

Article

Survival Prediction of Patients with Bladder Cancer after Cystectomy Based on Clinical, Radiomics, and Deep-Learning Descriptors

Di Sun ^{1,*} , Lubomir Hadjiiski ¹, John Gormley ¹, Heang-Ping Chan ¹, Elaine M. Caoili ¹, Richard H. Cohan ¹, Ajjai Alva ², Vikas Gulani ¹  and Chuan Zhou ¹

¹ Department of Radiology, University of Michigan, Ann Arbor, MI 48109, USA; lhadjisk@med.umich.edu (L.H.); jrgorm@umich.edu (J.G.); chanhp@med.umich.edu (H.-P.C.); caoili@med.umich.edu (E.M.C.); rcohan@med.umich.edu (R.H.C.); vikasgulani@med.umich.edu (V.G.); chuan@med.umich.edu (C.Z.)

² Department of Internal Medicine-Hematology/Oncology, University of Michigan, Ann Arbor, MI 48109, USA; ajjai@med.umich.edu

* Correspondence: disun@umich.edu

Simple Summary: Survival prediction of bladder cancer patients following cystectomy is essential for treatment planning. We propose a hybrid method that integrates clinical, radiomics, and deep-learning descriptors to improve survival prediction models. This approach demonstrates potential for more accurately predicting survival and prognosis in radical cystectomy treatment and in determining whether imaging adds additional predictive value over patients' clinical information.

Abstract: Accurate survival prediction for bladder cancer patients who have undergone radical cystectomy can improve their treatment management. However, the existing predictive models do not take advantage of both clinical and radiological imaging data. This study aimed to fill this gap by developing an approach that leverages the strengths of clinical (C), radiomics (R), and deep-learning (D) descriptors to improve survival prediction. The dataset comprised 163 patients, including clinical, histopathological information, and CT urography scans. The data were divided by patient into training, validation, and test sets. We analyzed the clinical data by a nomogram and the image data by radiomics and deep-learning models. The descriptors were input into a BPNN model for survival prediction. The AUCs on the test set were (C): 0.82 ± 0.06 , (R): 0.73 ± 0.07 , (D): 0.71 ± 0.07 , (CR): 0.86 ± 0.05 , (CD): 0.86 ± 0.05 , and (CRD): 0.87 ± 0.05 . The predictions based on D and CRD descriptors showed a significant difference ($p = 0.007$). For Kaplan–Meier survival analysis, the deceased and alive groups were stratified successfully by C ($p < 0.001$) and CRD ($p < 0.001$), with CRD predicting the alive group more accurately. The results highlight the potential of combining C, R, and D descriptors to accurately predict the survival of bladder cancer patients after cystectomy.

Keywords: bladder cancer; survival prediction; radical cystectomy; nomogram; radiomics; deep learning



Citation: Sun, D.; Hadjiiski, L.; Gormley, J.; Chan, H.-P.; Caoili, E.M.; Cohan, R.H.; Alva, A.; Gulani, V.; Zhou, C. Survival Prediction of Patients with Bladder Cancer after Cystectomy Based on Clinical, Radiomics, and Deep-Learning Descriptors. *Cancers* **2023**, *15*, 4372. <https://doi.org/10.3390/cancers15174372>

Academic Editors: Ke-Hung Tsui and Horng-Heng Juang

Received: 20 July 2023

Revised: 25 August 2023

Accepted: 28 August 2023

Published: 1 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bladder cancer is the tenth most common cancer worldwide [1], and the fourth most common cancer in men [2]. For muscle-invasive and recurrent non-invasive bladder cancer, radical cystectomy is a crucial treatment [3,4]. The procedure involves the complete removal of the bladder. In men, it also typically includes the removal of the prostate and seminal vesicles; in women, it often includes the removal of the uterus, ovaries, fallopian tubes, and part of the vagina [5].

The five-year survival rate serves as a valuable indicator of treatment effectiveness, varying across different types of cancer. Testicular cancer (97%), melanoma of the skin (92.3%), and prostate cancer (88%) have the highest estimated five-year survival [2]. Conversely, lung and bronchial cancer exhibit the lowest survival rate at only 22% [2]. In the

United States, the five-year relative survival rate for bladder cancer stands at 77% [2]. Specifically, for patients with bladder cancer who underwent radical cystectomy, the reported five-year survival rates range from 54.5% to 68% [6].

Survival prediction in bladder cancer can be performed based on data sources such as clinicopathological information [7], histological slides [8,9], gene expression, or molecular markers [10]. The analysis methods include machine-learning models like support vector machines, logistic regression, and random forest [11,12], as well as nomogram [13] and risk stratification [14]. Nomogram models, often developed based on large-size cohorts, can offer high-accuracy prediction capabilities.

CT urography (CTU) plays a significant role in bladder cancer diagnosis by providing 3D abdominal images that allow for the quantitative volumetric analysis of the bladder and lesion characteristics. The recent advances in deep-learning models offer new opportunities for extracting more comprehensive image features and the potential for improving the predictive models.

In this study, we combine clinical, histopathological features with CTU image features to harness the advantages of nomogram, radiomics, and deep-learning models. By introducing a hybrid approach combining clinical and imaging analytics, we aim to enhance the prediction of five-year survival for patients with bladder cancer after radical cystectomy, allowing better treatment planning and improving prognostic outcomes. The approach also has implications for the assessment of the added value of imaging in patient management and for how the added predictive value could be used to select the utilization of imaging.

2. Materials and Methods

The model development process is depicted in Figure 1. First, we collected a comprehensive dataset comprising clinical, histopathological information and CTU scans. One pair of pre- and post-treatment CTU scans was obtained for each patient. We then processed the collected data and developed models incorporating nomogram, radiomics, and deep-learning descriptors. The models were built with a training set and optimized with a validation set. After the models were developed, they were deployed to an independent test set to evaluate their survival prediction ability on unseen cases.

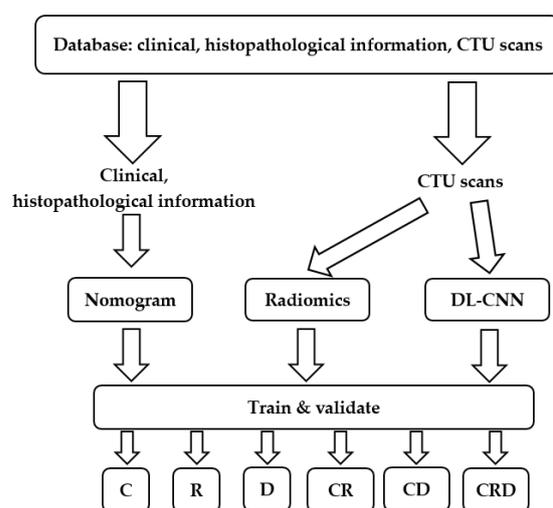


Figure 1. Predictive model development process of this study. We collected clinical, histopathological information and CTU scans. We built and validated survival prediction models based on individual or combined descriptors. C = clinical descriptors; R = radiomics descriptors; D = deep-learning descriptor; CR = clinical + radiomics descriptors; CD = clinical + deep-learning descriptors; CRD = clinical + radiomics + deep-learning descriptors. DL-CNN = Deep learning–convolutional neural network.

