Bladder Cancer Staging in CT Urography: Estimation and Validation of Decision Thresholds for a Radiomics-Based Decision Support System

Dhanuj Gandikota[1], Lubomir Hadjiiski[1], Heang-Ping Chan[1], Kenny H. Cha[2], Ravi Samala[1], Elaine M. Caoili[1], Richard H. Cohan[1], Alon Weizer[4], Ajjai Alva[3], Chintana Paramagul[1], Jun Wei[1], Chuan Zhou[1]

[1]Department of Radiology, University of Michigan, Ann Arbor, MI
[2]Center for Devices and Radiological Health, U.S. FDA, Silver Spring, MD
[3]Department of Internal Medicine-Hematology/Oncology, University of Michigan, Ann Arbor, MI
[4]Department of Urology, University of Michigan, Ann Arbor, MI

ABSTRACT

Stage T2 is the clinical threshold for deciding whether to treat bladder cancer with neoadjuvant chemotherapy. In this study we refined a radiomics-based decision support system (CDSS-S) to aid clinicians in staging of bladder cancer in CT urography (CTU). To train the CDSS-S, we used a data set of 84 bladder cancers from 76 CTU clinically staged cases, 43 cancers were below stage T2, and 41 were stage T2 or above. An independent test set comprising of 82 bladder cancers from 80 CTU clinically staged cases that were staged as T2 or above were also collected. Our Auto-Initialized Cascaded Level Sets (AI-CALS) segmentation pipeline was utilized to segment the lesions from which radiomics features were extracted. The training set was split on 2 balanced partitions. Four classifiers were studied: linear discriminant analysis (LDA), support vector machines (SVM), back-propagation neural networks (BPNN), and random forest (RAF) classifiers. Based on the likelihood scores for a training set, the decision threshold providing the highest classification accuracy for each classifier was determined. The classifier with the fixed decision threshold was then applied to the test set and the performance evaluated. The test classification accuracy for the LDA, SVM, BPNN, and RAF trained on Partition 1 was 0.95, 0.98, 0.88, and 0.89, respectively, and was 0.88, 0.94, 0.88, and 0.93, respectively, when trained on Partition 2. The test classification accuracy for the LDA, SVM, BPNN, and RAF trained on the entire training set was 0.94, 0.94, 0.94, and 0.89, respectively. The results show the potential of CDSS-S in bladder cancer stage assessment.

Keywords: Bladder Cancer Staging, Radiomics, Classification, Segmentation

1. INTRODUCTION

One of the most prominent forms of cancer that affects both men and women is urinary bladder cancer. In 2018, 81,190 new cases of bladder cancer with 17,240 deaths[1] were estimated. The accurate staging assessment of bladder cancer is pivotal in the decision for administering neoadjuvant chemotherapy to a patient.

Patients clinically diagnosed with a stage T2-T4 muscle invasive operable urothelial carcinoma of the bladder are prescribed neoadjuvant chemotherapy treatment. Multiple methods are utilized in the clinical staging of the bladder cancer. Clinical staging often consists of an extensive physical examination and a resection bimanual examination while under anesthesia. Pathology information is derived from TURBT (transurethral resection of bladder tumor). Cross-sectional imaging of the patient's bladder is also utilized in the clinical diagnosis. A computed tomography urography (CTU) slice of bladder cancer is shown in Figure 1. Neoadjuvant chemotherapy is recommended when muscle invasion is present in the pathology, a significant mass is discovered in the bimanual exam, and/or cross-sectional imaging indicates a disease of T2 or greater [2-5]. Poor renal function found in pre-existing conditions may exclude a patient from chemotherapy. Approximately 30% of patients are under-staged or over-staged in clinical staging. The clinical staging misdiagnosis may be attributed to the variability and subjectivity of clinicians in applying the available diagnostic

information. In our previous pilot study, we developed a decision support system for the staging of bladder cancer trained with pre-treatment clinical staging information [6-8].

