Artificial intelligence for predicting response to neoadjuvant chemotherapy for bladder cancer: A comprehensive systematic review and meta-analysis

Caio Vinícius Suartz[1], Lucas Motta Martinez[1], Maurício Dener Cordeiro[1], Hunter Ausley Flores[2], Sarah Kodama[3], Leonardo Cardili[1], José Maurício Mota[4], Fernando Morbeck Almeida Coelho[5], José de Bessa Junior[6], Cristina Pires Camargo[7], Jeremy Yuen-Chun Teoh[8], Shahrokh F. Shariat[9,10,11,12], Paul Toren[13], William Carlos Nahas[1], Leopoldo Alves Ribeiro-Filho[1]

[1]Division of Urology, Institute of Cancer of São Paulo, University of São Paulo, São Paulo, Brazil; [2]Department of Urology, University of Colorado, Aurora, CO, United States; [3]Virginia Commonwealth University School of Medicine, Richmond, VA, United States; [4]Division of Oncology, Institute of Cancer of São Paulo, University of São Paulo, São Paulo, Brazil; [5]Department of Radiology, University of São Paulo, São Paulo, Brazil; [6]State University of Feira de Santana, Feira de Santana, Bahia, Brazil; [7]Microsurgery and Plastic Surgery Laboratory, School of Medicine, University of São Paulo, São Paulo, Brazil; [8]Department of Surgery, S.H. Ho Urology Centre, Chinese University of Hong Kong, Hong Kong, China; [9]Department of Urology, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria; [10]Department of Urology, Weill Cornell Medical College, New York, NY, United States; [11]Department of Urology, University of Texas Southwestern, Dallas, TX, United States; [12]Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, Jordan; [13]CHU de Québec-Université Laval, Quebec City, QC, Canada

Cite as: Suartz CV, Martinez LM, Cordeiro MD, et al. Artificial intelligence for predicting response to neoadjuvant chemotherapy for bladder cancer: A comprehensive systematic review and meta-analysis. *Can Urol Assoc J* 2024 May 21; Epub ahead of print. http://dx.doi.org/10.5489/cuaj.8681

Published online May 21, 2024

Corresponding author: Dr. Caio Vinícius Suartz, Division of Urology, Institute of Cancer of São Paulo, University of São Paulo, São Paulo, Brazil; caio.v_suartz@hotmail.com

***

KEY MESSAGES

- The meta-analysis of studies focused on radiomics showed moderate accuracy in predicting neoadjuvant chemotherapy response in bladder cancer, with a sensitivity of 62% and specificity of 82%.
- Genetic data analysis in three studies showed promising results in predicting chemotherapy response but was limited by small sample sizes and methodology.
- A study using anatomopathological data identified key cellular characteristics for predicting treatment response, with accuracies ranging from 65–73%.
- The research indicates artificial intelligence as a potential tool to enhance diagnostic accuracy and reduce variability in physician assessments.

ABSTRACT

Introduction: Neoadjuvant cisplatin-based combination chemotherapy (NAC) followed by radical cystectomy is the standard of care for cisplatin-fit patients harboring muscle-invasive bladder cancer (MIBC). Prediction of response to NAC is essential for clinical decision-making regarding alternatives in case of non-response and bladder-sparing in case of

© 2024 Canadian Urological Association

complete response. This research aimed to assess the performance of machine learning in predicting therapeutic response following NAC treatment in patients with MIBC.
Methods: A systematic review adhering to the PRISMA guidelines was conducted until July 2023. The study integrated articles relating to artificial intelligence and NAC response in MIBC from various databases. The quality of articles was evaluated using the Quality Assessment Tool for Diagnostic Accuracy Studies 2 (QUADAS-2). A meta-analysis was subsequently performed on selected studies to determine the sensitivity and specificity of machine learning algorithms in predicting NAC response.
Results: Of 655 articles identified, 12 studies comprising 1523 patients were included, and four studies were eligible for meta-analysis. The sensitivity and specificity of the studies were 0.62 (95% confidence interval [CI] 0.50–0.72) and 0.82 (95% CI 0.72–0.89), respectively, with a heterogeneity score (I^2) of 38.5%. The machine learning algorithms used computed tomography, genetic, and anatomopathological data as input and exhibited promising potential for predicting NAC response.
Conclusions: Machine-learning algorithms, especially those using computed tomography, genetic, and pathologic data, demonstrate significant potential for predicting NAC response in MIBC. Standardization of methodologic data analysis and response criteria are needed as validation studies.

DRAFT

INTRODUCTION

Artificial intelligence (AI) is the ability of a machine to learn and display intelligence (1). AI has the potential to transform big data into clinical practice and can produce more accurate data, reduce inevitable errors (2), and make real-time predictions (3,4).
In the last decade, improvements in computational memory and processing speed have helped implement and refine artificial intelligence techniques for the oncological management of patients, especially utilizing machine learning (ML), which is a computer science field where algorithms learn and improve through data exposure. It involves using statistical methods to build models to identify patterns and make predictions from data. ML includes techniques like supervised learning (using labeled data) and unsupervised learning (finding Patterns in unlabeled data), enhancing our ability to handle large data volumes (5).

ML techniques have been applied across the bladder cancer management spectrum to improve the diagnostic accuracy of cystoscopy (6), computed tomography scanning for metastases, particularly micrometastases (i.e. lymph nodes) (7), and complete response to neoadjuvant chemotherapy (Figure 1).
Neoadjuvant cisplatin-based combination chemotherapy followed by radical cystectomy is the standard of care for cisplatin-fit muscle-invasive bladder cancer (MIBC) patients. The prediction of response to NAC is on the current clinical ward and could change the therapeutic paradigm by a large degree (8).

