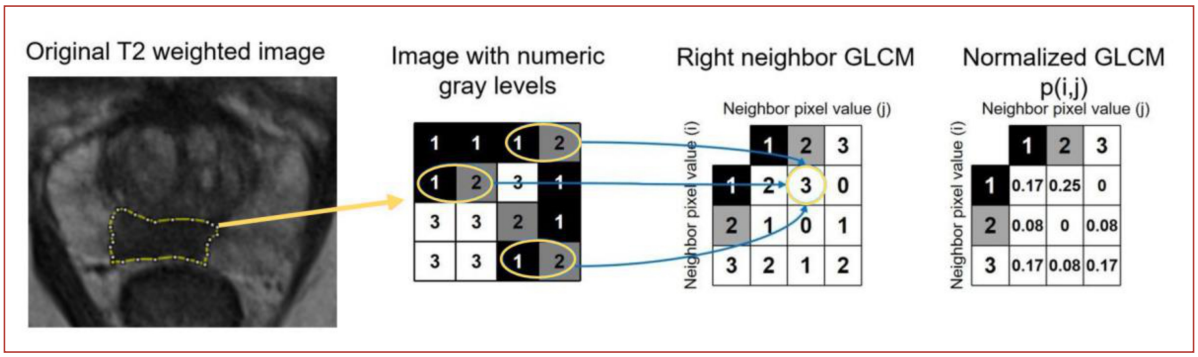


Figure 2: Illustration of the GLCM textural analysis method from an MRI T2 WI in 0°. The selected lesion numeric gray levels are first plotted, then each pixel's occurrence with its neighbors is counted, and lastly given a percentage accordingly. The example shows four pixels with a value of two, but only three pixels with a value of two as the right neighborhood for pixels with a value of one (three times), which accounts for 25% of the normalized GLCM.

After this, automatic classifier selection was performed on the data and Auto-WEKA plug-in options was primarily used. Lastly, multiple suggested best-performing classifiers were compared, revalidated, and retested manually with different parameter selections to obtain the best-performing classifier. the J48 classifier was finally selected. Table 1. A and B present a comparison between the three best-performing classifiers. Mathematical consideration: all mathematical formulas and relations were calculated automatically by WEKA platform. However, the used formulas were retested

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{Total Positive}} \quad \text{Specificity} = \frac{\text{True Negative}}{\text{Total Negative}}$$

(Sensitivity and specificity are inversely proportional)

$$\text{F measures} = \frac{\text{True Positive}}{\text{Total Positive} + \frac{1}{2}(\text{False Postive} + \text{False Negative})}$$

While Area under the receiver operating characteristic curves (AUROC) is calculated by integrating a function that best models true positive rate (TPR) as a function of false positive rate (FPR), which are calculated as:

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}} \quad \text{and FPR} = \frac{\text{FP}}{\text{FP+TN}}$$

where TP is the frequency of true positive results, FN is the frequency of false negative results, FP is the frequency of false positive results, and TN is the frequency of true negative results. The closer the AUROC curve approaches 1.0 the higher the accuracy of the test to diagnose the condition in question. The AUROC curves are plotted using "Threshold Curve" viewer option in WEKA after obtaining the results. The selected classifier underwent further evaluation with an attribute correlation evaluator ranker.

Obtaining the data did not require ethical board approval or patient's informed consent, as it has been made available for public use by TCIA. Fig. 3. show a simplified methodology.

ETHICAL APPROVAL

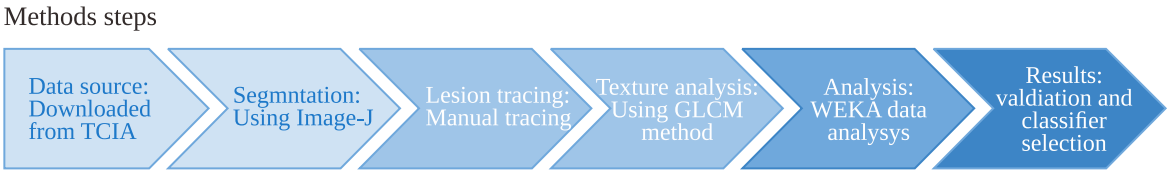


Figure 3: Show flow diagram for methodology used in six steps

RESULTS

J48 classifier yielded a tree comprising 4 leaves and 7 nodes, based on which the clinical significance was determined. To analyze the data, WEKA analysis was performed, and the following attributes were selected: ADC Entropy, ADC minimum, and ADC angular second moment. (Fig. 4). Moreover, 6 out of 38 attributes were selected for their high-class correlation, as generated by the attribute ranker, where the 3 highest features from each T2 WI and ADC images with the highest correlation by Entropy, Energy and Angular second moment for both imaging sequence were selected (Table 2). The proposed classifier can effectively discriminate clinically significant (malignant) prostrate lesions from the insignificant (benign) ones. Fig. 1