2.1. Patient Cohorts

With Institutional Review Board approval, histopathological information and CTU scans were collected for patients who had been diagnosed with bladder cancer. Patients were selected if they met all of the following criteria: (1) patients who underwent neoadjuvant chemotherapy and had at least one CTU scan before chemotherapy and one after chemotherapy; (2) patients who underwent radical cystectomy; (3) patients for whom follow-up information after surgery was available to determine their survival status.

We identified a total of 163 patients who satisfied the above criteria out of 337 patients. The study-population flow diagram is shown in Figure 2. Among the 163 patients, 79 patients were alive at the 5-year mark after receiving radical cystectomy, while 84 patients died before reaching the 5th year after cystectomy treatment. We split the data into three sets: 56% (92/163) training (55 alive; 37 deceased); 4% (7/163) validation (4 alive; 3 deceased); and 40% (64/163) test (20 alive; 44 deceased). We used a serial approach to assign the cases to the training, validation, and test datasets. We aimed at keeping a relatively large portion of the patients in the training and test sets. The bladder cancer cases were collected chronologically. The exam dates of the cases in the training and validation sets ranged from 2006 to 2012, and the cases in the test set ranged from 2015 to 2020. The serial approach simulates to some extent the “real world” clinical situation, where the model is built based on previous cases; after all weights and parameters are fixed, the model is applied to the new incoming cases.

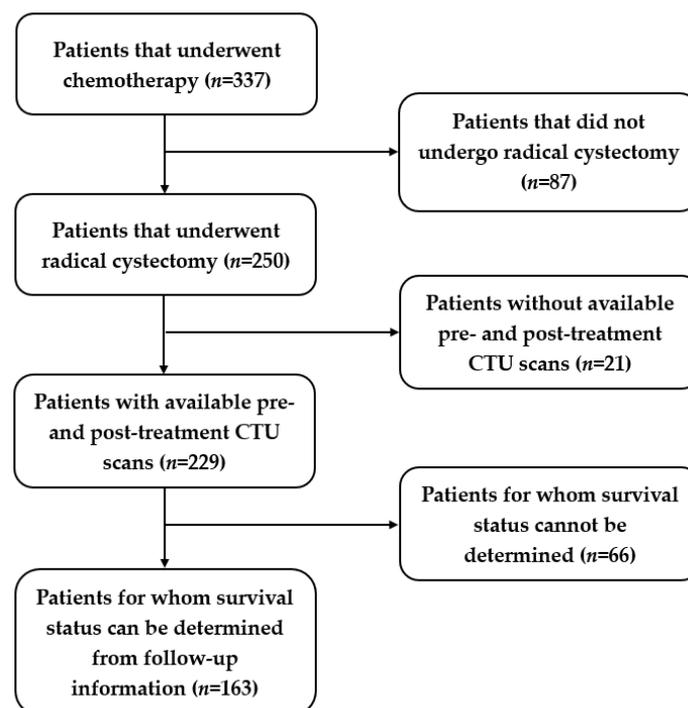


Figure 2. Study-population diagram. Out of 337 patients, 163 patients were identified for this study.

The collected clinical and histopathological information included five indices: post-surgery pathologic stage, lymphovascular invasion (LVI), pathologic node stage, whether patients underwent neoadjuvant chemotherapy, and whether patients underwent adjuvant radiotherapy. Chemotherapy plays a significant role in bladder cancer treatment influencing the survival of patients. The pre- and post-treatment CTU scan pairs were analyzed for changes in image features related to treatment response that are useful for survival prediction.

2.2. Nomogram

The histopathological information was analyzed using the nomogram developed by Shariat using a cohort of 731 patients [15]. The nomogram achieved a prediction accuracy of 0.791 in the study by Shariat [15], which was superior to that of the American Joint Committee on Cancer Staging [16] at 0.663 with $p = 0.001$.

The five clinical and histopathological indices described above were used as the input to the nomogram. For every patient, each of the five inputs was mapped to a “points” axis to determine how many points were contributed by each input. The sum of the points from the five inputs resulted in the estimation of the “total points” for the patient, which was then mapped to the n -year survival probability ($n = 5$ in the current study).

2.3. CTU Scans Processing

CTU scans for bladder cancer include a series of abdominal images, allowing for a comprehensive 3D inspection of the bladder. Figure 3a shows an example of a rendered 3D CTU scan. To extract radiomics and deep features from the CTU scans, the bladder cancer lesions were annotated by an experienced radiologist (R.H.C.) who has over 30 years of experience reading abdominal CT. The annotation process entailed several steps: (1) marking a volume of interest (VOI) for the lesion, (2) measuring the longest and perpendicular diameters of the tumor, (3) identifying the lesion location within the bladder, ureter, or other abdominal locations, and (4) evaluating the lesion type, edge characteristics, and likelihood of abnormality. While a patient could have multiple lesions, for this study, we limited the number of lesions to one per patient and selected the dominant lesion for feature extraction.