Patients who are unlikely to respond to NAC would receive alternative therapies and avoid unnecessary adverse events from an ineffective therapy. Those likely to experience

© 2024 Canadian Urological Association

complete responses would be given the choice of bladder-sparing strategies. Radiomic, genetic analysis, molecular markers, and histopathological characteristics have been used with algorithms to predict with variable accuracy the response to NAC in MIBC patients. In this study, we aimed to access the different AI-driven algorithms to predict NAC response in MIBC patients to establish an up-to-date catalog of studies to help advance future research in this space.

METHODS

Literature search

The study was conducted in strict compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) (9) statement on July 01, 2023, and it was registered in the PROSPERO international database of prospectively registered systematic reviews (CRD 42023463142).

A research question was established based on the Patient-Index test- Comparator-Outcome-Study design (PICOS) criteria, which is as follows (8): What is the diagnostic performance of machine learning in predicting therapeutic response to NAC in patients with MIBC?

The search strategy was: (Artificial intelligence OR deep learning OR machine learning) AND (neoplasm, urinary bladder OR urinary bladder neoplasm OR bladder tumors OR bladder tumor OR tumor, bladder OR tumors, bladder OR neoplasms, bladder OR bladder neoplasms OR bladder neoplasm OR neoplasm, bladder OR urinary bladder cancer OR cancer, urinary bladder OR malignant tumor of urinary bladder OR cancer of the bladder OR bladder cancer OR bladder cancers OR cancer, bladder OR cancer of bladder) AND (neoadjuvant therapy OR neoadjuvant chemotherapy OR treatment, neoadjuvant systemic OR systemic treatment, neoadjuvant OR chemotherapy treatment OR treatment, neoadjuvant).

We searched the following databases up to July 2023:

- Cochrane Central Register of Controlled Trials (CENTRAL) 2020, Issue 3, in the Cochrane Library;
- MEDLINE via Ovid (from 1946);
- Embase via Ovid (from 1974);
- LILACS (Latin American and Caribbean Health Science Information database, from 1982);
- Scopus, Elsevier's citation tool (from 2004);
- Web of Science/Web of Knowledge (Clarivate and Thomson Reuters) (from 1900);
- Education Resources Information Center (ERIC) (from 1966);

and the following trial registries:

- ISRCTN registry (http://www.isrctn.com);
- ClinicalTrials.gov (http://www.clinicaltrials.gov);
- Australian New Zealand Clinical Trials Registry (http://www.anzctr.org.au);
- World Health Organization International Clinical Trials Registry Platform (ICTRP) (apps.who.int/trialsearch/); and

© 2024 Canadian Urological Association

- EU Clinical Trials Register(http://www.clinicaltrialsregister.eu).

We also checked the bibliographies of the studies included for further references to relevant trials.

We included randomized clinical trials and cohort studies. Meeting abstracts, reviews, case reports, letters to the editor, and editorials were excluded.

Study screening and selection

Two independent authors screened all retrieved records. Discrepancies were resolved by discussion. Meeting abstracts, reviews, case reports, letters to the editor, and editorials were excluded. The full text of the screened papers was selected if found to be relevant to the present review.

Selection criteria

We included studies that met the following criteria : (1) patients (P) had muscle-invasive bladder cancer and underwent neoadjuvant chemo treatment; (2) index test (I) of pre-operative diagnostic test combined with machine learning to determine the outcome (O) of interest (sensitivity, specificity and accuracy for predicting neoadjuvant response) (3) pathological results based on radical cystectomy was used as the reference standard or comparator (C); (4) sufficient data were provided regarding sensitivity and specificity; and (5) publication type or study design (S) was original article. The studies were excluded if the model did not use artificial intelligence, models that studied artificial intelligence just for prognostic evaluation, and studies without NAC. Authors of the studies were contacted when needed to request information regarding 2 × 2 tables or other relevant information required for meta-analytical purposes.

Quality assessment and risk of bias

Studies were assessed for robustness of methodology using the scoring system of the APPRAISE-AI tool, which is structured to evaluate the quality of AI studies across six key domains: clinical relevance, data quality, methodological conduct, robustness of results, reporting quality, and reproducibility. These domains are further broken down into 24 specific items. Each item within these domains is assigned a certain number of points, contributing to the overall score. The maximum overall score that a study can achieve using the APPRAISE-AI tool is 100 points. Higher scores indicate stronger methodological or reporting quality. This point system allows for a detailed and nuanced assessment of each study, focusing on different aspects critical for the validity and reliability of AI research in the medical field (10).

Statistical analysis

Forest plots and summary receiver operating characteristic (ROC) curves were generated. The pooled Area Under the Curve (AUC), sensitivity, and specificity were used to assess the ability of radiomics to predict the treatment response of NAC in MIBC patients. Heterogeneity was visually checked and evaluated using I^2; an I^2 value of 0%–25% meant unremarkable heterogeneity, 25%–50% meant reduced heterogeneity, 50%–75% meant moderate heterogeneity, and >75% meant high heterogeneity. IBM SPSS Statistics (version

© 2024 Canadian Urological Association

24; IBM Corporation, Armonk, NY, USA) and Review Manager (version 5.4.1) were employed for statistical analyses.

A meta-analysis of investigations related to the prediction of the treatment response of NAC in MIBC patients was further performed. Two independent reviewers retrieved the data. A third reviewer assessed the internal validity. Only studies with similar designs and outcomes that provided a two-by-two contingency table or had enough data to reconstruct such a table were eligible for analysis.

