

Deep Learning Approach for Assessment of Bladder Cancer Treatment Response

Eric Wu¹, Lubomir M. Hadjiiski¹, Ravi K. Samala¹, Heang-Ping Chan¹, Kenny H. Cha¹, Caleb Richter¹, Richard H. Cohan¹, Elaine M. Caoili¹, Chintana Paramagul¹, Ajjai Alva², and Alon Z. Weizer³

Departments of ¹Radiology, ²Internal Medicine-Hematology/Oncology, and ³Urology, University of Michigan, Ann Arbor, MI

Corresponding Author:

Eric Wu

Department of Radiology, University of Michigan, 1500 E. Medical Center Drive, MIB C473, Ann Arbor, MI 48109-5842;

E-mail: ehwu@umich.edu

Key Words: deep-learning, transfer learning, treatment response, segmentation, CT, bladder

Abbreviations: Deep learning-convolutional neural network (DL-CNN), regions of interest (ROIs), area under the ROC curve (AUC), methotrexate, vinblastine, doxorubicin, and cisplatin (MVAC), computed tomography (CT), magnetic resonance imaging (MRI)

ABSTRACT

We compared the performance of different Deep learning-convolutional neural network (DL-CNN) models for bladder cancer treatment response assessment based on transfer learning by freezing different DL-CNN layers and varying the DL-CNN structure. Pre- and posttreatment computed tomography scans of 123 patients (cancers, 129; pre- and posttreatment cancer pairs, 158) undergoing chemotherapy were collected. After chemotherapy 33% of patients had T0 stage cancer (complete response). Regions of interest in pre- and posttreatment scans were extracted from the segmented lesions and combined into hybrid pre-post image pairs (h-ROIs). Training (pairs, 94; h-ROIs, 6209), validation (10 pairs) and test sets (54 pairs) were obtained. The DL-CNN consisted of 2 convolution (C1-C2), 2 locally connected (L3-L4), and 1 fully connected layers. The DL-CNN was trained with h-ROIs to classify cancers as fully responding (stage T0) or not fully responding to chemotherapy. Two radiologists provided lesion likelihood of being stage T0 posttreatment. The test area under the ROC curve (AUC) was 0.73 for T0 prediction by the base DL-CNN structure with randomly initialized weights. The base DL-CNN structure with pretrained weights and transfer learning (no frozen layers) achieved test AUC of 0.79. The test AUCs for 3 modified DL-CNN structures (different C1-C2 max pooling filter sizes, strides, and padding, with transfer learning) were 0.72, 0.86, and 0.69. For the base DL-CNN with (C1) frozen, (C1-C2) frozen, and (C1-C2-L3) frozen, the test AUCs were 0.81, 0.78, and 0.71, respectively. The radiologists' AUCs were 0.76 and 0.77. DL-CNN performed better with pretrained than randomly initialized weights.

INTRODUCTION

Bladder cancer is the fourth most common cancer in men. The American Cancer Society estimates that in 2018, 81 190 (men, 62 380; women, 18 810) new cases of bladder cancer will be diagnosed in the United States, with 17 240 (men, 12 520; women, 4720) deaths (1). Early treatment of bladder cancer is important to reduce morbidity and mortality, as well as reduce costs.

Radical cystectomy is considered the gold standard for treatment of patients with localized muscle-invasive bladder cancer. However, about 50% of such patients develop metastases within 2 years after cystectomy and subsequently die of the disease (2). Neoadjuvant chemotherapy of muscle-invasive operable bladder cancer has been shown to be beneficial for treating micrometastases and improving resectability of larger neoplasms before radical cystectomy (3-5). Chemotherapy involving methotrexate, vinblastine, doxorubicin, and cisplatin (MVAC) followed by radical cystectomy increases the probability of finding no residual cancer at surgery compared with

radical cystectomy alone and improves survival among patients with locally advanced bladder cancer (6, 7). In clinical trials, downstaging with drugs before surgery was shown to have significant survival benefits (7, 8). Current standard of care uses the neoadjuvant protocol consisting of 12 weeks of chemotherapy preceding radical cystectomy.

Although patients with advanced disease can benefit from neoadjuvant chemotherapy, there are drawbacks. Chemotherapy with the MVAC regimen has substantial toxicity and side effects (9). Significant toxicities, primarily leucopenia, culture-negative fever at the time of granulocytopenia, sepsis, and mucositis are associated with MVAC combination chemotherapy. Side effects such as nausea, vomiting, malaise, and alopecia are common. In addition, chemotherapy is expensive. However, because no reliable method yet exists for predicting the response of an individual case to chemotherapies such as MVAC, some patients may suffer from adverse reactions to the drugs without achieving beneficial effects, often also missing the opportunity for alternative therapy when their physical condition deteriorates.

