



COMMENTARY

Open Access

# Enhanced CT-based deep learning radiomics for bladder cancer prognosis

William C. Cho<sup>1\*</sup>

**Key words** Artificial intelligence (AI), Computed tomography (CT), Deep learning (DL), Muscle-invasive bladder cancer, Radiomics

Bladder cancer (BLCA) is the eighth most frequent cancer and the thirteenth leading cause of cancer death worldwide, with an estimated 635,000 new cases and 228,000 deaths in 2024 [1]. The clinical management of BLCA relies on one key question: has the tumor invaded the muscularis propria? This binary distinction, non-muscle-invasive vs. muscle-invasive binary classification, decides whether a patient undergoes organ-preserving endoscopic resection or radical cystectomy with systemic therapy. Yet the current gold standard, histopathological evaluation of transurethral resection specimens, has its constraints [2]. A biopsy may not capture the full extent of the tumor. The heat from transurethral resection equipment can distort tissue architecture. Variability among operators introduces inconsistency. These issues lead to understaging in many cases, with incidence varying by institution and patient population [3]. He *et al.* [4] have developed an artificial intelligence (AI)-driven solution to this longstanding clinical challenge.

Their work stands out for its methodological approach. The authors combined two complementary methods, traditional radiomics (which collects measurable features such as tumor size, shape, and interface characteristics) and deep learning (DL), particularly a convolutional neural network, which captures complex and high-dimensional representations that traditional feature extraction often overlooks [4,5]. This hybrid system resembles how an experienced clinician weighs measurable tumor characteristics against intuitive pattern-based assessment gained from years of practice. Internal validation showed an area under the receiver operating characteristic curve (AUC) of 0.807, while the external multi-center cohort achieved 0.783. These outcomes were

significantly better than those of the clinical model used alone. They also exceeded the predictive abilities of either the radiomic or DL components when used separately [5]. Nevertheless, several limitations warrant careful consideration. The DL backbone (ResNet101) was initially developed for natural image recognition. While its transferability is impressive, it is not fully optimized for computed tomography (CT) imaging. Future improvements will likely rely on designs specifically tailored for medical use. Some methodological points deserve consideration as well. The validation framework relies on a single training/validation split without cross-validation or bootstrap resampling. Performance metrics derived from a single split are highly sensitive to how that split was drawn. Without repeated sampling, there is no estimate of variance. On the one hand, the use of generative adversarial network (GAN)-based super-resolution also warrants close examination. GAN creates synthetic images that do not correspond to any actual patient scan. The boundary between synthetic and real data in the validation set is not clearly stated. If GAN-generated images are present in the validation set, reported performance will be inflated. On the other hand, the intensity standardization across centers used a fixed Hounsfield unit (HU) range of -75 to 175. This lower bound partially excludes the perivesical fat signal, which typically occupies -50 to -100 HU. Given that perivesical fat plane integrity is a key criterion for T3 staging, this truncation may limit the model's sensitivity to invasion-related features. The study's reliance on CT (the primary imaging modality for 2249 of 2909 patients) reflects clinical reality, yet CT's limited ability to resolve soft tissue compared to magnetic resonance imaging remains a fundamental limitation. The generalizability of the study population deserves attention. Most training and validation cohorts came from Chinese academic centers, representing mostly East Asian patients. The inclusion of The Cancer Genome Atlas (TCGA)-BLCA provided a preliminary test

This commentary refers to the article available online at <https://www.sciencedirect.com/science/article/pii/S2054936926000014>

\*Correspondence: William C. Cho, [chocs@ha.org.hk](mailto:chocs@ha.org.hk); [williamscho@gmail.com](mailto:williamscho@gmail.com)

<sup>1</sup>Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong SAR, China

of cross-ethnic applicability. The performance drop observed with TCGA-BLCA data (AUC=0.519) highlights the fragility of models when applied to different ethnic groups, imaging protocols, and clinical workflows. Strategies for adapting to domain differences and including multi-regional cohorts will be needed to enhance robustness, more extensive validation in large, diverse international cohorts is necessary.