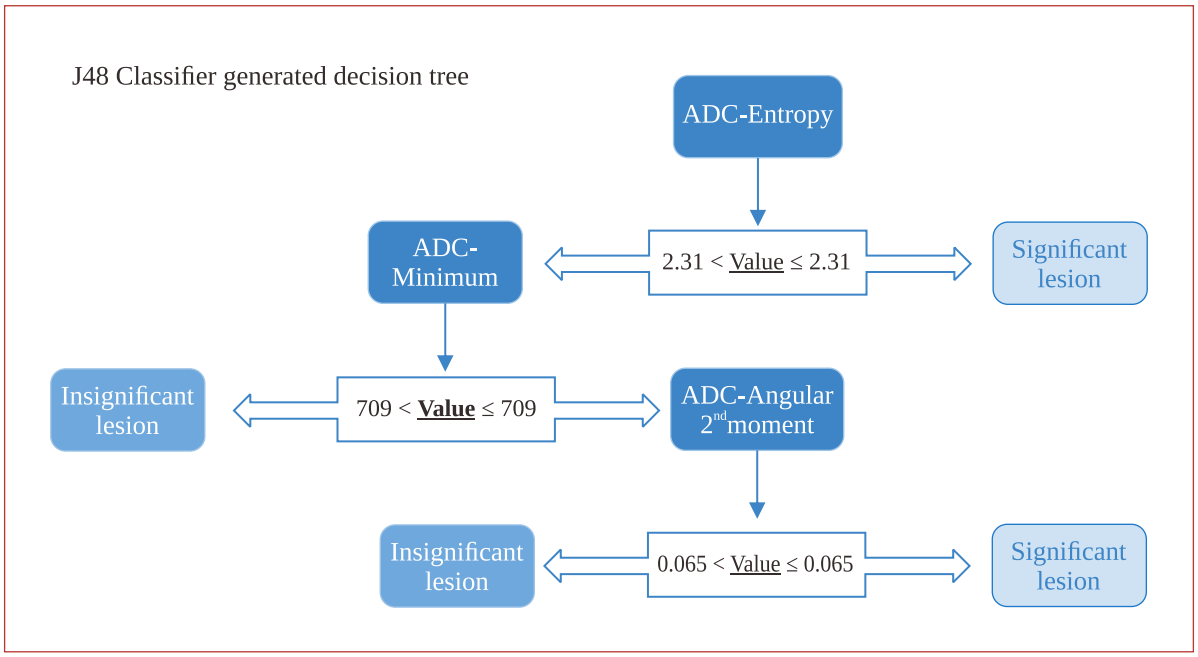
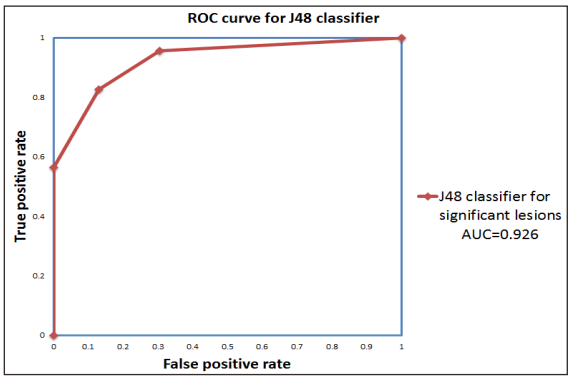
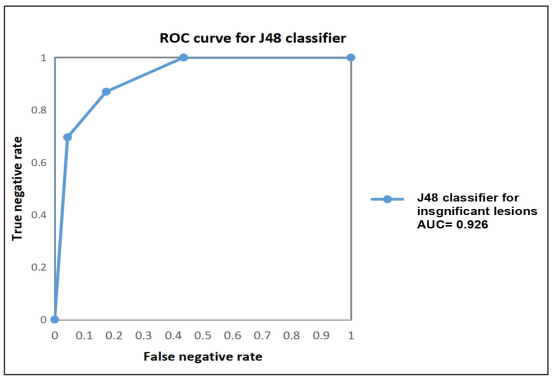


Figure 4: Demonstrate a decision tree that is generated by J48 algorithm. It shows 3 main attributes and their values used to determine lesions' clinical significance which are: ADC-Entropy, ADC-Minimum and ADC-angular 2nd moment respectively.

shows an example of a significant and insignificant lesion on T2 WI and ADC. According to the results of the WEKA analysis, 84.8% (39/46) lesions were correctly classified (19 significant and 20 insignificant), whereas 15.2% (7/46) were incorrectly classified (3 significant and 4 insignificant), with a confidence interval of 0.95 and kappa statistic of 0.7. The sensitivity, specificity, F-measure, and AUROC values were 82.6%, 87%, 84.4%, and 92.6%, respectively, for the significant (malignant) prostate lesions, and 87%, 82.6 %, 85.1%, and 92.6%, respectively, for the insignificant (benign) lesions. Fig. 5 shows a plot of the AUROC.