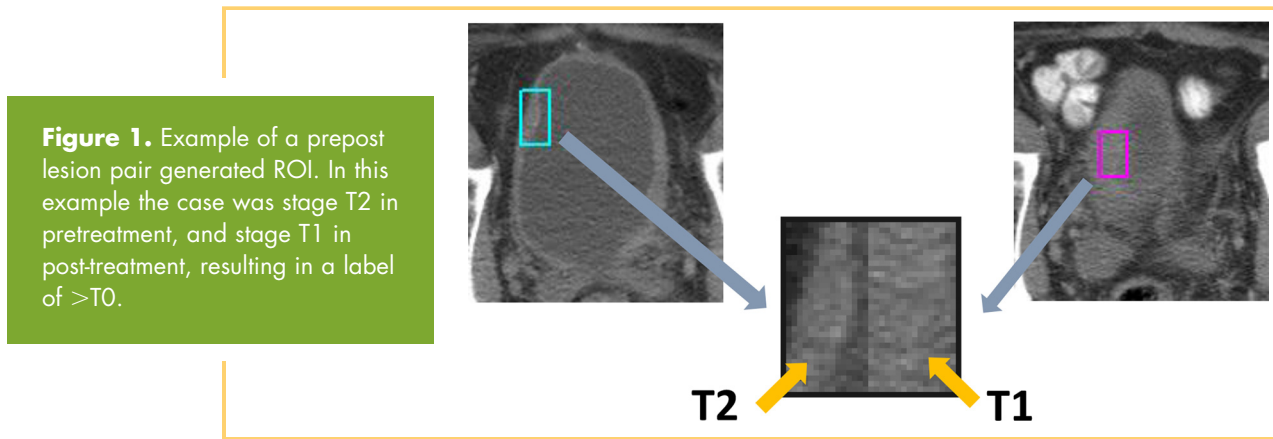


Figure 1. Example of a prepost lesion pair generated ROI. In this example the case was stage T2 in pretreatment, and stage T1 in post-treatment, resulting in a label of $>T_0$.

Early assessment of therapeutic efficacy and prediction of failure of the treatment would help physicians decide whether to discontinue chemotherapy at an early phase and thus reduce unnecessary morbidity and improve the quality of life of the patient, and reduce costs. The ultimate goal is to improve survival for those with a high risk of recurrence while minimizing toxicity to those who will have minimal benefit.

The development of an accurate predictive model for the effectiveness of a specific therapy and clinical evaluation of the predictive model are of critical importance for patients with bladder cancer. In addition, if a patient can be reliably identified as having complete response to treatment, the treatment option of preserving the bladder may be considered, which would drastically reduce the morbidity of the patient and improve his/her quality of life as compared to the current standard treatment by cystectomy.

Pathologic evaluation performed at the time of radical cystectomy is considered a “gold standard” for estimation of treatment response. However, this method cannot be used during the course of chemotherapy. Noninvasive evaluation of the treatment response can be performed during the course of chemotherapy (after 1 or 2 cycles) with computed tomography (CT) or magnetic resonance imaging (MRI) by measuring tumor size. CT provides accurate anatomical images of the tumor and is becoming the main tool for evaluation of bladder cancer.

We are developing a computerized decision support system (CDSS-T) for monitoring of bladder cancer treatment response. Machine learning techniques are used to integrate the image information into an effective predictive model. The purpose of the CDSS-T is to provide noninvasive, objective, and reproducible decision support for identifying nonresponders so that the treatment may be stopped early to preserve their physical condition or to identify full responders for organ preservation.

DL-CNN can be used to build pattern recognition models using large image data sets (10–12). There are an increasing number of DL-CNN applications in medical imaging field for lesion segmentation, characterization, and diagnosis of diseases in different organs (13).

Cha et al. (14) proposed DL-CNN-based method for treatment response assessment of bladder cancers. In their paper, the DL-CNN was trained directly on a pre- and posttreatment set of 82 patients with 87 bladder cancers and deployed on a test pre- and posttreatment set of 41 patients with 43 cancers.

In medical imaging where training image data sets are generally small, a commonly used approach for building robust DL-CNN models is transfer learning (15). This approach uses a large data set from a different domain (for example, natural scene images) to initially train the DL-CNN. Then most of the structures and the parameters of the DL-CNN are kept fixed and only a small part of the DL-CNN is retrained with the smaller data set from the specific domain of the task at hand, for which the model is designed. This approach has shown a lot of promise in a number of medical imaging applications (16–18).

In this study we have explored different DL-CNN models for bladder cancer treatment response assessment based on transfer learning by freezing different DL-CNN layers and varying the DL-CNN structure. We also compared the DL-CNN models to radiomics-based models.

METHODS

Data Set

Pre- and posttreatment CT scans of 123 patients (with 129 total cancers) undergoing chemotherapy were collected with IRB approval. In total, 33% of patients were determined to have T0 stage cancer (complete response) after chemotherapy.