Particularly encouraging is the model's ability to generalize across multiple institutions. Poor generalizability often challenges AI applications in medical imaging [6]. Against this backdrop, the authors' efforts to tackle data variability are commendable. Their use of GAN-based super-resolution, along with careful feature harmonization, serves as a useful methodological reference for future multi-center studies. The prospective human-AI interaction experiment, however, offers perhaps the most relevant clinical insight. AI assistance improved diagnostic accuracy among urologists, especially junior ones, addressing a real-world challenge where specialized uroradiology expertise is unavailable. The significant increase in sensitivity is particularly important. Missing a muscle-invasive BLCA diagnosis can lead to delayed treatment and disease progression, while over-calling may result in aggressive but potentially unnecessary surgery [7].

Beyond immediate applications, this methodological approach could be extended to other key questions in BLCA treatment. Predicting responses to neoadjuvant chemotherapy, forecasting immunotherapy outcomes, identifying patients at high risk for recurrence, and anticipating treatment resistance before it becomes obvious are all possible areas for study [8]. While the authors mentioned applicability to other malignancies, BLCA itself deserves priority since the infrastructure and expertise developed here can be directly utilized for these related tasks. Cervical cancer may represent another cancer type with a comparable clinical need for noninvasive risk stratification. Notably, CT imaging is already widely integrated into its standard care.

Integrating such AI tools into multidisciplinary tumor boards represents a natural step forward. Other studies have already started combining imaging-based risk assessments with molecular analysis (genomic, transcriptomic, or histopathological data) to refine treatment decisions [9,10]. These multimodal approaches, which integrate DL signatures with molecular biomarkers, offer a more complete picture of tumor biology and patient outcomes. The DL radiomics nomogram risk score, incorporated as a factor in a multivariate Cox model alongside clinical factors, shows independent prognostic value. This suggests the imaging characteristics reflect not only invasion depth but also the underlying

aggressiveness of the tumor.

Several specific considerations merit emphasis for clinicians evaluating this model's potential adoption. The authors did not stratify their analysis by patient subgroups that might derive particular benefit from AI-assisted interpretation, such as those with ambiguous imaging findings, equivocal biopsy results, or lesions at the extremes of size where visual assessment is most challenging. This represents a missed opportunity to identify the specific clinical scenarios where the model adds the greatest value. Similarly, the study did not separately analyze patients with prior intravesical therapy, variant histology, or multifocal disease. These subgroups are over-represented in real-world practice, and understanding model performance within them would substantially enhance clinical translation. The question of comparative performance against existing prediction models also remains incompletely addressed. While the authors validated their model on an external dataset, they did not benchmark it directly against previously published radiomic signatures or the vesical imaging-reporting and data system (VI-RADS) scoring in the same patient cohort. Without head-to-head comparison, clinicians cannot easily determine whether this hybrid approach offers meaningful advantages over simpler, more established methods. The authors' assertion that their model outperforms prior work rests largely on indirect comparisons across disparate populations with varying inclusion criteria, which should be interpreted with caution.

In conclusion, He *et al.* [4] have developed a multi-center-validated model that improves diagnostic accuracy for muscle invasion and prognostic evaluation for overall survival. It gives clinicians a non-invasive tool to navigate the complex treatment landscape of BLCA. The next challenge is to translate these promising results into routine clinical practice. This requires implementation studies, health-economic assessments, and, above all, building trust within the clinical community. The journey from algorithm to bedside is rarely simple, but studies like this one indicate a clear path forward.

## Abbreviations

AI: Artificial intelligence  
AUC: Area under the receiver operating characteristic curve  
BLCA: Bladder cancer  
CT: Computed tomography  
DL: Deep learning  
GAN: Generative adversarial network  
HU: Hounsfield unit  
TCGA: The Cancer Genome Atlas

## Acknowledgements

The author utilized Grammarly (available at <https://app.grammarly.com/>) solely for grammar and vocabulary adjustments. All conceptual

development, technical content, and critical discussion remain entirely human-generated. The author takes full responsibility for the accuracy of information and academic integrity of this work.

