OXFORD

JNCI J Natl Cancer Inst (2022) 114(2): djab179

doi: 10.1093/jnci/djab179
First published online September 2, 2021
Article

An Artificial Intelligence System for the Detection of Bladder Cancer via Cystoscopy: A Multicenter Diagnostic Study

Shaoxu Wu, MD,[1,2,†] Xiong Chen, MD,[1,†] Jiexin Pan, MD,[1] Wen Dong, MD,[1,2] Xiayao Diao, MD,[1] Ruiyun Zhang, MD,[3] Yonghai Zhang, MD,[4] Yuanfeng Zhang, MD,[4] Guang Qian, PhD,[5] Hao Chen, PhD,[6] Haotian Lin, PhD,[7,8] Shizhong Xu, BD,[1] Zhiwen Chen, MD,[9] Xiaozhou Zhou, MD,[9] Hongbing Mei, MD,[10] Chenglong Wu, MD,[10] Qiang Lv, MD,[11] Baorui Yuan, MD,[11] Zeshi Chen, MD,[1] Wenjian Liao, MD,[1] Xuefan Yang, MD,[1] Haige Chen, PhD,[3,‡] Jian Huang, PhD,[1,2,‡] Tianxin Lin, PhD[1,2,12,*,‡]

[1]Department of Urology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China; [2]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Guangzhou, China; [3]Department of Urology, Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China; [4]Department of Urology, Shantou Central Hospital, Shantou, China; [5]Peng Cheng Laboratory, Shenzhen, China; [6]Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, China; [7]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Centre, Sun Yat-sen University, Guangzhou, China; [8]Centre for Precision Medicine, Sun Yat-sen University, Guangzhou, China; [9]Departmemt of Urology, The First Hospital Affiliated to Army Medical University, Chongqing, China; [10]Department of Urology, Shenzhen Second People's Hospital, Shenzhen, China; [11]Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China; and [12]State Key Laboratory of Oncology in Southern China, Guangzhou, China

[†]These authors contributed equally.

[‡]These authors contributed equally to this work as last authors.

[*]**Correspondence to:** Tianxin Lin, PhD, Department of Urology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, 107th Yanjiangxi Rd, Guangzhou, China (e-mail: lintx@mail.sysu.edu.cn).

ARTICLE

Abstract

Background: Cystoscopy plays an important role in bladder cancer (BCa) diagnosis and treatment, but its sensitivity needs improvement. Artificial intelligence has shown promise in endoscopy, but few cystoscopic applications have been reported. We report a Cystoscopy Artificial Intelligence Diagnostic System (CAIDS) for BCa diagnosis. **Methods:** In total, 69 204 images from 10 729 consecutive patients from 6 hospitals were collected and divided into training, internal validation, and external validation sets. The CAIDS was built using a pyramid scene parsing network and transfer learning. A subset (n = 260) of the validation sets was used for a performance comparison between the CAIDS and urologists for complex lesion detection. The diagnostic accuracy, sensitivity, specificity, and positive and negative predictive values and 95% confidence intervals (CIs) were calculated using the Clopper-Pearson method. **Results:** The diagnostic accuracies of the CAIDS were 0.977 (95% CI = 0.974 to 0.979) in the internal validation set and 0.990 (95% CI = 0.979 to 0.996), 0.982 (95% CI = 0.974 to 0.988), 0.978 (95% CI = 0.959 to 0.989), and 0.991 (95% CI = 0.987 to 0.994) in different external validation sets. In the CAIDS vs urologists' comparisons, the CAIDS showed high accuracy and sensitivity (accuracy = 0.939, 95% CI = 0.902 to 0.964; sensitivity = 0.954, 95% CI = 0.902 to 0.983) with a short latency of 12 seconds, much more accurate and quicker than the expert urologists. **Conclusions:** The CAIDS achieved accurate BCa detection with a short latency. The CAIDS may provide many clinical benefits, from increasing the diagnostic accuracy for BCa, even for commonly misdiagnosed cases such as flat cancerous tissue (carcinoma in situ), to reducing the operation time for cystoscopy.