After the chemotherapy treatment, each patient underwent cystectomy. The final cancer stage after treatment was determined on the basis of the pathology obtained from the bladder at the time of the surgery. The pathological cancer stage was used as the reference standard for response to treatment: complete response (stage T0) or not complete response (stage $>T_0$).

The CT scans were acquired with GE Healthcare LightSpeed MDCT scanners (120 kVp; 120–280 mA). The pixel size range was 0.586 to 0.977 mm and the slice thickness range was 0.5 to 7.5 mm.

The lesions on the pre- and posttreatment scans were segmented using our previously developed autoinitialized cascaded level sets system (19). ROIs of pre- and posttreatment scans of these patients were extracted from segmented lesions as 32×16 -pixel images, and pre- and posttreatment images of patients were combined to make hybrid pre-post image pairs in the form of 32×32 -pixel image ROIs. Figure 1 gives an example of a $>T_0$ lesion pair and how it is generated. Multiple ROIs were extracted from pre- and posttreatment images of the lesion and combined to obtain a number of hybrid pre-post image pairs for the same lesion. Each hybrid ROI was labeled as T0 (complete

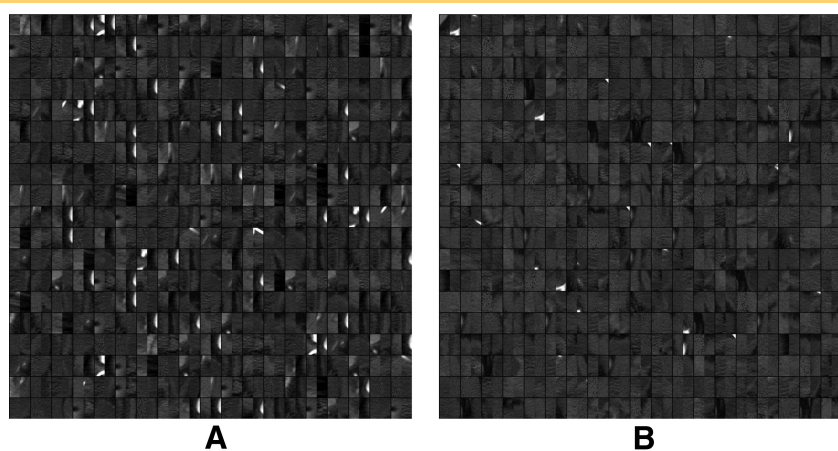


Figure 2. Subset of 6209 total regions of interest (ROIs) used in training set. Cases with complete response (T0) to treatment (A). Cases that did not fully respond (>T0) to treatment (B).

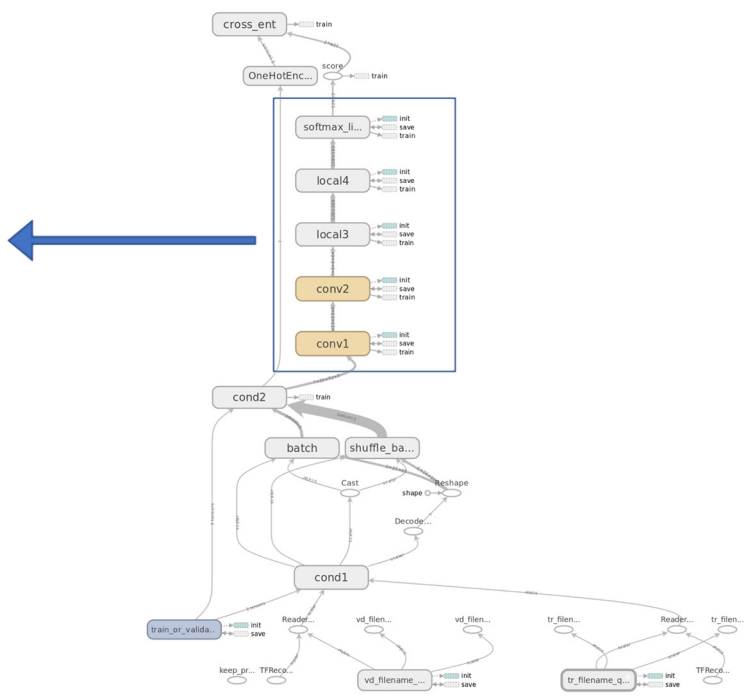
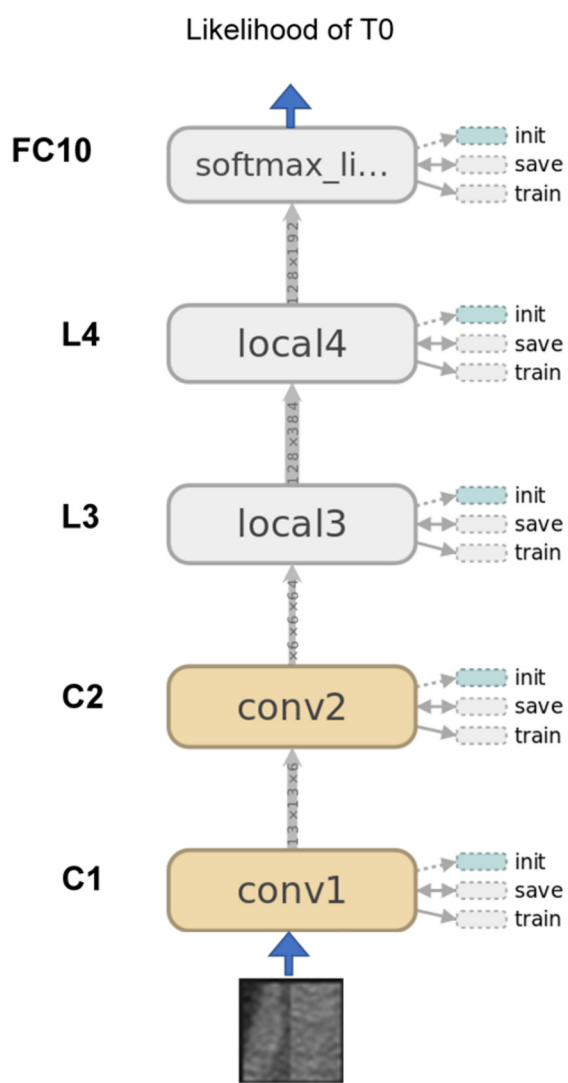