We opted to analyze sensitivity and specificity over AUC as primary metrics due to their direct reflection of the machine learning models' diagnostic accuracy, providing clear, actionable insights into patient outcomes. These metrics are fundamental in determining the true positive and negative rates, essential for the accurate categorization of responders and non-responders to treatment—critical in the clinical setting. Additionally, our meta-analysis concentrated exclusively on radiomic studies to capitalize on the homogeneity and comparability of data, which is pivotal for the integrity of meta-analytic conclusions. Radiomics—a burgeoning field in oncological imaging—offers quantifiable insights from medical images, presenting a more standardized and replicable dataset across studies for predictive modeling.

Our study prioritizes sensitivity and specificity over AUC for their clarity in evaluating machine learning predictions of chemotherapy response in bladder cancer, ensuring clinical relevance. We exclusively analyzed radiomic studies in our meta-analysis to maintain data homogeneity, reflecting the precision required for robust predictive modeling in personalized medicine.

In cases where multiple models were presented, the model with the most significant accuracy was chosen for our study. A coupled forest plot of sensitivity and specificity was created using RevMan (version 5.4.1). The ROC curve, combined sensitivity, and combined specificity were used to assess the ability of machine learning algorithms to predict the response of NAC.

RESULTS

Literature screening

The literature search retrieved 655 papers. One hundred and fifty-seven duplicate studies were automatically excluded. After title and abstract screening of the 498 unique references, 472 records were excluded due to irrelevance to the study’s aim. The full texts of the remaining 26 studies were assessed for eligibility. Finally, 12 studies were selected and included. Figure 2 shows the PRISMA flowchart of the literature search. In Table 1, we summarized information about the selected articles, such as institution, year of publication, country, study design, total number of patients, and type of data used.

Evaluation criteria for neoadjuvant chemotherapy

The response evaluation of NAC in all included studies assessing imaging was based on the Response Evaluation Criteria in Solid Tumors 1.1 (RECIST 1.1) (11). All studies used the anatomopathological pT0 status of RC as a definition of pathologic response, with most

© 2024 Canadian Urological Association

considering ypT0 (9/12) as the response and three considering ypT1 (3/12) (ie, downstaging to non-muscle invasive).

Study evaluation

The quality of studies was assessed by the APPRAISE-AI tool (11), which is presented in Table 2. A range of scores indicates varying levels of research quality in using AI as a predictive tool following neoadjuvant chemotherapy for bladder cancer. The overall APPRAISE-AI scores out of 100 for the studies vary, with some studies like Lu 2023, Mohanad 2021, and Wu 2019 receiving high scores (67, 67, and 61, respectively), indicating a high quality of research. Conversely, studies such as Parmar 2021 and Sun 2022, with scores of 47 and 52, fall into the moderate quality category. This variability underscores the evolving nature of AI applications in this clinical setting and the lack of methodological standardization. The detailed analysis provided by the APPRAISE-AI tool allows us to delve into specific aspects such as clinical relevance, data quality, methodological conduct, robustness of results, reporting quality, and reproducibility. Notably, all studies scored well in clinical relevance, suggesting a strong alignment of the AI tools with clinical needs in bladder cancer treatment. However, the scores in categories like robustness of results and methodological conduct varied more significantly, highlighting areas where future research could focus to enhance the quality of the studies.

Ten studies examined were conducted in individual and 2 multi-institutional institutions. The absence of sample size calculations in the analyzed articles is a significant methodological gap that prevents the evaluation of statistical power and generalization of results. In addition, there is a lack of transparency in the treatment of missing data, which may compromise the integrity and reliability of the study conclusions. Regarding the potential data leakage, that is, contamination between the training and test sets, the articles did not provide enough information to ensure the absence of this practice. The assessment of bias, particularly in relation to gender differences in model performance, was not adequately addressed in the studies, leaving open questions of equity and applicability. Finally, the unavailability of models or datasets creates significant barriers to external validation and replication of studies by other researchers.

The neoadjuvant therapy regimen varied between studies, which impairs the comparison between models and applicability in other neoadjuvant treatment regimens. In the future, new studies with standardized treatments should be performed and validated, allowing the extension of the conclusions to other cohorts of patients.

Meta-analysis

Four studies were selected for the quantitative meta-analysis since the other studies involving radiomics did not present enough data to reconstruct a 2x2 contingency table. The meta-analysis did not include the four studies involving circulating tumor DNA (ctDNA), epigenetics, and anatomopathological criteria, as their methodologies were too specific to allow comparison.

Based on our analysis, the meta-analytical sensitivity and specificity for response to NAC were 0.62 (95% CI: 0.50 - 0.72) and 0.82 (95% CI: 0.72 - 0.89), respectively, as

© 2024 Canadian Urological Association

evidenced by the corresponding Forest Graphs in Figure 3. The I^2 score was 38.5%, representing a low heterogeneity level within eligible studies with heterogeneity. The ROC curve is provided in Figure 4. Given that we had less than ten eligible articles in our meta-analysis, the Egger test was not applicable, as suggested by the Cochrane guidelines.

This systematic review summarized 12 prediction models for response to NAC in MIBC patients. The models were grouped according to the data type utilized in the machine learning algorithms and included several promising predictors, with some models showing encouraging predictive potential. However, nine studies considered responders patients with ypT0, while three studies considered responders all patients with ≤ypT1. In addition, most studies used small samples, and external validation in independent studies was insufficient.

Machine learning with computerized tomography

Eight retrospective cohorts used abdominal computed tomography images as input for machine learning algorithms to predict response to NAC in patients with MIBC. Deep Learning-Convolutional Networks and Random Forests were the most used machine learning approaches. The number of patients in the training cohorts ranged from 19 to 87, while the validation cohort ranged from 9 to 123. Five studies mainly used MVAC as NAC. Seven studies considered ypT0 as responders, and one study (21) considered ≤ypT1 as responders. The accuracy of the algorithms ranged from 0.69 +/- 0.08 to 0.80 +/- 0.04.