#### Author contributions

WCC wrote, reviewed, and edited the manuscript. The author read and approved the final manuscript.

#### Funding

Not applicable.

#### Availability of data and materials

Not applicable.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The author declares that there is no competing interests.

#### References

1. Sung H, Filho AM, Laversanne M, Ferlay J, Siegel RL, Soerjomataram I, *et al.* Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *CA Cancer J Clin.* 2026;76(4):e70090. <https://doi.org/10.3322/caac.70090>.
2. de Angelis M, Scilipoti P, Santangelo A, Longoni M, Lonati C, Tremolada G, *et al.* The impact of histopathological evaluation at transurethral resection of bladder tumour on survival in radical cystectomy candidates. *BJU Int.* 2025;136(2):336–43. <https://doi.org/10.1111/bju.16714>.
3. Uğur R, Demiriz B, Nişli O, Keskin ET, Şahin S, Baytekin HF, *et al.* Who is at risk of residual tumor and pathologic upstaging at a second TURBT in non-muscle-invasive bladder cancer?. *Ann Surg Oncol.* 2026. <https://doi.org/10.1245/s10434-026-20268-1>.
4. He YB, Hu J, Liu Z, Xiao ZC, Liu JH, Liang HS, *et al.* Non-invasive evaluation of muscle invasion and survival prognosis in bladder cancer using enhanced CT-based deep learning radiomics: a multi-center real-world cohort study. *Mil Med Res.* 2026;13(1):100001. <https://doi.org/10.1016/j.mmr.2026.100001>.
5. El-Atier RA, Saraya MS, Saleh AI, Rabie AH. Accurate bladder cancer diagnosis using ensemble deep learning. *Sci Rep.* 2025; 15(1):12880. <https://doi.org/10.1038/s41598-025-95002-0>.
6. Suleman MU, Mursaleen M, Khalil U, Saboor A, Bilal M, Khan SA, *et al.* Assessing the generalizability of artificial intelligence in radiology: a systematic review of performance across different clinical settings. *Ann Med Surg (Lond).* 2025;87(12):8803–11. <https://doi.org/10.1097/MS9.0000000000004166>.
7. Li S, Chen R, Raj A, Xue N, Zhao F, Shen X, *et al.* Impact of the time of surgical delay on survival in patients with muscle-invasive bladder cancer. *Front Oncol.* 2022;12:1001843. <https://doi.org/10.3389/fonc.2022.1001843>.
8. Bai Z, Osman N, Brendel M, Tangen CM, Flaig TW, Thompson IM, *et al.* Predicting response to neoadjuvant chemotherapy in muscle-invasive bladder cancer via interpretable multimodal deep learning. *NPJ Digit Med.* 2025;8(1):174. <https://doi.org/10.1038/s41746-025-01560-y>.
9. Wang D, Liu C, Wang X, Liu X, Lan C, Zhao P, *et al.* Automated machine-learning framework integrating histopathological and radiological information for predicting IDH1 mutation status in glioma. *Front Bioinform.* 2021;1:718697. <https://doi.org/10.3389/fbinf.2021.718697>.
10. Muneer A, Waqas M, Saad MB, Showkatian E, Bandyopadhyay R, Xu H, *et al.* From classical machine learning to emerging foundation models: review on multimodal data integration for cancer research. *Artif Intell Rev.* 2026;59:119. <https://doi.org/10.1007/s10462-026-11522-9>.

<https://doi.org/10.1016/j.mmr.2026.100065>

**Cite this article as:** Cho WC. Enhanced CT-based deep learning radiomics for bladder cancer prognosis. *Mil Med Res.* 2026;13(1):100065.