Figure 3. TensorFlow graph of base deep learning-convolutional neural network (DL-CNN) structure with different layers marked. The hybrid pre-post lesion pair ROIs were input to the DL-CNN, which then predicted a likelihood score of a complete response (T0) as an output.

Table 1. Modifications in Layers C1 and C2 for Each Structure Variation

	Base	DL-CNN-1	DL-CNN-2	DL-CNN-3
C1				
Convolution				
Size	5 × 5	5 × 5	5 × 5	5 × 5
Stride	1	1	2	1
Max Pooling				
Size	3 × 3	5 × 5	3 × 3	3 × 3
Stride	2	2	2	2
Padding	Valid	Valid	Valid	Same
C2				
Convolution				
Size	5 × 5	5 × 5	5 × 5	5 × 5
Stride	1	1	1	1
Max Pooling				
Size	3 × 3	2 × 2	2 × 2	4 × 4
Stride	2	1	1	2

response after treatment) or >T0 (the cancer did not respond completely after treatment) as determined by pathology.

The data set was split into training, validation, and test sets. The training set consisted of 77 lesions from 73 patients, where 19 lesions were stage T0, and 58 lesions were stage >T0. The 77 lesions formed 94 lesion pairs, and 6209 hybrid ROIs were generated. The validation set consisted of 10 lesions (stage T0, 5; stage >T0, 5) that formed 10 pre- and posttreatment cancer pairs and generated 521 hybrid ROIs. The test set was composed of 42 lesions from 41 patients, where 12 lesions were stage T0, and 30 lesions were stage >T0. The 42 lesions formed 54 pre- and posttreatment cancer pairs. Figure 2 displays 2 mosaics of different pre-post lesion pairs used in the training, with the left mosaic (Figure 2A) containing T0 pairs and the right (Figure 2B) containing >T0 pairs.

Two experienced radiologists, blinded to the clinical treatment outcome, also evaluated each pair of pre- and posttreatment CT scans in the test data set, displayed on 2 medical-grade monitors side by side, and provided ratings for the likelihood of the posttreatment lesions being stage T0 cancer.

Network Structures

The DL-CNN structure used in this study was based on AlexNet (10) and implemented and validated in the TensorFlow framework. The base structure of the DL-CNN consisted of 2 convo-

lution layers (C1 and C2) followed by 2 locally connected layers (L3 and L4) and a fully connected layer (FC10). The output from the DL-CNN was trained to classify cases as fully responding (stage T0) or not fully responding (stage > T0) to chemotherapy based on the hybrid ROIs. Within C1 and C2, convolution filtering with 64 “5 × 5” kernels and a stride of 1 was performed, followed by local response normalization and max pooling with a 3 × 3 filter of stride 2. Layer L3 consisted of 64 “3 × 3” kernels, and L4 consisted of 32 “3 × 3” kernels. The output from L4 was input to the FC10, which was a softmax linear layer. The FC10 layer produced a numerical likelihood score from 0 to 1, with 0 corresponding to a stage > T0 case, and 1 corresponding to a stage T0 case. Figure 3 shows a labeled map of the DL-CNN generated by TensorBoard, a visualization tool for TensorFlow.