Cha et al. (2016) published the first article exploring the possibility of predictive models based on radiomics that may be able to distinguish between patients who experienced response (ypT0) to NAC and those who did not, based on the analysis of pre-and post-treatment CT images (12). The tested models of radiometry involved a pattern recognition method (DL-CNN). The evolution of machine learning techniques has progressively improved in accuracy since 2016, with increasingly efficient and elaborate algorithms for training and validation.

Our meta-analysis showed a specificity of 82% (0.73 - 0.89), which demonstrates the great capacity of the algorithms to identify patients who, based on their likelihood of complete response, can be chosen for bladder-sparing despite the modest sensitivity of 62% (0.50 - 0.72).

In a study conducted by Cha et al. (2019), radiologists demonstrated an accuracy of 0.74 (range: 0.66 - 0.78) without the aid of machine learning. When supported by machine learning, accuracy statistically improved to 0.77 (range: 0.73-0.81), $p<0.001$ (15). These findings are similar to those of Hadjiiski et al. (2020), who reported an average AUC of 12 physicians examined in scenarios with and without machine learning assistance (16).

Moreover, there is notable variability in accuracy without machine learning, with a standard deviation of 26,53. The integration of machine learning markedly mitigated this variability, resulting in a more uniform accuracy across physicians, evidenced by a marginally reduced standard deviation of 21,59 (16). Currently, radiomics is mainly helpful to auxiliary doctors in radiological diagnosis. ML can help non-experts' performance and even achieve the same accuracy as experienced doctors (26,22).

Parmar et al. performed preoperative tomography to predict the response to neoadjuvant chemotherapy, developing an algorithm to characterize a signature for predicting

© 2024 Canadian Urological Association

the response to neoadjuvant chemotherapy in muscle-invasive bladder cancer before any treatment (21).

Machine learning with genetic data

Three studies employed genetic analyses to predict NAC responses. These comprised a retrospective case-control study (19), a retrospective cohort study (18), and a prospective study (23). Investigations focusing on DNA have shown promising outcomes with satisfactory accuracies. However, these studies often have limited participants and need external validation for clinical applicability.

Mohanad et al. (2022) conducted a retrospective case-control study involving 100 patients—70 with MIBC and 30 with unaltered urothelium. Their primary aim was to investigate the epigenetic inactivation of DNA Damage Repair (DDR) genes as predictive biomarkers and to identify patients likely to benefit from NAC (19). The model combined the expression of Retinoblastoma Binding protein 8 (RBBP8) and MutS homolog 4 (MSH4), ascertained via qRT-PCR, with clinical features such as tumor size, grade, stage, and lymph node metastasis.

Algorithms like Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF) were utilized. The most efficacious predictive model was the KNN, with an accuracy of 90.05 ± 4.5%, a sensitivity of 87.8%, and a specificity of 90.7%. A combination of KNN and RF models yielded an accuracy of 90.0 ± 3.4%, a sensitivity of 92.98%, and a specificity of 81.4%. Nevertheless, while the model's predictive strength is commendable, it did not distinguish between patients treated with neoadjuvant and adjuvant chemotherapy (19).

The second study, by Hepburn et al. (2021), was a retrospective cohort investigation that utilized a 41-gene panel to predict NAC responses. The study compared the performance of a machine learning algorithm with that of a microarray differential gene expression analysis. Patients were deemed responders if they exhibited anatomopathological staging of radical cystectomy ≤ T1. The Rank-Guided Iterative Feature Elimination (RGIFE) algorithm, designed to iteratively remove and evaluate feature blocks, was used. The most predictive of the gene panels generated was "panel 1," encompassing nine genes. SHapley Additive exPlanations (SHAP) were employed to ascertain the contribution of each gene. Notably, the HIST1H4F gene consistently appeared as a response predictor, up-regulated in responders, while the CNGB1 gene was identified as up-regulated in non-responders (18).

The final study, an exploratory analysis of the prospective research SWOG S1314, contrasted the effects of neoadjuvant chemotherapy with gemcitabine and cisplatin versus dense-dose MVAC. The primary objective was to delineate the role of cell-free DNA (cfDNA) methylation in predicting neoadjuvant responses using a machine-learning approach. Plasma samples were procured before and after one chemotherapy cycle. Using Random Forest, a classifier named Methylation-based Response score (mR-score) was devised and crafted employing Elastic Net. In the prediction analysis, the mR-score, built with Random Forest, presented a receiving area operating characteristic under the curve AUC of 0.636 (95% CI [0.498 - 0.773]), while the Elastic Net-derived model attained an AUC of 0.639 (95% CI [0.503 - 0.774]).

© 2024 Canadian Urological Association

It is worth noting that the sample size of this study was limited, and the training of the models focused on overall cfDNA methylation, of which bladder-derived DNA formed a minor fraction, ranging between 0-3%. Consequently, the mR-score was not dictated by residual disease, as evidenced by its lack of correlation with circulating bladder DNA (23).

Machine learning with anatomopathological data

In a retrospective study conducted by Mi et al. (2021), a total of 129 MIBC who underwent NAC were investigated. Transurethral resection of the bladder anatomopathological specimens were obtained from these patients prior to NAC and were subsequently analyzed using tissue microarrays, incorporating hematoxylin-eosin staining and immunohistochemistry. The captured images were meticulously examined, focusing on three cellular categories: cancer cells, lymphocytes, and stromal cells (17).