The gold standard in current clinical practice to obtain the accurate staging of a patient's bladder cancer is the pathological staging after cystectomy (the removal of the bladder). An objective decision support system trained with the pathological staging after cystectomy, may be useful for assisting clinicians in making more accurate and consistent staging assessments. Our goal is to develop a quantitative computerized decision support system (CDSS-S) that may be used to aid clinicians in the staging of bladder cancer in CTU and reducing staging errors.

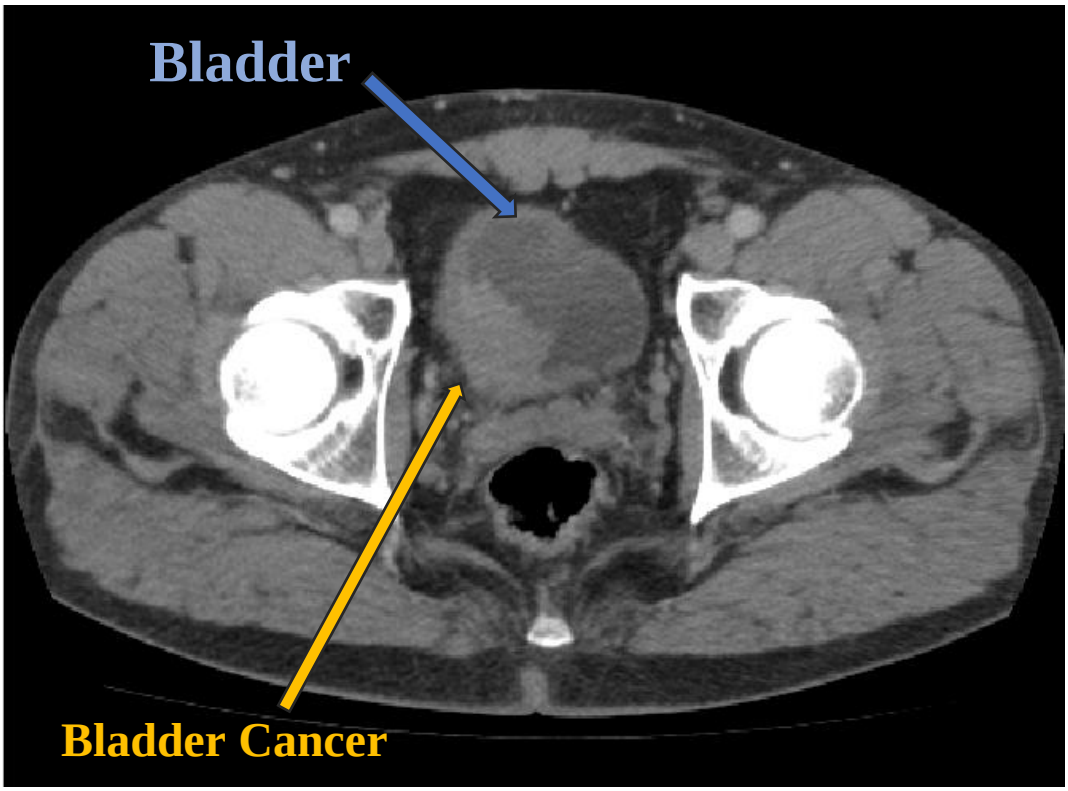

Figure 1: CT slice of the urinary bladder with a large malignant lesion.

2. MATERIALS AND METHODS

2.1 Data set

With IRB approval, we retrospectively collected a training set comprising of 84 bladder cancers from 76 CTU cases imaged before treatment. An independent test set of CTU scans prior to undergoing chemotherapy treatment was also collected comprising of 82 bladder cancers from 80 cases. All cases had clinical staging and also cystectomy-determined stages as reference standard. The lesions were categorized into two classes: (A) staged at or above T2 or (B) staged below T2. Of the 84 bladder cancers in the training set, 43 of the lesions were below stage T2, and 41 were stage T2 or above. The 82 cancers from the 80 patients in the test set were staged as T2 or higher.