We first trained the DL-CNN with randomly initialized weights. We then explored the use of transfer learning. The DL-CNN with pretrained weights from the CIFAR10 image set were used. The CIFAR10 image set consists of 10 classes (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck) and 60 000 total 32 × 32 images collected by Krizhevsky et al. Each class contains 6000 images (20). We also performed alterations to the DL-CNN structure to study its effect on the DL-CNN performance. The modifications of the structures took place in layers C1 and C2, and these involved the filter size, filter

Table 2. Test AUC Values for DL-CNN Models with Modified Structures

DL-CNN Type	Base DL-CNN Structure (Random Weights)	Base DL-CNN Structure (Pretrained Weights)	DL-CNN-1	DL-CNN-2	DL-CNN-3
AUC	0.73 ± 0.08	0.79 ± 0.07	0.72 ± 0.08	0.86 ± 0.06	0.69 ± 0.09

Table 3. Test AUC Values for DL-CNN Models with Transfer Learning and Different Frozen Layers

DL-CNN Type	Base DL-CNN Structure (Pretrained Weights)	C1 Frozen	C1, C2 Frozen	C1, C2, L3 Frozen
AUC	0.79 ± 0.07	0.81 ± 0.07	0.78 ± 0.08	0.71 ± 0.08

stride, and padding type of the convolutions and max pooling performed in each layer. Three different structures were studied (DL-CNN-1, DL-CNN-2, and DL-CNN-3), and the modifications performed can be observed in Table 1.

In addition, we trained the network with one (C1) or more (C1, C2, L3) layers frozen. Freezing a layer during training prevents its weights from being altered, and it may be necessary to preserve the starting weights for some layers of the network to optimize training results (21). All of the experiments with frozen layers used the CIFAR10 transfer learning and the original DL-CNN network structure.

Training and Testing Process

The DL-CNN models were trained first for 10 000 epochs by using the training data set. For every 100 epochs, the trained DL-CNN model was deployed on the validation set. The area under the ROC curve (AUC) was calculated as a performance measure, and the validation AUC results were recorded. To reduce the likelihood of overfitting, a line plot of the validation AUC results was created and a training epoch number around where the validation AUCs peaked (usually around 2000 epochs) was selected. The final DL-CNN model was trained on the combined training set (comprising the merged training and validation sets) up to the selected epoch. The trained DL-CNN model was then deployed on the test set and the AUC was estimated.

Training for 10 000 epochs for 1 experiment typically took about 8.3 hours with an NVidia GeForce GTX 1080 Ti GPU. Final training with the combined set took about 1.7 hours. Deployment on the test set took less than 1 minute per case.

Evaluation

The AUC results of our experiments were compared with those of the 2 radiologists, as well as those from 2 radiomics feature-based classification methods (RF-SL and RF-ROI) by Cha et al. (14). The radiomics-based methods involved predicting the response of cases based on the estimated changes in automatically extracted features (including morphological, gray level, and texture features) between lesions in pre- and posttreatment scans. Cha et al. (14) also evaluated the performance of a similarly structured DL-CNN. The results of the variations in the DL-CNN structure and the transfer learning schemes were compared with those of the base structure. We generated ROC curves for each experiment and used 2 statistical significance tests, ROC-kit from the University of Chicago, and the DeLong Test, to estimate the statistical significance of the differences between AUC values of the corresponding experiments. In addition, using the ROC curves, we calculated the sensitivity and accuracy of the test results at specificity of 80%, and statistical significance of the differences was also estimated. The specificity of