Globally, bladder cancer (BCa) is the 10th most prevalent cancer (1). Cystoscopy is the standard diagnostic and surveillance tool for BCa. Patients with suspicious lesions found on cystoscopy will undergo transurethral resection of bladder tumors (TURBt) for pathological diagnosis and staging. Approximately 75% of BCa cases are nonmuscle-invasive BCa and can be managed with TURBt (2,3). Follow-up with cystoscopy (every 3-6 months) is necessary to detect recurrence of BCa (4).

White light cystoscopy is the standard procedure for BCa diagnosis. Despite its excellent sensitivity in identifying papillary

Received: February 8, 2021; **Revised:** June 1, 2021; **Accepted:** September 1, 2021

lesions, flat cancerous tissue (carcinoma in situ [CIS]) and very small tumors can easily be missed by urologists, leading to high rates of misdiagnosis and incomplete TURBt (5). Currently, the misdiagnosis rate under white light is reported to be 30% (6,7), and the reported incomplete rate of TURBt is up to 50% (8,9), which contributes to high early recurrence and progression rates of BCa.

Therefore, new technologies should be developed to improve the sensitivity of cystoscopy. Fluorescence cystoscopy and narrow-band imaging have been developed and applied in various countries, contributing to increased detection of bladder lesions (10,11). However, the adoption of fluorescence cystoscopy and narrow-band imaging remains modest.

Artificial intelligence (AI) has been applied in many medical scenarios (12,13). For example, AI can automatically extract microimaging structures and identify pixel-level features undetectable by the human eye, and it has been used in medical imaging diagnosis, including in endoscopic diagnostic systems (14). When trained on large datasets of clinical endoscopic images or videos, AI can assist endoscopists in diagnosing cancers accurately and efficiently (15). However, regarding the AI-based detection of BCa via cystoscopy, only single-center and small-sample-size studies have been conducted with limited clinical impact (16,17).

Therefore, our aim was to develop a computer-assisted diagnosis system, the Cystoscopy Artificial Intelligence Diagnostic System (CAIDS), for detecting bladder lesions using real-world cystoscopic images collected from more than 10 000 patients at 6 hospitals.

Methods

Study Design and Participants

This multicenter study was implemented in 6 hospitals in China: 4 national hospitals, namely, Sun Yat-sen Memorial Hospital (SYSMH), Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine (RJH), the First Hospital Affiliated to Army Medical University (AMUFH), and the First Affiliated Hospital of Nanjing Medical University (NJFH); and 2 municipal hospitals, namely, Shenzhen Second People's Hospital (SZSH) and Shantou Central Hospital (STCH). The study design is shown in Figure 1.

All images were collected from consecutive patients who underwent cystoscopic examinations or TURBt. Cystoscopic examinations were scheduled for those with hematuria, or urothelial carcinomas detected by imaging, or fistulae, or follow-up for BCa patients. TURBt was performed for those with BCa detected by imaging or cystoscopy. Only patients with malignant cells in bladder lavage fluid or voided urine cytology were scheduled for random biopsies. Tumors were histologically confirmed by biopsy, TURBt, or radical cystectomy samples and staged in accordance with the 2016 TNM staging system (18). The grading system used in the study was based on the 2016 grading system of the World Health Organization (19). All patients were re-assessed for at least 6 months to avoid misdiagnosis. Those who had a BCa detected during follow-up were included in the tumor group. Images of histologically confirmed tumors were defined as tumor images, and cystoscopic images of patients without visible lesions or with lesions histologically excluded as tumors were defined as control images. This study was approved by the institutional review board of each participating hospital, and the requirement for informed consent was waived.

The collected images constituting the 2 largest datasets (from SYSMH and RJH) were divided into a training set and an internal validation set based on the data of operation at a ratio of 8:2 (80% of the images were assigned as training images, and the other 20% were assigned as validation images). To evaluate the generalizability of the developed system, we used the images collected from the other 4 hospitals as external validation sets.

Cystoscopy and Image Quality Control

All cystoscopic images were captured following a standard procedure under white light. The images were divided into a tumor group and a control group based on the histological results. Cystoscopic images of low quality due to poor insufflation, blur, halation or blood were excluded.