From this analysis, 1,187 features were derived and integrated with demographic and clinical data. These consolidated data points were then processed using machine learning algorithms, specifically the Support Vector Machine and Random Forest techniques. The most salient features were discerned based on their frequency of occurrence (17). Notably, diversified cell morphology, elevated stromal load, distinct cell orientation, and a decreased CD8-Foxp3 ratio were found to be correlated with resistance to NAC. Conversely, rapid cell proliferation, close cellular proximity, elongated cellular structure, and the presence of luminal MIBC were identified as indicators of a higher likelihood of response to NAC (17).

It is important to highlight that the cohort size for this study was relatively limited, and the validation process did not incorporate immunohistochemical markers. Nonetheless, the predictive model showcased a commendable accuracy in determining the response to NAC in MIBC patients, with accuracy scores spanning from 65% to 73%. Fusing anatomopathological features with other clinical attributes holds significant promise for future research endeavors (17).

DISCUSSION

Our results indicate that some pre-treatment prediction models, particularly those based on CT scan features, genetics, and pathology, may have significant clinical value in the future. Models must be further developed and externally validated in independent prospective cohorts with sufficient sample size. Specifically, there is a need to standardize specimen acquisition, study design, and study endpoints.

One area of promise is magnetic resonance imaging with radiomics to predict response to NAC and assess the specimen for the complete response. In the future, the models should be integrated into a multi-omics concept, increasing accuracy, and allowing for an individualized precision approach.

It is important to underline the contribution of machine learning in increasing our understanding of responses to neoadjuvant treatment in bladder tumors. Unlike traditional statistical methods, which focus predominantly on numerical predictions, machine learning models excel at discovering complex and nonlinear relationships within large datasets, offering a deeper and differentiated understanding of treatment outcomes. The analysis of

© 2024 Canadian Urological Association

additive Shapley explanations (SHAP), as exemplified in Hepburn's study, elucidates this aspect by providing transparent insights into the decision-making processes of AI models (18)

Random Forest and Convolutional Neural Networks were among the most utilized models in predicting the response to neoadjuvant chemotherapy in bladder cancer, mainly when dealing with computed tomography, genetic data, and histopathological images. The versatility of Random Forest, an ensemble learning method, lies in its ability to handle large datasets with multiple variables, making it ideal for genetic data analysis. On the other hand, CNNs excel in image processing, an essential aspect when analyzing computed tomography and histopathological images. Their deep learning architecture enables them to automatically and efficiently extract and learn feature hierarchies, crucial for identifying subtle patterns and characteristics in medical images indicative of chemotherapy response. This synergy of Random Forest's adeptness at handling genetic data and CNN's proficiency in image analysis makes them highly effective for predicting neoadjuvant chemotherapy response in patients with bladder cancer.

In addition, our review reveals a critical gap between the advanced capabilities of these statistical models and their practical implementation in clinical settings. Addressing this gap requires a joint effort to validate and translate these AI-driven insights into tangible clinical practices, eliminating the division between theoretical models and real-world medical applications. This shift toward a more interpretative and application-focused approach to AI research is not just a technological breakthrough but a paradigm shift in understanding and responding to complex medical challenges in oncology.

CONCLUSIONS

Studies that used machine learning to evaluate the efficacy of NAC in MIBC patients are still limited, including a small number of patients/cases, and lack external validation. The clinical potential of using artificial intelligence in this context is promising but needs a clear strategy for validation and standardization of the methodological analysis of the data.

DRAFT

© 2024 Canadian Urological Association

REFERENCES

1. Kozikowski M, Suarez-Ibarrola R, Osiecki R, et al. Role of radiomics in the prediction of muscle-invasive bladder cancer: A systematic review and meta-analysis. *Eur Urol Focus* 2020;8:728-38. https://doi.org/10.1016/j.euf.2021.05.005
2. Binder N, Dette H, Franz J, et al. Data mining in urology: Understanding real-world treatment pathways for lower urinary tract systems via exploration of big data. *Eur Urol Focus* 2020;8:391-3. https://doi.org/10.1016/j.euf.2022.03.019
3. Klén R, Salminen AP, Mahmoudian M, et al. Prediction of complication related death after radical cystectomy for bladder cancer with machine learning methodology. *Scand J Urol* 2019;53:325-31. https://doi.org/10.1080/21681805.2019.1665579
4. Laukhtina E, Rajwa P, Mori K, et al. Accuracy of frozen section analysis of urethral and ureteral margins during radical cystectomy for bladder cancer: A systematic review and diagnostic meta-analysis. *Eur Urol Focus* 2022;8752-60. https://doi.org/10.1016/j.euf.2021.05.010
5. Hastie T, Tibshirani R, Friedman JH (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, Springer. https://doi.org/10.1007/978-0-387-84858-7
6. Ikeda A, Nosato H, Kochi Y, et al. Support system of cystoscopic diagnosis for bladder cancer based on artificial intelligence. *J Endurol* 2020;34:352-8. https://doi.org/10.1089/end.2019.0509
7. Wu S, Hong G, Xu A, et al. Artificial intelligence-based model for lymph node metastases detection on whole slide images in bladder cancer: A retrospective, multicentre, diagnostic study. *Lancet Oncol* 2023;24:360-70. https://doi.org/10.1016/S1470-2045(23)00061-X
8. Witjes JA, Bruins HM, Carrión A, et al . European Association of Urology Guidelines on muscle-invasive and metastatic bladder cancer: Summary of the 2023 guidelines. *Eur Urol* 2024;85:17-31. https://doi.org/10.1016/j.eururo.2023.08.016
9. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ* 2015;350:g7647. https://doi.org/10.1136/bmj.g7647
10. Kwong JCC, Khondker A, Lajkosz K, et al. APPRAISE-AI tool for quantitative evaluation of ai studies for clinical decision support. *JAMA Netw Open* 2023;6:e2335377. https://doi.org/10.1001/jamanetworkopen.2023.35377
11. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228-47. https://doi.org/10.1016/j.ejca.2008.10.026
12. Cha KH, Hadjiiski LM, Samala RK, et al. Bladder cancer segmentation in ct for treatment response assessment: Application of deep-learning convolution neural network-a pilot study. *Tomography* 2016;2:421-9. https://doi.org/10.18383/j.tom.2016.00184