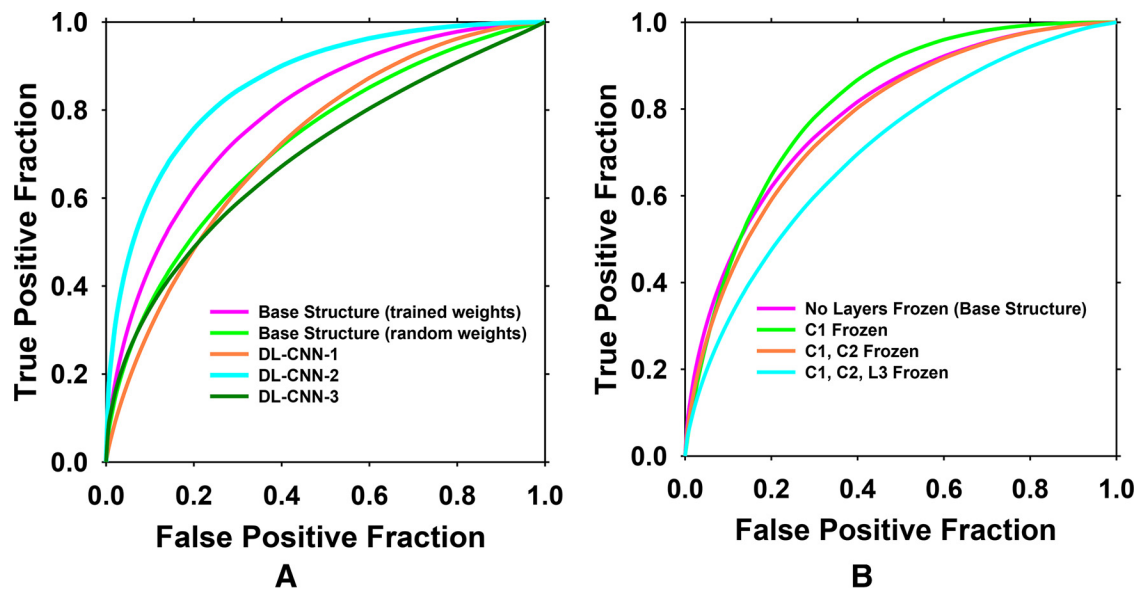


Figure 4. Test ROC curves of different DL-CNN models. ROC graph comparing base DL-CNN model (base structure) to DL-CNN models with modified structure (A). ROC graph comparing base DL-CNN model (base structure) with pretrained weights but no frozen layers to DL-CNN models with frozen layers (B).

Table 4. Test AUC Values for Radiologists and Methods Used in Cha et al. Study

DL-CNN Type	Base DL-CNN Structure (Random Weights)	Radiologist 1	Radiologist 2	DL-CNN (Cha)	RF-SL	RF-ROI
AUC	0.73 ± 0.08	0.76 ± 0.08	0.77 ± 0.08	0.73 ± 0.08	0.77 ± 0.08	0.69 ± 0.08

80% was selected by an experienced urologist (A.W.), as a possible clinically meaningful value.

RESULTS

The AUCs for our experiments are shown in Tables 2 and 3, and the ROC curves are shown in Figure 4. For the base DL-CNN structure with randomly initialized weights, the test AUC for T0 prediction was 0.73 ± 0.08. For the base DL-CNN structure, with transfer learning using CIFAR10 pretrained weights and no frozen training layers, the test AUC was 0.79 ± 0.07. The test AUCs for the DL-CNN-1, DL-CNN-2, and DL-CNN-3 modified structures (with transfer learning and no frozen layers) were 0.72 ± 0.07, 0.86 ± 0.06, and 0.69 ± 0.09, respectively. The only statistical significance difference observed was between DL-CNN-2 and DL-CNN-3 ($P = .007$, DeLong; $P = .006$, ROC-kit).

With the first layer (C1) of the base DL-CNN frozen, the test AUC was 0.81 ± 0.07. With the first 2 layers (C1 and C2) frozen, the test AUC was 0.78 ± 0.08. With the first 3 layers (C1, C2, and L3) frozen, the test AUC was 0.71 ± 0.08. None of the differences in AUC between the DL-CNN with frozen layers and the base structure with no layers frozen reached statistical significance.

Table 4 shows the AUC of the base DL-CNN with randomly initialized weights versus the radiologists and methods from the Cha et al. study (14). The AUCs of radiologist 1 and radiologist 2 were 0.76 ± 0.08 and 0.77 ± 0.08, respectively. The AUCs of the radiomics-based methods RF-SL and RF-ROI were 0.77 ± 0.08 and 0.69 ± 0.08, respectively. The network structure used in the study by Cha et al. achieved an AUC of 0.73 ± 0.08.

Table 5 shows the sensitivity and accuracy of each model at a specificity of 80%. The corresponding sensitivities ranged from 41.7% to 75.0%, while the corresponding accuracies ranged from 64.1% to 78.9%. Neither of the differences in sensitivities and accuracies between models reached statistical significance.

DISCUSSION

The results of this study show the feasibility of DL-CNN in estimating bladder cancer treatment response in CT. The DL-CNN performed better with pretrained weights from the CIFAR-10 image set than with randomly initialized weights, while the AUC from the randomly initialized weights matched that of the network structure used in the previous Cha et al. study (14). The base DL-CNN and its modified structures all performed similarly to the radiologists, and in a few cases, performing better with higher AUCs. The AUCs of the base DL-CNN and its variations were comparable to the AUCs of the radiomics-based methods from the Cha et al. study. Only 1 network variation (DL-CNN-2) resulted in a statistically significant improvement in performance compared to the base structure.

Figure 5 shows examples of pre- and postlesion pairs predicted correctly and incorrectly by the base DL-CNN with CIFAR10 weights.

The performance of the DL-CNN generally decreased as more training layers were frozen. Freezing layer C1 resulted in a slight, but not statistically significant, improvement in performance. According to a study by Yosinski et al. (22), the first layer of neural networks trained on natural images aims, in general, to capture more universal features (such as edges and curves), while proceeding layers aim to capture features more specific to the input image set (in this case, bladder lesions). As a result, allowing the first layer to train and change its weights may have minimal or adverse effects on the results of the training. Such a phenomenon may have been observed in our experiments, given the performance increase in our network with layer C1 frozen.