Twelve experienced urologists from SYSMH were divided into 6 groups for image quality assessment and labeling. All images were randomly divided among the 6 groups. In each group, 1 urologist assessed and labeled the images, and the other urologist double-checked the labeled images. If they had different opinions about the delineation, the image was sent to a professor with at least 20 years of experience in urology for redelineation.

Development of the CAIDS Algorithm

We built the CAIDS framework using a pyramid scene parsing network (PSPNet) trained on ImageNet and the training set. We chose the PSPNet architecture because it was superior to the FCN, SegNet, DilatedNet, and CascadeNet architectures on the ADE20K dataset (shown in the Supplementary Methods, available online) (20).

The proposed PSPNet consists of 3 main components: a preprocessing module, a feature extraction module, and a pyramid pooling module. The framework is shown in Figure 2.

The preprocessing procedure is a crucial step that helps enhance the quality of the data to promote the extraction of meaningful insights. The preprocessing module applies normalization and data augmentation to all training samples through the following methods: 1) the images are cropped along their longer edges into square images with the original center points fixed—for example, an image of 640×480 resolution will be cropped to 480×480 by removing 80×480 pixels on the left and right sides; 2) each image is resized to a 256×256 resolution; 3) to increase the limited available data to a meaningfully larger and more diverse amount, each sample in the training set is augmented into 4 transformed versions, including the original image, 1 flipped vertically, 1 flipped horizontally, and 1 rotated between -10 and 10 degrees; and 4) the pixel values of each image are normalized to [0,1].

The feature extraction module, a ResNet-101 model with weights pretrained on ImageNet, extracts the feature maps embedded in the raw data. Pretraining using the ImageNet database has 2 advantages: it can overcome the disadvantages of imbalanced data (as in our case) and avoid overfitting.

The pyramid pooling module, consisting of a pyramid of 4 parallel levels, can collect more representative levels of information compared with global pooling by concatenating different feature maps at different scales. The pyramid structure is important because its various pooling kernels cover the whole, half, and small portions of each image, resulting in receptive fields of 4 different sizes. Hence, both global (corresponding to