© 2024 Canadian Urological Association

13. Cha KH, Hadjiiski L, Chan HP, et al. Bladder cancer treatment response assessment in ct using radiomics with deep-learning. *Sci Rep* 2017;7:8738. https://doi.org/10.1038/s41598-017-09315-w
14. Wu E, Hadjiiski LM, Samala RK, et al. Deep learning approach for assessment of bladder cancer treatment response. *Tomography* 2019;5:201-8. https://doi.org/10.18383/j.tom.2018.00036
15. Cha KH, Hadjiiski LM, Cohan RH, et al. Diagnostic accuracy of ct for prediction of bladder cancer treatment response with and without computerized decision support. *Acad Radiol* 2019;26:1137-45. https://doi.org/10.1016/j.acra.2018.10.010
16. Hadjiiski LM, Cha KH, Cohan RH, et al. Intraobserver variability in bladder cancer treatment response assessment with and without computerized decision support. *Tomography* 2020;6:194-202. https://doi.org/10.18383/j.tom.2020.00013
17. Mi H, Bivalacqua TJ, Kates M, et al. Predictive models of response to neoadjuvant chemotherapy in muscle-invasive bladder cancer using nuclear morphology and tissue architecture. *Cell Rep Med* 2021;2:100382. https://doi.org/10.1016/j.xcrm.2021.100382
18. Hepburn AC, Lazzarini N, Veeratterapillay R, et al. Identification of CNGB1 as a predictor of response to neoadjuvant chemotherapy in muscle-invasive bladder cancer. *Cancers (Basel)* 2021;13:3903. https://doi.org/10.3390/cancers13153903
19. Mohanad M, Yousef HF, Bahnassy AA. Epigenetic inactivation of DNA repair genes as promising prognostic and predictive biomarkers in urothelial bladder carcinoma patients. *Mol Genet Genomics* 2022;297:1671-87. https://doi.org/10.1007/s00438-022-01950-x
20. Choi SJ, Park KJ, Heo C, et al. Radiomics-based model for predicting pathological complete response to neoadjuvant chemotherapy in muscle-invasive bladder cancer. *Clin Radiol* 2021;76:627.e13-627.e21. https://doi.org/10.1016/j.crad.2021.03.001
21. Parmar A, Qazi AA, Stundzia A, et al. Development of a radiomic signature for predicting response to neoadjuvant chemotherapy in muscle-invasive bladder cancer. *Canad Urol Assoc J* 2022;16:E113-9. https://doi.org/10.5489/cuaj.7294
22. Sun D, Hadjiiski L, Alva A, et al. Computerized decision support for bladder cancer treatment response assessment in ct urography: Effect on diagnostic accuracy in multi-institution multi-specialty study. *Tomography* 2022;8:644-56. https://doi.org/10.3390/tomography8020054
23. Lu YT, Plets M, Morrison G, et al. Cell-free DNA methylation as a predictive biomarker of response to neoadjuvant chemotherapy for patients with muscle-invasive bladder cancer in SWOG S1314. *Eur Urol Oncol* 2023;6:516-24. https://doi.org/10.1016/j.euo.2023.03.008

© 2024 Canadian Urological Association

FIGURES AND TABLES

Figure 1. Clinical application of artificial intelligence.