A. Area under receiver operating characteristic curve (AUROC) for clinically significant lesions is 92.6%



B. Area under receiver operating characteristic curve (AUROC) for clinically insignificant lesions is 92.6%

Figure 5: Demonstrate the receiver operating characteristic curves for clinically significant lesions (A) and clinically insignificant lesions (B).

DISCUSSION

Our method utilizing only T2 WI and ADC sequences outperformed the best performing method in the PROSTATEx challenges, which achieved an area under curve (AUC) value of 89% compared to the 92.6% achieved using our method. However, our study used training data alone, which may affect the overall results.[17]

We conducted multiple trials testing different classifiers using automatic and manual dataset analysis approaches and size-based filtration trails. The random forest, J48, and support vector machine classifiers were mainly used in addition to other classifiers such as multilayer perceptrons and meta vote. The J48 classifier achieved the best result. Moreover, for the size-based filtration different size cutoffs were tested to determine a better CAD capability. We found that a cut off of 64 mm2 in ADC sequences yielded better diagnostic results. Nevertheless, increasing the analyzed lesion's size did not enhance the results significantly. While lowering it below this size caused a significant decrease in the overall statistical values, which in particular can be attributed to the insufficient amount of extracted texture to be analyzed from these smaller lesions. Another limitation is the dataset itself, where we excluded more lesions due to incomplete image sequence acquisition, as some had a missing T2 WI or ADC image sequence. By contrast, some had multiple image files in the T2 WI sequence or incorrect coordination from the provided guide. Nevertheless, the presented classifier can adequately discriminate clinically significant from nonsignificant prostate lesions. However, the dataset did not include patient demographics and details concerning significant lesion advancement or staging, the conclusion drawn may be limited.

CONCLUSION

In conclusion, J48 classifier performed well in discriminating clinically significant malignant prostate lesions. However, focusing on future studies, using the same datasets with the same or different textural analysis techniques is essential to validate further and improve the results and AI field, while encouraging more dataset publication with more relevant clinical data would be extremely helpful. Texture extraction and J48 classifier do not require high-end computers to process pixels, and data can be deployed in portable computers.

CONFLICT OF INTEREST

Author and co-authors have no conflicts of interest to disclose.

ACKNOWLEDGMENTS

None

FUNDING SOURCES

No funding source/s for this study.

REFERENCES

1. Canadian Cancer Statistics Advisory Committee 2020. Prostate cancer statistics. [online] Canadian Cancer Society. Available at: <https://cancer.ca/en/cancer-information/cancer-types/prostate/statistics>.
2. American Cancer Society. Cancer Facts & Figures 2020. Atlanta: American Cancer Society; 2020.
3. Giger ML, Chan HP, Boone J. Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. Med Phys. 2008;35(12):5799–5820. doi:10.1118/1.3013555.
4. Wang S, Burtt K, Turkbey B, Choyke P, Summers RM. Computer aided-diagnosis of prostate cancer on multiparametric MRI: a technical review of current research. Biomed Res Int. 2014;2014:789561. doi:10.1155/2014/789561
5. Lemaître G, Martí R, Freixenet J, Vilanova JC, Walker PM, Meriaudeau F. Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. Comput Biol Med. 2015;60:8–31. doi:10.1016/j.compbiomed.2015.02.009
6. Wibmer A, Hricak H, Gondo T, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. Eur Radiol. 2015;25(10):2840–2850. doi:10.1007/s00330-015-3701-8
7. Niaf E, Rouvière O, Mège-Lechevallier F, Bratan F, Lartizien C. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Phys Med Biol. 2012;57(12):3833–3851. doi:10.1088/0031-9155/57/12/3833
8. Haralick R. Statistical and structural approaches to texture. Proceedings of the IEEE. 1979;67(5):786–804. doi:10.1109/proc.1979.11328
9. Clark K, Vendt B, Smith K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging. 2013;26(6):1045–1057. doi:10.1007/s10278-013-9622-7
10. Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman. "ProstateX Challenge data", The Cancer Imaging Archive (2017). DOI: 10.7937/K9TCIA.2017.MURS5CL

11. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. IEEE Trans Med Imaging. 2014;33(5):1083–1092. doi:10.1109/TMI.2014.2303821

12. Haralick R, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Trans Syst Man Cybern. 1973;SMC-3(6):610–621. doi:10.1109/tsmc.1973.4309314

13. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676–682. Published 2012 Jun 28. doi:10.1038/nmeth.2019.

14. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–675. doi:10.1038/nmeth.2089.

15. Hall M, et al. Correlation-based feature selection for machine learning, PhD Thesis, 1999 New Zealand Department of Computer Science, Waikato University.

16. Eibe Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & DATA, M. (2005, June). Practical machine learning tools and techniques. In Data Mining (Vol. 2, No. 4).

17. Armato SG 3rd, Huisman H, Drukker K, et al. PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. J Med Imaging (Bellingham). 2018;5(4):044501. doi:10.1117/ 1.JMI.5.4.044501.