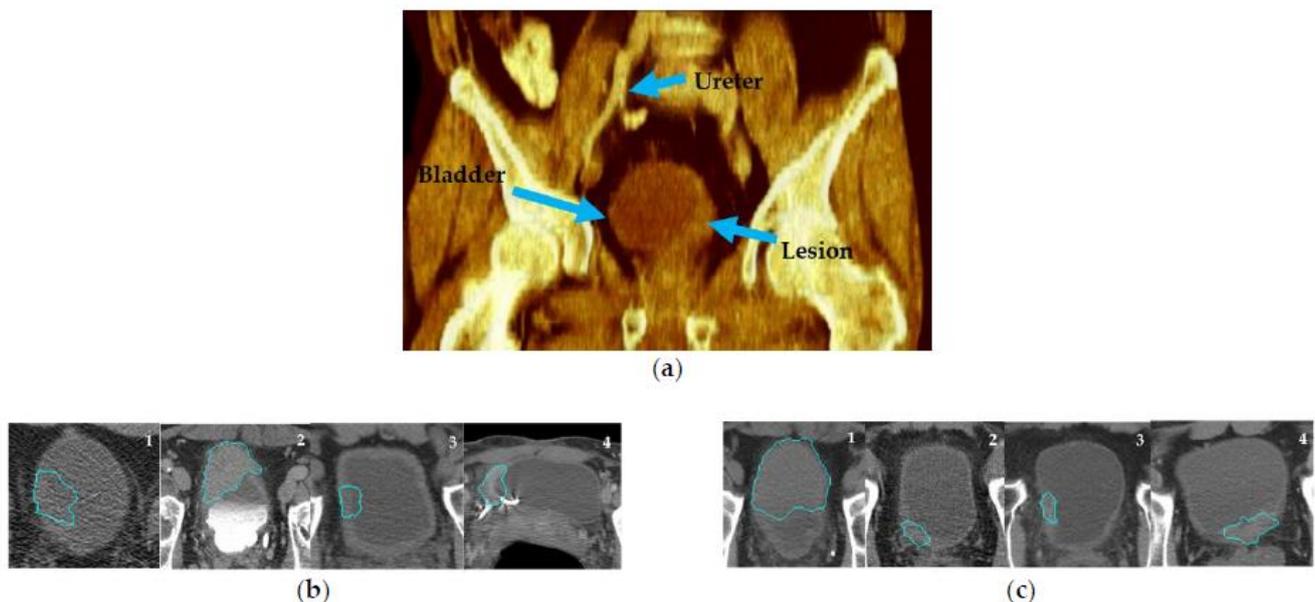


Figure 3. CTU scan examples. (a) A 3D volume rendering CTU scan showing the ureter, bladder, and bladder lesion. (b) CTU scans that are potentially challenging for lesion segmentation. (b)-1: without contrast; (b)-2: delayed-contrast phase; (b)-3: bladder wall thickening; (b)-4: image artifact. (c) Examples of CTU scans in early contrast phase and without artifacts. The blue outlines are the AI-CALS segmentation outlines.

Lesion segmentation was obtained by our in-house developed semi-automatic algorithm AI-CALS [17] utilizing the marked VOIs. The segmentation provided 3D contours outlining the lesion, from which we performed feature extraction. In a previous study, we showed that the AI-CALS algorithm achieved an average volume error of $4.9 \pm 38.3\%$ when compared to the manual outlines drawn by an experienced radiologist [17]. Segmenting lesions in CTU scans can be challenging due to various factors, such as patient position-

ing, image artifacts, and excreted contrast in the bladder. Ideally, since bladder cancer is expected to enhance in the early contrast phase, these images can help to distinguish neoplasm from the rest of the bladder tissue. However, some patients underwent CTU with different imaging protocols from referral sites including only non-contrast images (Figure 3b-1) for various reasons such as renal failure or patient refusal of contrast. Other protocols consisted of only delayed contrast phase imaging (Figure 3b-2) which could decrease the conspicuity of the lesion from surrounding structures. Figure 3b-3 shows a case with bladder-wall thickening after treatment, in which it is difficult to ascertain whether the lesion still existed. Additionally, artifacts in CTU scans can also complicate segmentation (Figure 3b-4). Despite these challenges, the performance of AI-CALS was satisfactory in most cases. Figure 3c illustrates CTU examples with early contrast phase and minimal artifacts, for which the segmentation was in good agreement with the manual reference.

2.4. Radiomics Model

Radiomics features derived from medical images have the potential to reveal intricate tumoral patterns and characteristics that may not be perceivable by the naked eye. The radiomics features (RF) developed in our group [18–20] demonstrated a good performance across various tasks, including lung nodule classification [21], mammographic mass characterization [22], and bladder cancer treatment response assessment [23].

A set of 91 RF features was extracted, including morphological, texture, and intensity-based features. Thus far, these features are primarily limited to 2D analysis. In this study, we extracted RF features from the central slice of the lesion, including the RF features from pre-treatment scans, referred to as f_{pre} , and the RF features from post-treatment scans, referred to as f_{post} . To capture the changes in features between pre- and post-treatment scans, the difference features f_{diff} were calculated by:

$$f_{diff} = \frac{f_{pre} - f_{post}}{f_{pre}}, \quad (1)$$

We therefore obtained a total of 273 radiomics features. Feature selection can eliminate the redundant variables, reduce the time and space requirements for data processing, and reduce the risk of the “curse of dimensionality” when the training data size is limited [24–26]. We employed a mutual information (MI) method for feature selection.

MI is capable of measuring the arbitrary dependencies between random variables and has been widely used for feature selection in machine learning [27]. One of the advantages of MI is its ability to assess the information content of features in nonlinear relations. Equation (3) shows that MI is formulated as the difference in the entropies between the variable itself (Equation (2)) and the variable conditioned on the target value. The MI score ranges from 0 to ∞ , where zero signifies complete independence between the feature and the target, while higher scores indicate greater relevance such that the corresponding feature may be useful for the learning task.

For a feature vector f , the entropy is:

$$H(F) = - \int P(f) \log P(f) df \quad (2)$$

Given the condition of C , the MI between variables c and f is:

$$I(F;C) = H(F) - H(F|C) \quad (3)$$

In order to select the useful set of features, a threshold was imposed on the calculated MI. Pearson correlation was applied to estimate the feature correlation.

2.5. Deep-Learning Model

To extract deep features from the paired pre- and post-treatment scans, we employed a hybrid-ROI strategy [28]. This approach, illustrated in Figure 4a, involved creating a

hybrid-ROI that is composed of one ROI of size 32×16 pixels from the pre-treatment scan on the left side and one ROI of the same size from the post-treatment scan on the right side. The resulting hybrid ROI has a size of 32×32 pixels. We utilized a sliding-window technique to extract the ROIs from the VOI. To ensure a balanced dataset and prevent dominance or bias caused by CTU scan pairs with large lesions, we imposed a threshold to limit the number of hybrid ROIs from one scan pair. All hybrid ROIs from the same case were assigned the same case label, survived over 5 years = 1, otherwise = 0. A subset of hybrid ROIs from the training set is shown in Figure 4b.