2.2 Segmentation and Feature Extraction of Bladder Lesions

Segmentation proves difficult for lesions found within the bladder as they often varied in size with uneven boundaries. Additionally, lesion boundaries observed in the non-contrast enhanced-regions of the bladder are difficult to discern with high accuracy because of the low contrast between the lesion and its surroundings. We utilized our previously developed auto-initialized cascaded level sets (AI-CALS) segmentation pipeline. A radiologist-provided bounding box enclosing each lesion in the CTU volume is used as input for the AI-CALS segmentation pipeline. Preprocessing is applied to the input volume of interest and used to estimate the initial contour. The cascaded 3-D/2-D level set was initialized with the initial contour to obtain a 3-D lesion outline as shown in Figure 2. Eighty-nine radiomics features were extracted from each of the 168 segmented lesions. The feature set included lesion volume, 64 shape descriptors, and 24 gray level descriptors, from which useful features were selected using the training set.

2.3 Classification

Partition 1 contained 22 cancers staged below T2 and 20 cancers staged T2 or higher. Partition 2 contained 21 cancers staged below T2 and 21 cancers staged T2 or higher. Four types of classifiers, linear discriminant analysis (LDA), support vector machine (SVM), back-propagation neural network (BPNN) and random forest (RAF), were trained to merge the selected radiomics features and estimate a likelihood score for prediction of T2 stage. The classifiers were trained on each training partition, as well as on the entire training set, and then evaluated on the independent test set.

For each training partition, LDA with stepwise feature selection was used to select the best subset of features with Fin and Fout values optimized for each partition. Three features were selected for Partition 1 and seven features were selected for Partition 2. The selected feature set was used by the LDA, SVM, and BPNN classifiers to merge into a likelihood score with weights trained with the corresponding training partition. The BPNN contained a single output node and a single hidden layer. The RAF classifier, implemented by WEKA, was trained on each partition using all 89 features with 100 trees and 5 features per tree specified at the input. In addition, each of the four classifiers was trained on the entire training set that combined Partition 1 and Partition 2. The LDA classifier with stepwise feature selection selected 5 features when trained on the combined training set with Fin and Fout values used from the training partitions.

For a given classifier, based on the training likelihood scores obtained on a given training set, a decision threshold that maximized the classification accuracy was determined. The classification accuracy was defined as the sum of true positives and true negatives divided by the total number of cancers. The decision threshold was then applied to the independent test set and the performance of the CDSS-S was evaluated in terms of the classification accuracy on the test cases.

3. RESULTS

For the CDSS-S utilizing LDA, the training classification accuracy on Partition 1 and Partition 2 was 0.88 and 1.0, respectively, while the test classification accuracy was 0.95 and 0.88, respectively. For the CDSS-S utilizing SVM, the training classification accuracy on Partition 1 and Partition 2 was 0.88 and 1.0, respectively, with corresponding test classification accuracy of 0.98 and 0.94, respectively. The training classification accuracy for the BPNN based CDSS-S trained on Partition 1 and Partition 2 was 0.93 and 1.0 respectively, while the test classification accuracy was 0.88 for both. For the RAF based CDSS-S, the training classification accuracy on Partition 1 and Partition 2 was 1.0 for both, while the test classification accuracy was 0.89 and 0.93, respectively. The training classification accuracy for the LDA, SVM, BPNN, and RAF based CDSS-Ss trained on the entire combined training set was 0.92, 0.94, 0.93, 1.0, respectively, while the corresponding test classification accuracy was 0.94, 0.94, 0.94, and 0.89, respectively.

Tables 1-4: Training and testing results for each CDSS-S model. Table 1 used LDA, Table 2 used SVM, Table 3 used BPNN, and Table 4 used RAF. Within these tables, classification accuracy is denoted by CA, sensitivity is denoted by SE and specificity is denoted by SP. The classifier scores were normalized to between 1 and 100. The number of true positives, true negatives, false positives, and false negatives are denoted by TP, TN, FP, and FN, respectively.