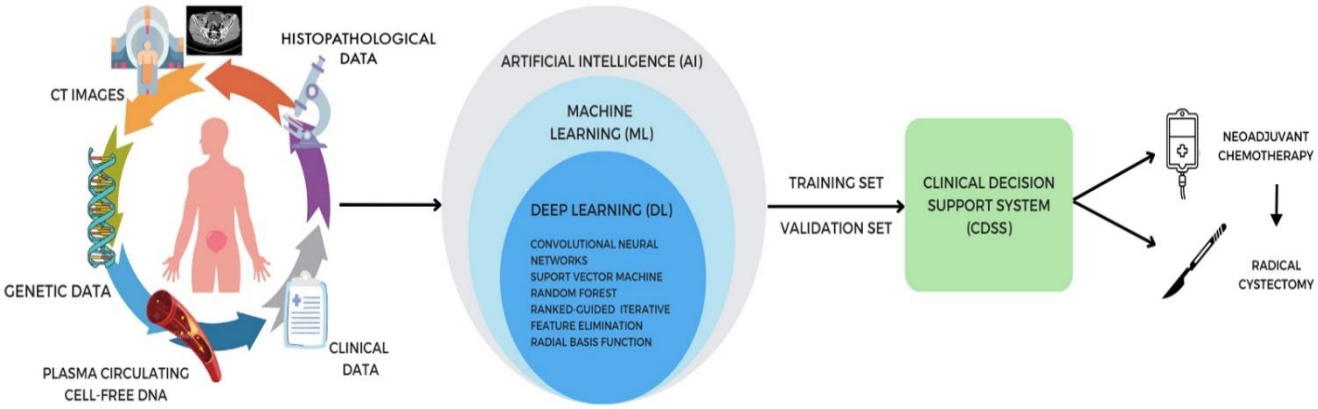
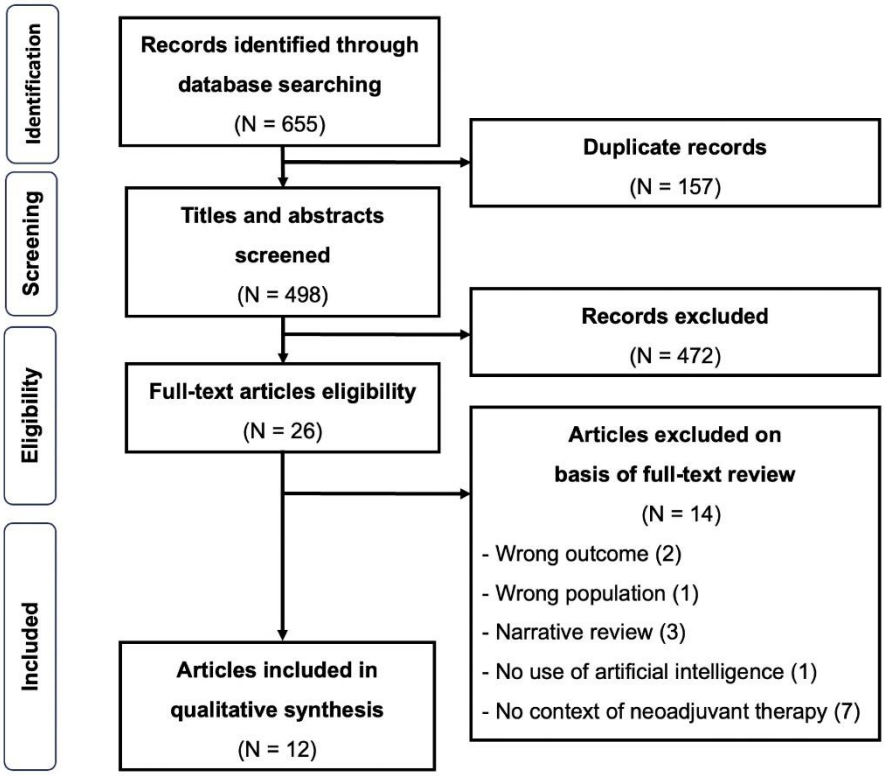
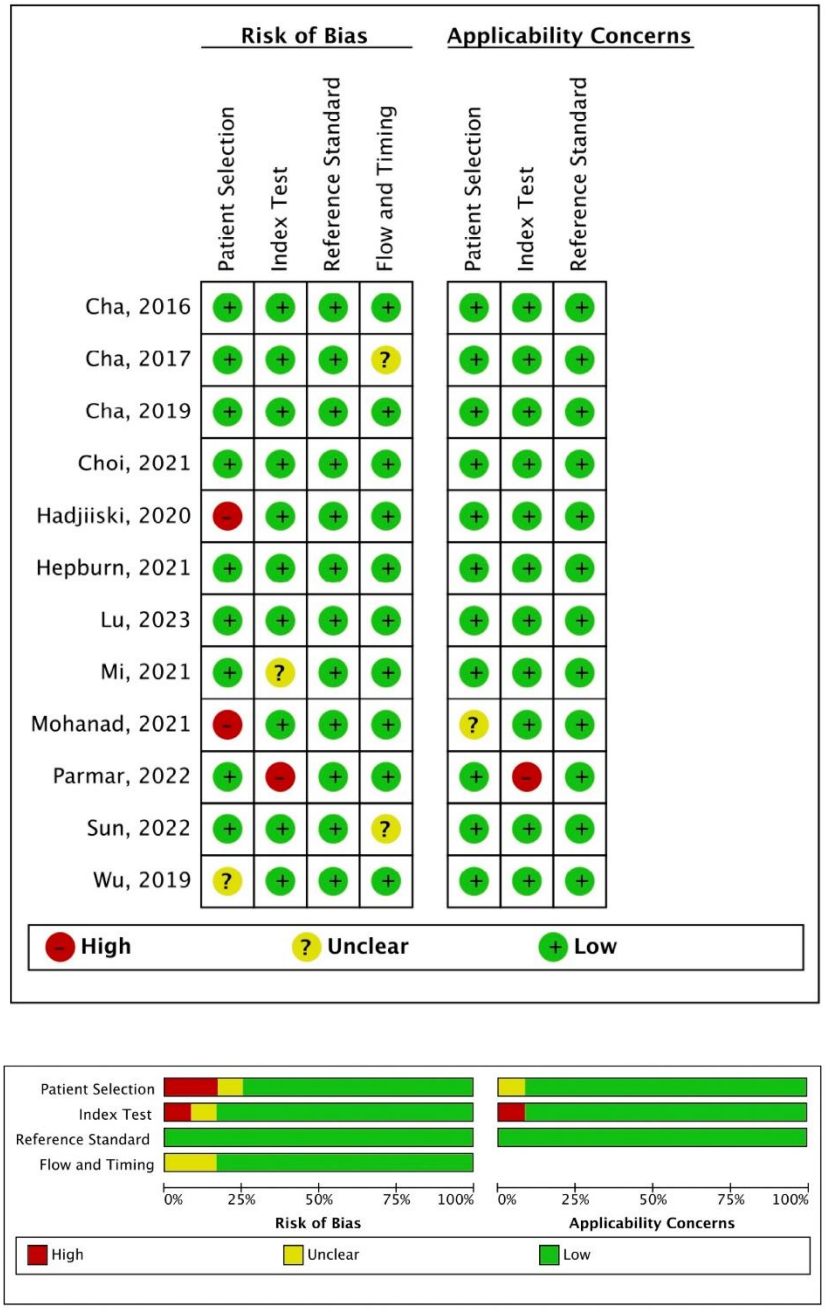


Figure 2. PRISMA flowchart of the literature research.