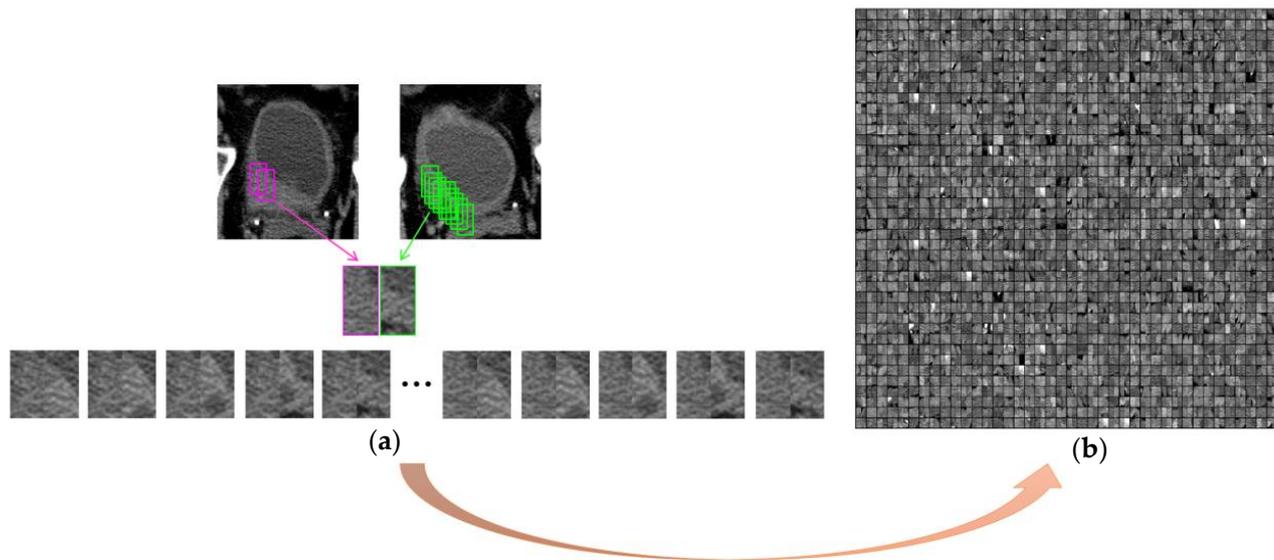


Figure 4. Formation of hybrid ROIs. (a) Use a sliding-window technique to extract ROIs from the VOI. A hybrid ROI is formed by combining one ROI from the pre-treatment scan on the left side (pink frame) and one ROI from the post-treatment scan on the right side (green frame). Different combinations result in a number of hybrid ROIs from one pre- and post-treatment scan pair. (b) A subset of hybrid ROIs of the training set shown in a matrix. Each hybrid ROI (32×32 pixels) was a separate sample during deep-learning model training.

Deep features were extracted from the hybrid ROIs by a Convolutional Neural Network (CNN) [29]. The structure of the CNN is depicted in Figure 5, comprising two convolutional layers, C1 and C2, each of which is accompanied by a local response normalization layer and a max-pooling layer; two locally connected layers, L3 and L4; and a fully connected layer, FC10. The C1, C2, and the max-pooling layers could be considered as a deep-feature extractor. The last three fully connected layers synthesized the deep features into a likelihood score for each input image (hybrid ROI). The likelihood scores of all the hybrid ROIs from one patient were combined into a single likelihood score that reflected the survival likelihood of that patient.

2.6. Classification

The features from the three types of descriptors of each case, i.e., the selected radiomics features by the MI, the survival likelihood score from the deep-learning model, and the “points” derived from each of the five indices of the nomogram, were fed into a Back-Propagation Neural Network (BPNN) [30] to classify the likelihood of survival of each patient. The performance of the descriptors when they were used individually was compared to the utilization of the descriptors in combination. The structure of the BPNN is shown in Figure 6, where the parameters of the BPNN were selected as guided by the validation set.

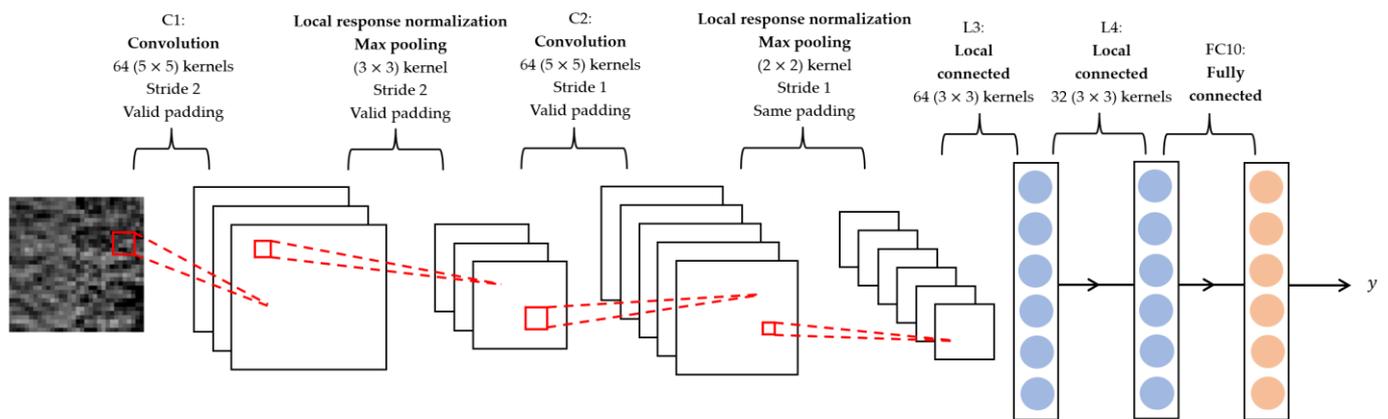


Figure 5. Architecture of CNN. The input image has a size of 32×32 pixels. The CNN was composed of two convolution layers (C1 and C2), each of them was followed by a local response normalization layer and a max-pooling layer, two locally connected layers (L3 and L4), and one fully connected layer (FC10). The kernel, stride, and padding setting of each layer are illustrated. The output is a likelihood score.

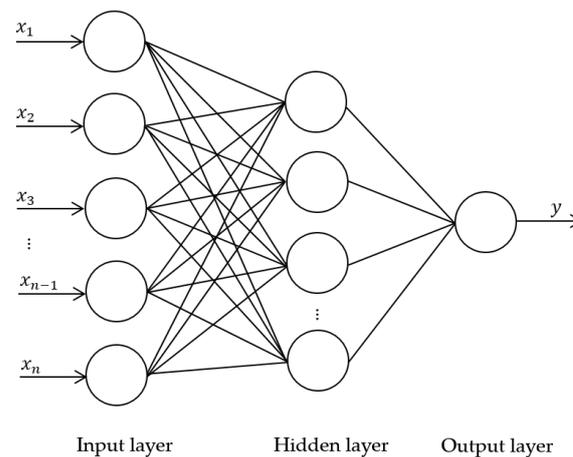


Figure 6. Structure of BPNN. The input x_i ($i = 1, 2, 3, \dots, n$) are the descriptors. The hidden layer contains 13 nodes. The output y is the likelihood score assessing the survival of each patient.

2.7. Statistical Analysis

The classification performance of the descriptors was evaluated using the area under the receiver operating characteristics (ROC) curve (AUC). We compared three combinations of descriptors for survival prediction (C vs. CRD, R vs. CRD, and D vs. CRD); thus, the critical α value for statistical significance was adjusted to be $\alpha = 0.017$ ($\alpha = 0.05/3$) according to the Bonferroni correction for multiple hypothesis testing [31,32].