LDA	Train AUC	Train CA	TP	TN	FP	FN	Train SE	Train SP	Thresh	Test CA(SE)	TP	FN
Part1	0.92	0.88	19	18	4	1	0.95	0.82	45.3	0.95	78	4
Part2	1	1	21	21	0	0	1	1	51.5	0.88	72	10
Comb	0.95	0.92	40	37	6	1	0.98	0.86	42.8	0.94	77	5

SVM	Train AUC	Train CA	TP	TN	FP	FN	Train SE	Train SP	Thresh	Test CA(SE)	TP	FN
Part1	0.92	0.88	19	18	4	1	0.95	0.81	31.0	0.98	80	2
Part2	1	1	21	21	0	0	1	1	37.0	0.94	77	5
Comb	0.99	0.94	40	39	4	1	0.98	0.91	42.0	0.94	77	5

BPNN	Train AUC	Train CA	TP	TN	FP	FN	Train SE	Train SP	Thresh	Test CA(SE)	TP	FN
Part1	1	0.93	20	19	3	0	1	0.86	41.8	0.88	72	10
Part2	1	1	21	21	0	0	1	1	41.8	0.88	72	10
Comb	0.95	0.93	40	38	5	1	0.98	0.88	41.1	0.94	77	5

RAF	Train AUC	Train CA	TP	TN	FP	FN	Train SE	Train SP	Thresh	Test CA(SE)	TP	FN
Part1	1	1	20	22	0	0	1	1	34.0	0.89	73	9
Part2	1	1	21	21	0	0	1	1	32.0	0.93	76	6
Comb	1	1	41	43	0	0	1	1	25.0	0.89	73	9