Similar trends were observed by Samala et al. (23) for the task of classification of malignant and benign breast masses on mammograms and tomosynthesis.

In our statistical significance tests, we found that one of our structure modifications, DL-CNN-2 (with the highest AUC value of all structures), achieved statistically significant improvement in performance compared to DL-CNN-3 (with the lowest AUC value of all structures). We will perform further testing to confirm the validity of our results and measure the performance of the structure with a larger data set.

There are limitations in this study. We are currently working with a relatively small data set in training, validation and testing of our DL-CNN models, which may also be a reason for achiev-

Table 5. Test Sensitivity and Accuracy of DL-CNN Models at a Specificity of 80%

	Sensitivity (%)	Accuracy (%)
Base Structure (Pretrained weights)	59.5%	64.1%
Base Structure (Random Weights)	41.7%	71.5%
Structure Modifications		
DL-CNN-1	50.0%	73.3%
DL-CNN-2	75.0%	78.9%
DL-CNN-3	50.0%	73.3%
Layer Freezing		
C1 Frozen	58.3%	75.2%
C1, C2 Frozen	58.3%	75.2%
C1, C2, L3 Frozen	58.3%	75.2%

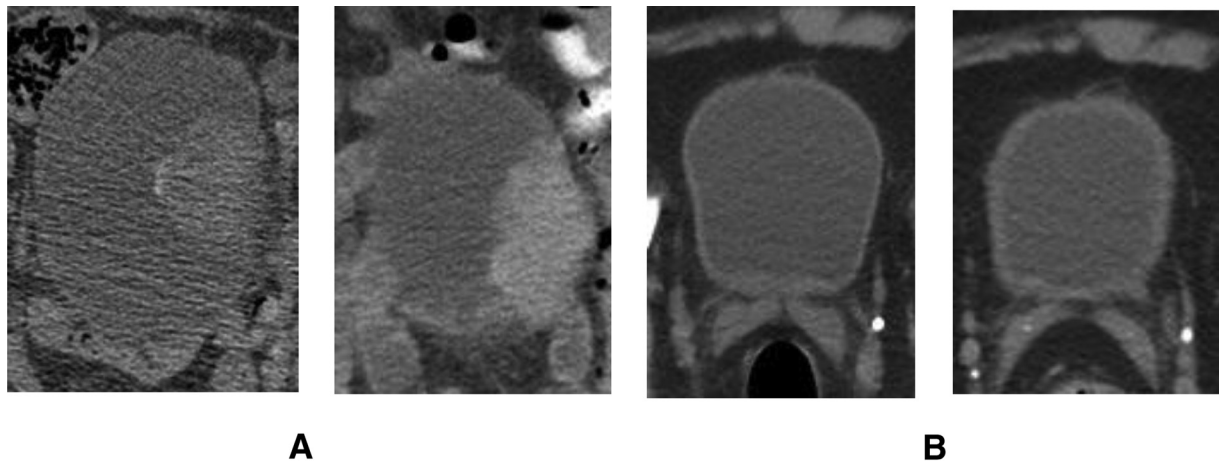


Figure 5. Examples of cases that the DL-CNN predicted correctly and incorrectly. The base DL-CNN with transferred weights correctly predicted this lesion as $>T0$ (A). The base DL-CNN with transfer learning incorrectly predicted this T0 lesion as $>T0$ (B).

ing statistical significance for only 1 comparison. In the future, we will continue to collect a larger data set with new cases (both T0 and non-T0) in our networks. Another limitation is that we have evaluations from only 2 radiologists on the test set. Additional classifications from different radiologists would be needed to study the variability in the accuracy of such readings.

Our network was trained using the CIFAR-10 data set, which produces favorable results, but is not relevant in the field of medical imaging. A better approach for training with transfer learning would be to use CT scan images, ideally bladder scans, as pretrained weights. Several networks pretrained using CT scans exist, and we may, in the future, explore the use of such networks in training with our data set.

The pixel sizes of the CT scans used in our data set vary in the range of 0.586 to 0.977 mm², and slice thicknesses vary from 0.5 to 7.5 mm. While the nonuniform nature of the scans may be seen as a limitation, in that it may bias the training results, learning different sizes would help the network better handle variability which would be present in real clinical applications. While scans would ideally take place under the same conditions using the same scanner, this is very difficult to achieve in

clinical settings. Nevertheless, we may try in the future to match voxel sizes of scans using methods such as interpolation.

It is important to accurately assess a bladder cancer's response to treatment based on pre- and posttreatment lesion scans to determine what further treatment a patient will require, if any at all. While our current network structure has shown to classify cases with considerable accuracy, we will further improve the model and validate its generalizability in unknown cases. Because of the small data set, we used DL-CNNs of relatively small structures in this study. We will investigate if deeper DL-CNN models such as GoogLeNet Inception (24) and ResNet (25) may provide better performance when a large data set becomes available.