To further analyze the model performance on predicting survival outcomes, we conducted a Kaplan–Meier analysis based on a log-rank test [33] to generate the survival curves.

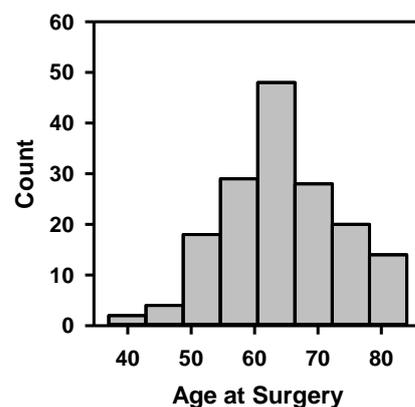
3. Results

3.1. Cohort Statistics

The demographic information and cancer stage of the 163 patients are shown in Table 1, including gender, race, tobacco use, clinical stage of cancer, and pathological stage of cancer after radical cystectomy. The distribution of patients' age at surgery is depicted in Figure 7. These statistics show that the dataset covered a wide range of patient demographics.

Table 1. Patient characteristics ($n = 163$).

Attributes		#
Gender	Male	131
	Female	32
Race	White	142
	Black/African Am.	13
	Am. Indian/Native	1
	Asian	2
	Unknown	5
Tobacco use	Current	40
	Former	88
	Never	34
	Unknown	1
Post-surgery stage	pT0	35
	pTa/pTi/pTis	15
	pT1	16
	pT2	36
	pT3	45
	pT4	16
Lymphovascular invasion (LVI)	Yes	61
	No	102
Pathologic node stage	N0	112
	N1	24
	N2	23
	N3	4
Neoadjuvant chemotherapy	Yes	163
	No	0
Adjuvant radiotherapy	Yes	0
	No	163

**Figure 7.** Histogram of patients' age at surgery.

3.2. Radiomics Features

The MI-selected radiomics features were evaluated by Pearson correlation on the training set. First, we compared the feature vector with the target vector. In our analysis, we aimed for a strong correlation between feature and target, as this would allow us to predict the target from the feature. Then we compared the feature vectors among themselves. It is preferable to have the selected features less correlated to make the model more robust. However, features with some correlation as a group may also be useful in the model design.

Out of the 12 selected features, we plotted a heatmap showing the correlation values between each pair of features (Figure 8). The majority of the correlation values fell within

the range of ± 0.3 , indicating weak correlations. Weak correlations implied that the selected features were providing mostly distinct and independent information, which can contribute to the predictive power of the model with a low redundancy. On the other hand, the features with a certain level of correlation may contribute to the predictive power when combined with other features.

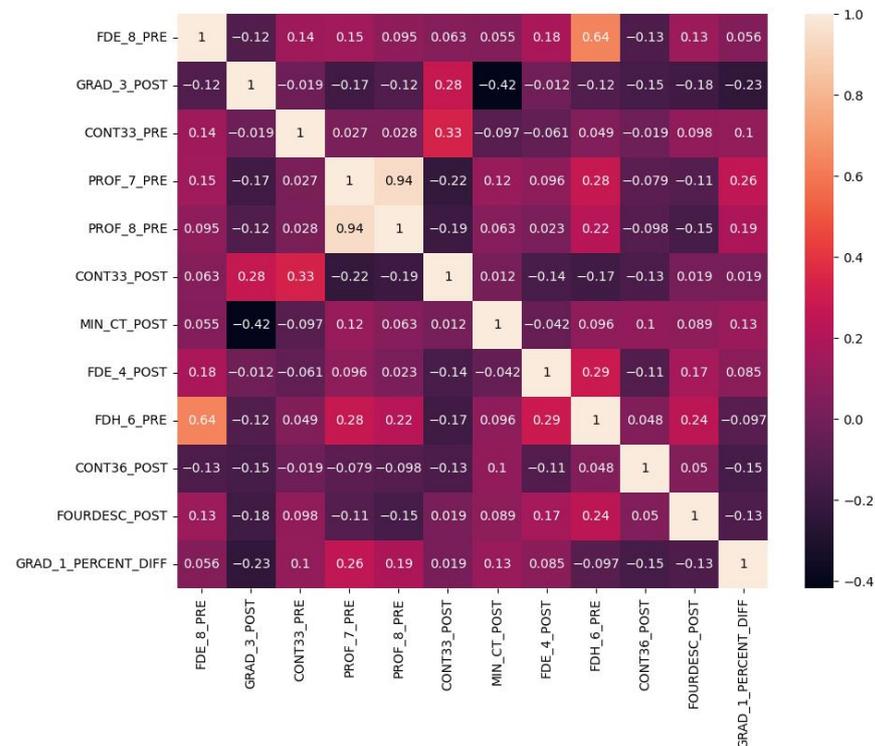


Figure 8. Heatmap of Pearson correlation between any pair of the MI-selected radiomics features. Most of the correlation values were within ± 0.3 which indicates a weak correlation.

3.3. Survival Prediction

After the training of the models with the training set, we selected the best-performing clinical, radiomics, deep-learning, and combined models using the validation set. For the combined models, BPNN training required about 40 iterations. We deployed the selected models on the test set to evaluate their ability to predict patient survival. The ROC curves of the individual and combined descriptors are plotted in Figure 9. The clinical (C), radiomics (R), and deep-learning (D) descriptors achieved AUCs of 0.82 ± 0.06 , 0.73 ± 0.07 , and 0.71 ± 0.07 , respectively. By combining the R and D descriptors with the clinical C descriptors, the AUCs of the models were improved to 0.86 ± 0.05 for CR, 0.86 ± 0.05 for CD, and 0.87 ± 0.05 for CRD. The ROC curves of C and CRD are plotted in one figure (Figure 9c) to show their differences in classification performance. The classifications based on D and CRD descriptors achieved a significant level of difference ($p = 0.007$) (Table 2). Other pairs of comparisons did not reach statistical significance but showed strong trends towards improvement when used in combination.

To further investigate the impact of incorporating radiomics and deep-learning descriptors into clinical descriptors on survival prediction, we conducted Kaplan–Meier analysis on C and CRD descriptors. To categorize the patients in the test set into two groups (deceased and alive groups), we determined the cutoff values of the scores generated by BPNN for C and CRD descriptors by selecting the point of least misclassification for the C and CRD models separately based on the training and validation sets. Figure 10 displays the Kaplan–Meier curves of the deceased and alive groups predicted by C ($p < 0.001$) and CRD descriptors ($p < 0.001$). For the alive group, the CRD-generated curve was superior to the C-generated curve.