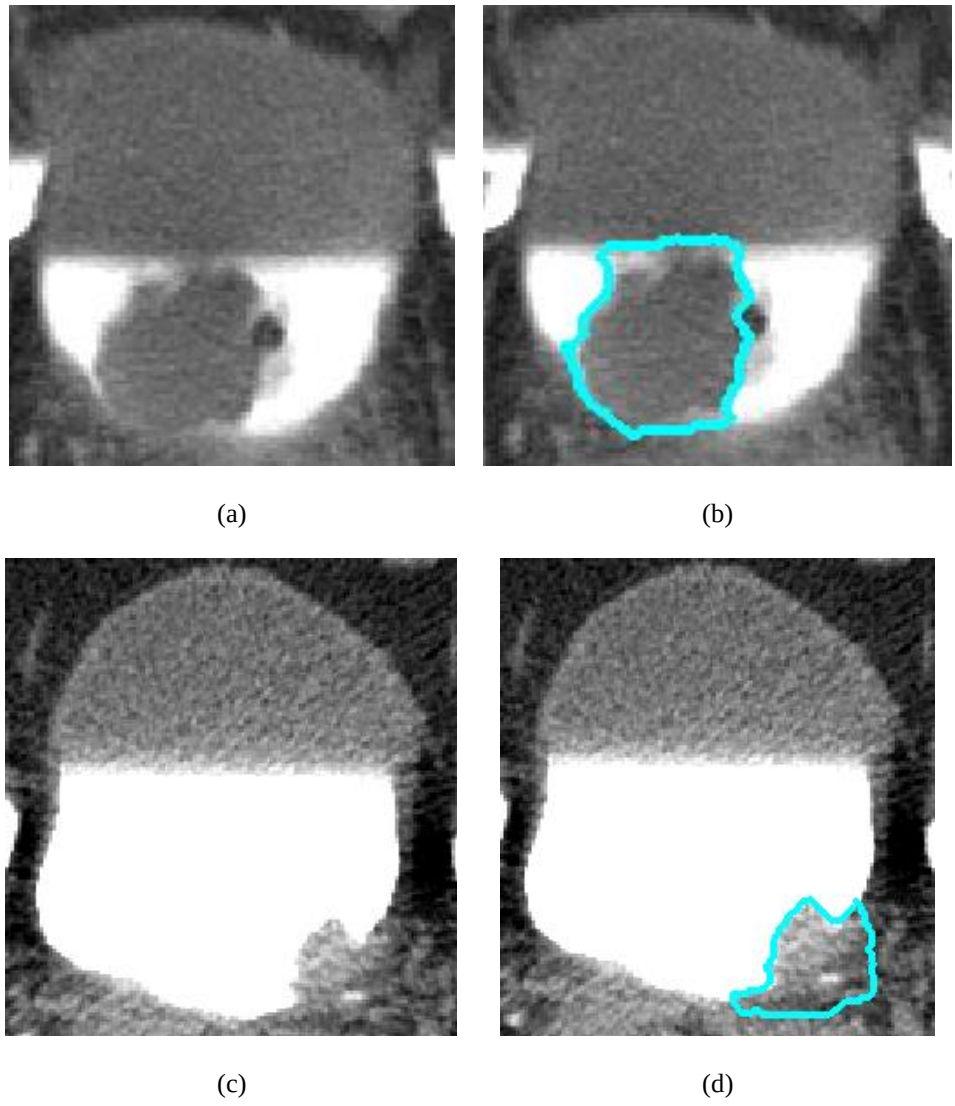

Fig. 2. Examples of CTU scans of bladder cancers from the independent test set. The blue outlines are the AI-CALS segmentation. Using the classifiers trained on the combined set of 84 bladder cancers, the following scores from the test set were obtained. The stage T2 cancer in (a), (b) was classified accurately as T2 or above, with scores of LDA=100, SVM=43.7, BPNN=100, and RAF=70.8, above the threshold of each classifier. The stage T2 cancer in (c), (d) was misclassified with scores of LDA=26.4, SVM=19.5, BPNN=23, and RAF=9.6, below the threshold of each classifier.

4. CONCLUSION

This study demonstrated the potential of using automatically extracted radiomic features from CTU and decision thresholds to build a statistical predictive model for the staging of bladder cancer. The high classification accuracy in pre-treatment CTU of bladder cancer cases validated the performance of the predictive models. Further work includes the collection of a larger data set and improvement in the predictive model accuracy through the inclusion of clinical and molecular data.

ACKNOWLEDGMENT

This work is supported by National Institutes of Health Grant number U01CA179106.

REFERENCES

[1] American Cancer Society, www.cancer.org 2018, "Cancer Facts & Figures 2018" (2018).
[2] Grossman H. B., Natale R. B., Tangen C. M., et al. "Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer," N Engl J Med 349, 859-866 (2003).
[3] Sternberg C. N., "The treatment of advanced bladder cancer," Ann Oncol 6, 113-126 (1995).
[4] Hadjiiski L., Mukherji S. K., Gujar S. K., et al. "Treatment response assessment of head and neck cancers on CT using computerized volume analysis," AJNR 31, 1744-1751 (2010).
[5] Meeks JJ., Bellmunt J., Bochner BH., et al. "A systematic review of neoadjuvant and adjuvant chemotherapy for muscle invasive bladder cancer," Eur Urol 62, 5230533 (2012).
[6] Hadjiiski L., Chan H-P., Caoili E. M., Cohan R. H., Wei J., Zhou C. "Auto-initialized cascaded level set (AI-CALS) segmentation of bladder lesions on multi-detector row CT urography," Acad Radiol 20, 148-155 (2013).
[7] Garapati SS, Hadjiiski L, Cha K, Chan HP, Caoili EM, Cohan RH, Weizer A, Alva A, Paramagul C, Wei J, Zhou C. "Automatic staging of bladder cancer on CT urography." Proc. SPIE Medical Imaging; 9785: 97851G (2016).
[8] Garapati SS, Hadjiiski L, Cha KH, Chan HP, Caoili EM, Cohan RH, Weizer A, Alva A, Paramagul C, Wei J, Zhou C. "Urinary Bladder Cancer Staging in CT Urography using Machine Learning." Med Phys; 44(11): 5814-5823 (2017).