© 2024 Canadian Urological Association

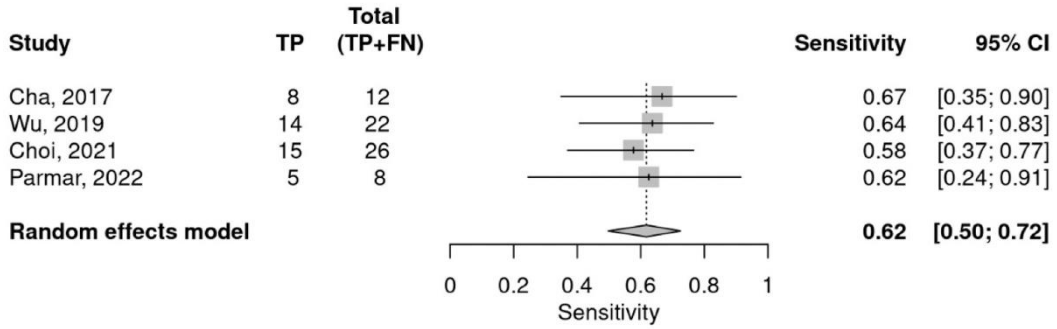
Figure 3.



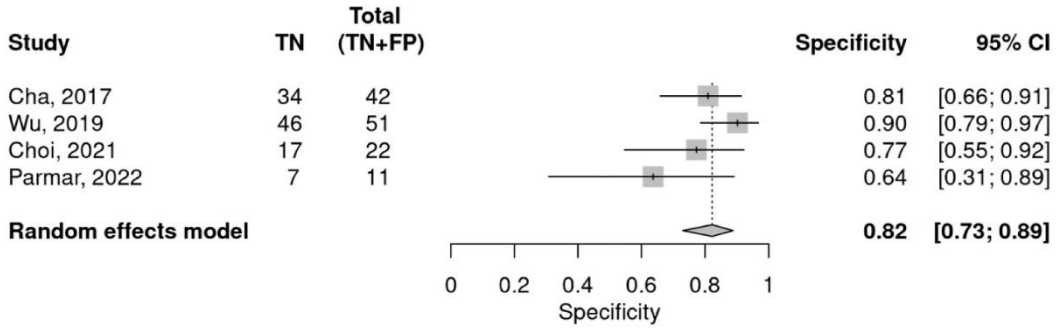
© 2024 Canadian Urological Association

Figure 4. Forest plots. (A) sensitivity; (B) specificity; (C) global analyses

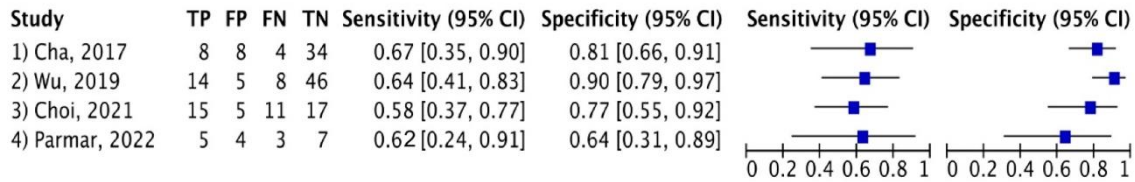
A



B



C



© 2024 Canadian Urological Association

Table 1. Characteristics of eligible studies.

Study	Year	Country	Study design	Number of patients	Type of data
Cha et al[12]	2016	USA	Retrospective cohort	91	CT images
Cha et al[13]	2017	USA	Retrospective cohort	123	CT images
Wu et al[14]	2019	USA	Retrospective cohort	123	CT images
Cha et al[15]	2019	USA	Retrospective cohort	205	CT images
Hadjiiski et al[16]	2020	USA	Retrospective cohort	199	CT images
Mi et al[17]	2021	USA	Retrospective cohort	122	Imaging of biopsy pathology specimens pre neoadjuvant treatment
Hepburn et al[18]	2021	UK	Retrospective cohort	129	Genetic data Microarray gene expression data
Mohanad et al[19]	2021	Egypt	Retrospective case-control	100	RNA-sequencing data
Choi et al[20]	2021	Korea	Retrospective cohort	135	CT images
Parmar et al[21]	2021	Canada	Retrospective cohort	19	CT images
Sun et al[22]	2022	USA	Retrospective cohort	205	CT images
Lu et al[23]	2023	USA	Prospective	72	Plasma circulating cell-free DNA

CT: computed tomography

DRAFT

© 2024 Canadian Urological Association

Table 2. Evaluation of the quality of studies with APPPRAISE-AI score

	Cha 2016	**Cha 2017**	**Cha 2019**	**Choi 2021**	**Hadjiiski 2020**	**Hepburn 2021**	**Lu 2023**	**Mi 2021**	**Mohanad 2021**	**Parmar 2021**	**Sun 2022**	**Wu 2019**
Overall APPRAISE-AI score (out of 100)	54	55	60	63	60	60	67	55	67	47	52	61
Quality based on overall APPRAISE-AI score	Moderate	Moderate	High	High	High	High	High	Moderate	High	Moderate	Moderate	High
Clinical relevance (out of 4)	4	4	4	4	4	4	4	4	4	3	3	3
Data quality (out of 24)	13	15	14	15	15	12	18	15	18	12	13	15
Methodological conduct (out of 20)	13	12	12	13	13	14	14	12	10	8	14	12
Robustness of results (out of 20)	4	3	9	9	6	3	3	3	9	5	6	9
Reporting quality (out of 12)	9	10	10	10	10	12	12	10	10	8	11	10
Reproducibility (out of 20)	11	11	11	12	12	15	16	11	16	11	5	12

The overall APPRAISE-AI score was graded as follows: very low quality, 0–19; low quality, 20–39; moderate quality, 40–59; high quality, 60–79; very high quality, 80–100.

DRAFT

© 2024 Canadian Urological Association