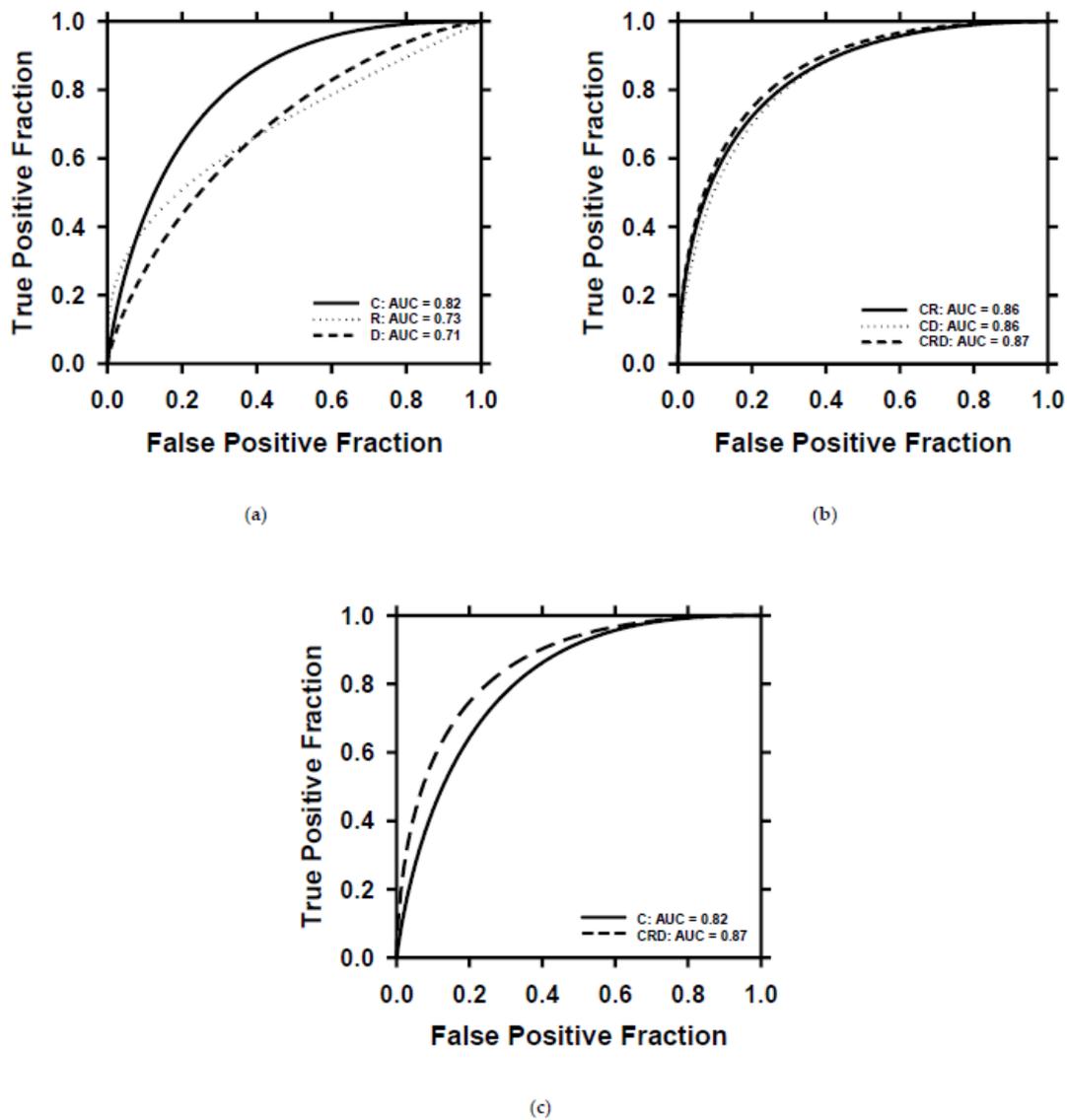


Figure 9. ROC curves of survival prediction on the test set by using different descriptors. (a) ROC curves and AUC values for the individual descriptor. (b) ROC curves and AUC values for combined descriptors. (c) Direct comparison of the ROC curves of C and CRD descriptors. C = clinical descriptors; R = radiomics descriptors; D = deep-learning descriptor; CR = clinical + radiomics descriptors; CD = clinical + deep-learning descriptors; CRD = clinical + radiomics + deep-learning descriptors.

Table 2. The *p*-values of survival prediction based on individual and combined descriptors. The critical α value for statistical significance was adjusted to be $\alpha = 0.017$ ($\alpha = 0.05/3$) according to the Bonferroni correction for multiple hypothesis testing.

Comparison	<i>p</i> -Value (Adjusted $\alpha = 0.017$)
C vs. CRD	0.153
R vs. CRD	0.056
D vs. CRD	0.007 *

* Achieved a significant difference.

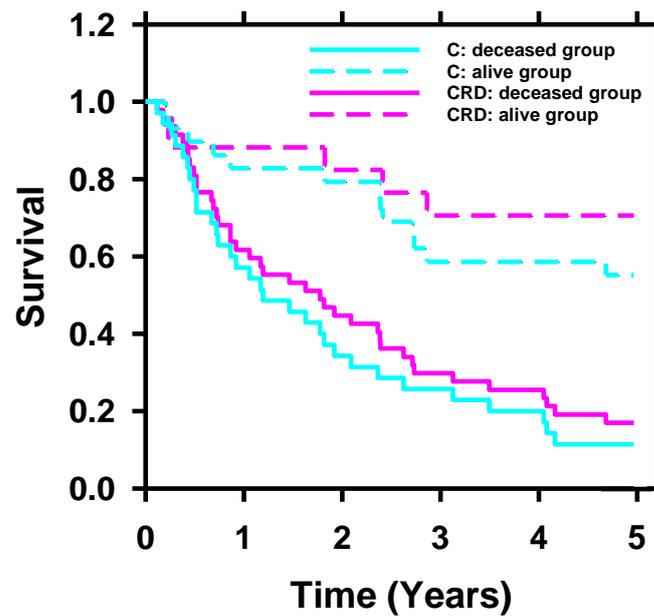


Figure 10. Kaplan–Meier survival probability analyzed with C and CRD descriptors. Stratification of the two groups (deceased and alive) achieved statistical significance using C ($p < 0.001$) and CRD ($p < 0.001$) descriptors. For the alive group, the survival curve of CRD was superior to that of C.