In conclusion, our results showed that DL-CNN can effectively predict the response of a bladder cancer lesion to chemotherapy, with many of our experiments comparing favorably to the performance of the radiologists. Adjusting the structure of the base network and freezing certain layers of the network during training may further improve the performance. This study suggests that the DL-CNN may be useful in conjunction with medical professionals as decision support for bladder cancer treatment response assessment.

ACKNOWLEDGMENTS

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

Disclosures: No disclosures to report.

REFERENCES

1. Key Statistics for Bladder Cancer. 2018. <https://www.cancer.org/cancer/bladder-cancer/about/key-statistics.html>.
2. Sternberg CN. The treatment of advanced bladder cancer. *Ann Oncol.* 1995;6: 113–126.
3. Fagg SL, Dawson-Edwards P, Hughes MA, Latief TN, Rolfe EB, Fielding JW. CIS-Diamminedichloroplatinum (DDP) as initial treatment of invasive bladder cancer. *Br J Urol.* 1984;56:296–300.
4. Raghavan D, Pearson B, Coorey G, Woods W, Arnold D, Smith J, Donovan J, Langdon P. Intravenous CIS-platinum for invasive bladder cancer – safety and feasibility of a new approach. *Med J Aust.* 1984;140:276–278.
5. Meeks JJ, Bellmunt J, Bochner BH, Clarke NW, Daneshmand S, Galsky MD, Hahn NM, Lerner SP, Mason M, Powles T, Sternberg CN, Sonpavde G. A systematic review of neoadjuvant and adjuvant chemotherapy for muscle-invasive bladder cancer. *Eur Urol.* 2012;62:523–533.

Conflict of Interest: The authors have no conflict of interest to declare.

6. Advanced Bladder Cancer Meta-analysis Collaboration. Neoadjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis. *Lancet*. 2003;361:1927–1934.
7. Grossman HB, Natale RB, Tangen CM, Speights VO, Vogelzang NJ, Trump DL, deVere White RW, Sarosdy MF, Wood DP Jr., Raghavan D, Crawford ED. Neoadjuvant chemotherapy plus cystectomy compared with cystectomy alone for locally advanced bladder cancer. *N Engl J Med*. 2003;349:859–866.
8. Splinter TAW, Scher HI, Denis L, Bukowski R, Simon S, Klimberg I, Soloway M, Vogelzang NJ, van Tinteren H, Herr H. The prognostic value of the pathological response to combination chemotherapy before cystectomy in patients with invasive bladder cancer. *J Urol*. 1992;147:606–608.
9. Witjes JA, Wullink M, Oosterhof GO, de Mulder P. Toxicity and results of MVAC (methotrexate, vinblastine, adriamycin and cisplatin) chemotherapy in advanced urothelial carcinoma. *Eur Urol*. 1997;31:414–419.
10. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Paper presented at: 26th Annual Conference on Neural Information Processing Systems 25 (NIPS 2012); December 03–08, 2012; Advances in Neural Information Processing Systems. 2012;2:1097–1105. Lake Tahoe, NV.
11. LeCun Y, Bengio Y, Hinton GE. Deep learning. *Nature*. 2015;521:436–444.
12. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85–117.
13. Lijens G, Kooi T, Bejnordi BE, Setio AAA, Ciampi F, Ghahfarooian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
14. Cha KH, Hadjiiski L, Chan HP, Weizer AZ, Alva A, Cohan RH, Caoili EM, Paramagul C, Samala RK. Bladder cancer treatment response assessment in CT using radiomics with deep-learning. *Sci Rep*. 2017;7:8738.
15. Pan SJ, Yang QA. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22:1345–1359.
16. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging*. 2016;35:1285–1298.
17. Samala RK, Chan HP, Hadjiiski LM, Helvie MA, Cha KH, Richter CD. Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. *Phys Med Biol*. 2017;62:8894–8908.
18. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J Med Imaging*. 2016;3:034501.
19. Hadjiiski LM, Chan HP, Caoili EM, Cohan RH, Wei J, Zhou C. Auto-initialized cascaded level set (AI-CALS) segmentation of bladder lesions on multi-detector row CT urography. *Acad Radiol*. 2013;20:148–155.
20. Krizhevsky A. Learning Multiple Layers of Features from Tiny Images. Toronto: University of Toronto; 2009:60.
21. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Wei J, Cha K. Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. *Med Phys*. 2016;43:6654–6666.
22. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems 27 (NIPS '14). 2014: 3320–3328.
23. Samala RK, Chan HP, Hadjiiski L, Helvie MA, Richter CD, Cha KH. Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. *IEEE Trans Med Imaging*. 2018:1.
24. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. *IEEE Computer Vision and Pattern Recognition (CVPR)*. 2015:1–9.
25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27–30, 2016:770–778. Las Vegas, NV.