4. Discussion

All three descriptors—C, R, and D—showed strength in classifying the test dataset. However, the clinical descriptors demonstrated a higher performance compared to the deep-learning and radiomics descriptors. One of the main factors contributing to this result is that the nomogram model utilized in the clinical descriptors was developed using a larger training set, while the deep-learning and radiomics models used here were trained on a smaller dataset. Deep-learning models are known to achieve a better generalization and robustness when they are trained with larger and more representative datasets [34].

With the integration of radiomics and deep-learning descriptors alongside the clinical descriptors, the CRD descriptors can stratify the deceased and alive groups more effectively. The lack of statistical significance in the difference between the AUCs of the C and CRD descriptors, and the modest improvement in the stratification ability in the survival curves between the deceased and alive groups by CRD compared to that of the C descriptors, may be attributed to the small size of the test set, which could limit the statistical power to detect the small differences, and also the strong performance of the clinical descriptors alone. While this is a preliminary study and requires significant further investigation, the result has implications for the utilization of this approach in the future, as it could be used to assess whether an imaging test has added value over clinical data alone. If an imaging test is expected to add predictive value for the clinical task at hand, it will strengthen the utility of the imaging test for patient care.

The results also point to the importance of clinical information in the utilization of imaging, as this information can lead to the improved performance of imaging in answering the clinical question. This is well accepted in clinical radiology, but clinical information is not typically utilized in radiomics or machine-learning computational models for imaging tasks. This study demonstrated that the clinical data may contain important complementary information to improve the performance of these models.

Our study has some limitations. The dataset was relatively small and collected from a single site. We need to collect multisite cases and larger datasets to build more reliable and generalizable models. We also included non-contrast CT images which could affect the performance of the R and D models. On the other hand, the presence of non-contrast CT

images represents the real clinical situation and the models may have been trained to be more robust with the heterogeneous image features.

Despite the limitations, this study demonstrated that CTU scans potentially can provide useful information to improve the survival prediction of patients with bladder cancer after radical cystectomy. Radiomics and deep-learning models can extract effective image features from CTU scans. Combined with a nomogram, the prediction ability of the hybrid model can outperform the models using the individual types of descriptors.

5. Conclusions

While larger datasets are needed, this study demonstrates that combining radiomics and deep-learning descriptors with clinical descriptors holds promise for improving the prediction of the 5-year survival of bladder cancer patients after radical cystectomy. The accurate assessment of 5-year survival offers potential benefits with patient counseling and postoperative surveillance strategies.

Author Contributions: Conceptualization, L.H., V.G., H.-P.C., R.H.C., E.M.C. and D.S.; methodology, L.H., H.-P.C. and V.G.; software, L.H.; validation, D.S., L.H., J.G. and H.-P.C.; formal analysis, D.S. and L.H.; investigation, J.G., L.H., R.H.C. and E.M.C.; data curation, D.S., J.G. and L.H.; writing—original draft preparation, D.S.; writing—review and editing, D.S., L.H., H.-P.C., V.G., R.H.C., E.M.C., A.A., C.Z. and J.G.; visualization, D.S.; supervision, L.H.; project administration, L.H. and V.G.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institutes of Health, grant number U01-CA232931.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Michigan (HUM00044051, 15 November 2010).

Informed Consent Statement: Patient consent was waived due to the designation of this study as a Health Insurance Portability and Accountability Act (HIPAA)-compliant retrospective cohort study.

Data Availability Statement: Data available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Deep learning-convolutional neural network (DL-CNN); Clinical descriptors (C); radiomics descriptors (R); deep-learning descriptor (D); clinical + radiomics descriptors (CR); clinical + deep-learning descriptors (CD); clinical + radiomics + deep-learning descriptors (CRD); CT Urography (CTU); lymphovascular invasion (LVI); volumes of interest (VOI); radiomics features (RF); mutual information (MI); Convolutional Neural Network (CNN); Back-Propagation Neural Network (BPNN); receiver operating characteristics (ROC); area under ROC curve (AUC).

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)] [[PubMed](#)]
2. American Cancer Society. *Cancer Facts & Figures 2023*; American Cancer Society: Atlanta, GA, USA, 2023.
3. Gschwend, J.E.; Fair, W.R.; Vieweg, J. Radical cystectomy for invasive bladder cancer: Contemporary results and remaining controversies. *Eur. Urol.* **2000**, *38*, 121–130. [[CrossRef](#)] [[PubMed](#)]
4. Chang, S.S.; Cookson, M.S. Non-muscle-invasive bladder cancer: The role of radical cystectomy. *Urology* **2005**, *66*, 917–922. [[CrossRef](#)] [[PubMed](#)]
5. Modh, R.A.; Mulhall, J.P.; Gilbert, S.M. Sexual dysfunction after cystectomy and urinary diversion. *Nat. Rev. Urol.* **2014**, *11*, 445–453. [[CrossRef](#)]
6. Stein, J.P.; Lieskovsky, G.; Cote, R.; Groshen, S.; Feng, A.-C.; Boyd, S.; Skinner, E.; Bochner, B.; Thangathurai, D.; Mikhail, M. Radical cystectomy in the treatment of invasive bladder cancer: Long-term results in 1,054 patients. *J. Clin. Oncol.* **2001**, *19*, 666–675. [[CrossRef](#)]
7. Eisenberg, M.S.; Boorjian, S.A.; Cheville, J.C.; Thompson, R.H.; Thapa, P.; Kaushik, D.; Frank, I. The SPARC score: A multifactorial outcome prediction model for patients undergoing radical cystectomy for bladder cancer. *J. Urol.* **2013**, *190*, 2005–2010. [[CrossRef](#)]

8. Gavriel, C.G.; Dimitriou, N.; Brieu, N.; Nearchou, I.P.; Arandjelović, O.; Schmidt, G.; Harrison, D.J.; Caie, P.D. Assessment of immunological features in muscle-invasive bladder cancer prognosis using ensemble learning. *Cancers* **2021**, *13*, 1624. [[CrossRef](#)]
9. Zheng, Q.; Yang, R.; Ni, X.; Yang, S.; Xiong, L.; Yan, D.; Xia, L.; Yuan, J.; Wang, J.; Jiao, P. Accurate Diagnosis and Survival Prediction of Bladder Cancer Using Deep Learning on Histological Slides. *Cancers* **2022**, *14*, 5807. [[CrossRef](#)]
10. Seiler, R.; Ashab, H.A.D.; Erho, N.; van Rhijn, B.W.; Winters, B.; Douglas, J.; Van Kessel, K.E.; van de Putte, E.E.F.; Sommerlad, M.; Wang, N.Q. Impact of molecular subtypes in muscle-invasive bladder cancer on predicting response and survival after neoadjuvant chemotherapy. *Eur. Urol.* **2017**, *72*, 544–554. [[CrossRef](#)]
11. Bhambhani, H.P.; Zamora, A.; Shkolyar, E.; Prado, K.; Greenberg, D.R.; Kasman, A.M.; Liao, J.; Shah, S.; Srinivas, S.; Skinner, E.C. Development of robust artificial neural networks for prediction of 5-year survival in bladder cancer. In *Urologic Oncology: Seminars and Original Investigations*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 193.e7–193.e12.
12. Gavriel, C.G.; Dimitriou, N.; Brieu, N.; Nearchou, I.P.; Arandjelović, O.; Schmidt, G.; Harrison, D.J.; Caie, P.D. Identification of immunological features enables survival prediction of muscle-invasive bladder cancer patients using machine learning. *BioRxiv* **2020**. [[CrossRef](#)]
13. Riestler, M.; Taylor, J.M.; Feifer, A.; Koppie, T.; Rosenberg, J.E.; Downey, R.J.; Bochner, B.H.; Michor, F. Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer. *Clin. Cancer Res.* **2012**, *18*, 1323–1333. [[CrossRef](#)]
14. Shariat, S.F.; Bolenz, C.; Godoy, G.; Fradet, Y.; Ashfaq, R.; Karakiewicz, P.I.; Isbarn, H.; Jeldres, C.; Rigaud, J.; Sagalowsky, A.I. Predictive value of combined immunohistochemical markers in patients with pT1 urothelial carcinoma at radical cystectomy. *J. Urol.* **2009**, *182*, 78–84. [[CrossRef](#)] [[PubMed](#)]
15. Shariat, S.F.; Karakiewicz, P.I.; Palapattu, G.S.; Amiel, G.E.; Lotan, Y.; Rogers, C.G.; Vazina, A.; Bastian, P.J.; Gupta, A.; Sagalowsky, A.I. Nomograms provide improved accuracy for predicting survival after radical cystectomy. *Clin. Cancer Res.* **2006**, *12*, 6663–6676. [[CrossRef](#)] [[PubMed](#)]
16. Edge, S.B.; Compton, C.C. The American Joint Committee on Cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann. Surg. Oncol.* **2010**, *17*, 1471–1474. [[CrossRef](#)]
17. 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 multidetector row CT urography. *Acad. Radiol.* **2013**, *20*, 148–155. [[CrossRef](#)]
18. Way, T.W.; Hadjiiski, L.M.; Sahiner, B.; Chan, H.P.; Cascade, P.N.; Kazerooni, E.A.; Bogot, N.; Zhou, C. Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours. *Med. Phys.* **2006**, *33*, 2323–2337. [[CrossRef](#)]
19. Way, T.W.; Sahiner, B.; Chan, H.P.; Hadjiiski, L.; Cascade, P.N.; Chughtai, A.; Bogot, N.; Kazerooni, E. Computer-aided diagnosis of pulmonary nodules on CT scans: Improvement of classification performance with nodule surface features. *Med. Phys.* **2009**, *36*, 3086–3098. [[CrossRef](#)] [[PubMed](#)]
20. Sahiner, B.; Chan, H.P.; Petrick, N.; Helvie, M.A.; Goodsitt, M.M. Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis. *Med. Phys.* **1998**, *25*, 516–526. [[CrossRef](#)]
21. Kirby, J.S.; Armato, S.G.; Drukker, K.; Li, F.; Hadjiiski, L.; Tourassi, G.D.; Clarke, L.P.; Engelmann, R.M.; Giger, M.L.; Redmond, G.; et al. LUNGx Challenge for computerized lung nodule classification. *J. Med. Imaging* **2016**, *3*, 044506.
22. Sahiner, B.; Chan, H.P.; Petrick, N.; Helvie, M.A.; Hadjiiski, L.M. Improvement of mammographic mass characterization using spiculation measures and morphological features. *Med. Phys.* **2001**, *28*, 1455–1465. [[CrossRef](#)]
23. Sun, D.; Hadjiiski, L.; Alva, A.; Zakharia, Y.; Joshi, M.; Chan, H.-P.; Garje, R.; Pomerantz, L.; Elhag, D.; Cohan, R.H. Computerized decision support for bladder cancer treatment response assessment in CT urography: Effect on diagnostic accuracy in multi-institution multi-specialty study. *Tomography* **2022**, *8*, 644–656. [[CrossRef](#)]
24. Chan, H.P.; Sahiner, B.; Wagner, R.F.; Petrick, N. Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers. *Med. Phys.* **1999**, *26*, 2654–2668. [[CrossRef](#)] [[PubMed](#)]
25. Sahiner, B.; Chan, H.P.; Petrick, N.; Wagner, R.F.; Hadjiiski, L. Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size. *Med. Phys.* **2000**, *27*, 1509–1522. [[CrossRef](#)] [[PubMed](#)]
26. Way, T.W.; Sahiner, B.; Hadjiiski, L.M.; Chan, H.P. Effect of finite sample size on feature selection and classification: A simulation study. *Med. Phys.* **2010**, *37*, 907–920. [[CrossRef](#)]
27. Battiti, R. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. Neural Netw.* **1994**, *5*, 537–550. [[CrossRef](#)] [[PubMed](#)]
28. Cha, K.H.; Hadjiiski, L.; Chan, H.-P.; Weizer, A.Z.; Alva, A.; Cohan, R.H.; Caoili, E.M.; Paramagul, C.; Samala, R.K. Bladder cancer treatment response assessment in CT using radiomics with deep-learning. *Sci. Rep.* **2017**, *7*, 8738. [[CrossRef](#)] [[PubMed](#)]
29. Wu, E.; Hadjiiski, L.M.; Samala, R.K.; Chan, H.-P.; Cha, K.H.; Richter, C.; Cohan, R.H.; Caoili, E.M.; Paramagul, C.; Alva, A. Deep learning approach for assessment of bladder cancer treatment response. *Tomography* **2019**, *5*, 201–208. [[CrossRef](#)]
30. Goh, A.T. Back-propagation neural networks for modeling complex systems. *Artif. Intell. Eng.* **1995**, *9*, 143–151. [[CrossRef](#)]
31. Chen, S.-Y.; Feng, Z.; Yi, X. A general introduction to adjustment for multiple comparisons. *J. Thorac. Dis.* **2017**, *9*, 1725. [[CrossRef](#)]
32. Bland, J.M.; Altman, D.G. Multiple significance tests: The Bonferroni method. *BMJ* **1995**, *310*, 170. [[CrossRef](#)]

33. Altman, D.G. *Practical Statistics for Medical Research*; CRC Press: Boca Raton, FL, USA, 1990.
34. Samala, R.K.; Chan, H.-P.; Hadjiiski, L.; Helvie, M.A.; Richter, C.D.; Cha, K.H. 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**, *38*, 686–696. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.