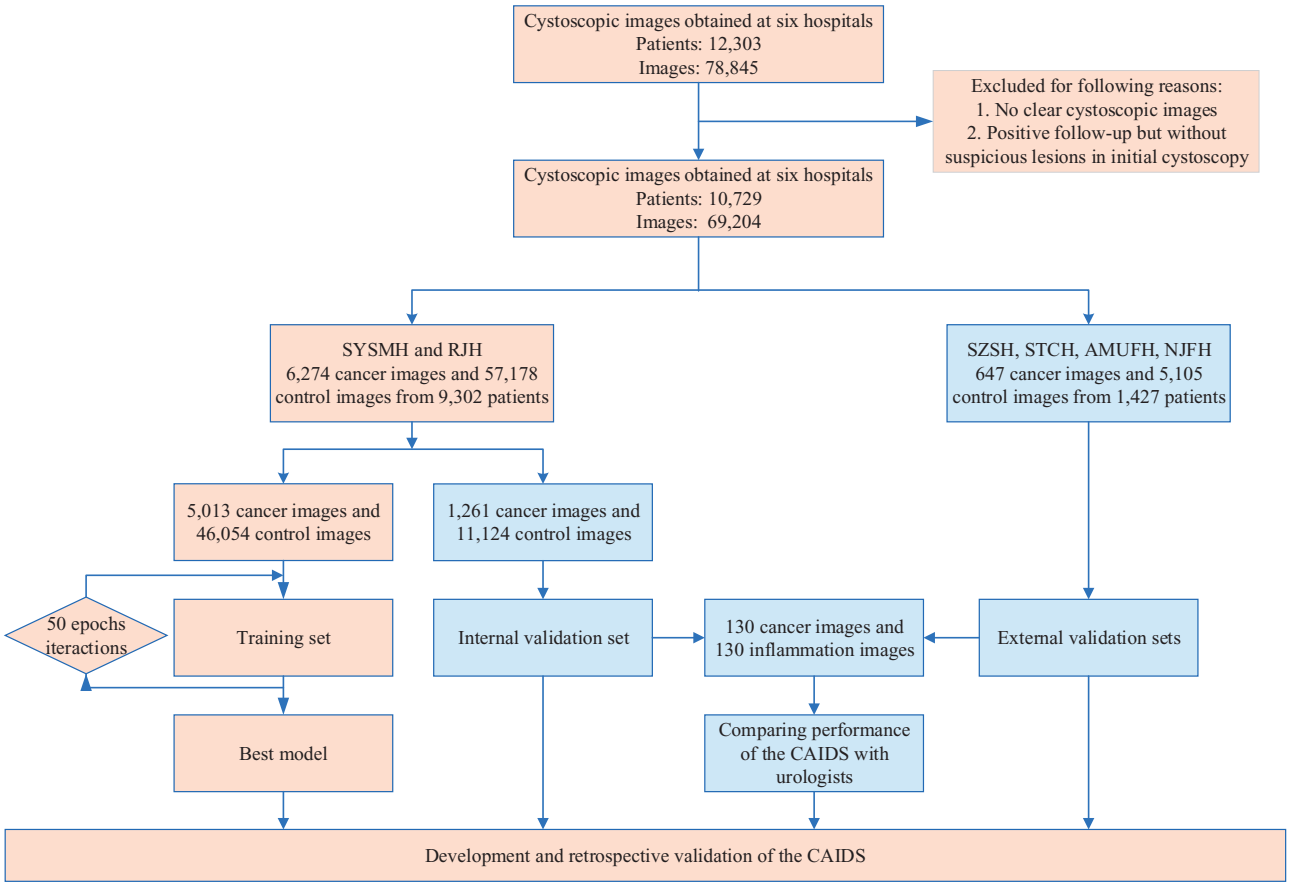

Figure 1. Flowchart of the study design for the development and validation of the Cystoscopy Artificial Intelligence Diagnostic System (CAIDS). AMUFH = The First Hospital Affiliated to Army Medical University; NJFH = The First Affiliated Hospital of Nanjing Medical University; RJH = Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine; STCH = Shantou Central Hospital; SYSMH = Sun Yat-sen University Memorial Hospital; SZSH = Shenzhen Second People's Hospital.

ARTICLE

larger receptive fields) and local (corresponding to smaller receptive fields) descriptors can contribute to the final prediction.

The CAIDS algorithm is presented in detail in the Supplementary Methods (available online).

Validation of the CAIDS Algorithm in the Validation Sets

The diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predicted value (NPV), and their 95% confidence intervals (CIs) were calculated using the Clopper-Pearson method.

The receiver operating characteristic (ROC) curves and precision-recall (PR) curves were plotted, and the areas under these curves were calculated to describe the diagnostic performance of the CAIDS. The F1 score was calculated as the harmonic mean of precision and recall, which is defined as $2 \times \text{precision} \times \text{recall}/(\text{precision} + \text{recall})$.

Then a subgroup of 260 images of complex lesions (130 CIS or small tumor images and 130 images of chronic inflammation randomly selected from the validation sets) was used to compare the performance of CAIDS with that of urologists in identifying BCa. "Small tumor" was defined as $<40 \times 40$ pixels, which could be easily misdiagnosed by human urologists. Nine urologists with varying degrees of expertise (3 trainees with approximately 2 years of experience, 3 competent urologists with approximately 7 years of experience, and 3 expert urologists with at least 20 years of experience) were asked to independently diagnose the 260 images. The time they spent on diagnosis was recorded. One month later, the 260 images were arranged into a different order randomly, and the 9 urologists were asked to make a diagnosis after being given the diagnosis estimated probability for each of the 260 images by the CAIDS. The pathological results for the images were not revealed to any of the 9 urologists, nor were they given the composition of the testing set, and none were involved in other parts of this study.

Visual Explanation

To further understand what and where the CAIDS finds important features, we visualized the predictions for the cystoscopic images through heat maps to confirm whether the information that it uses for classification is reasonable.

Statistical Analysis

All statistical tests were conducted using the R software environment, version 3.6.1. The ROC curves were plotted with the pROC package (21). The PR curves were plotted with the precrec package.

Results

In total, 69 204 images from 10 729 patients were used to develop and validate the CAIDS. For SYSMH and RJH, 63 452 images from

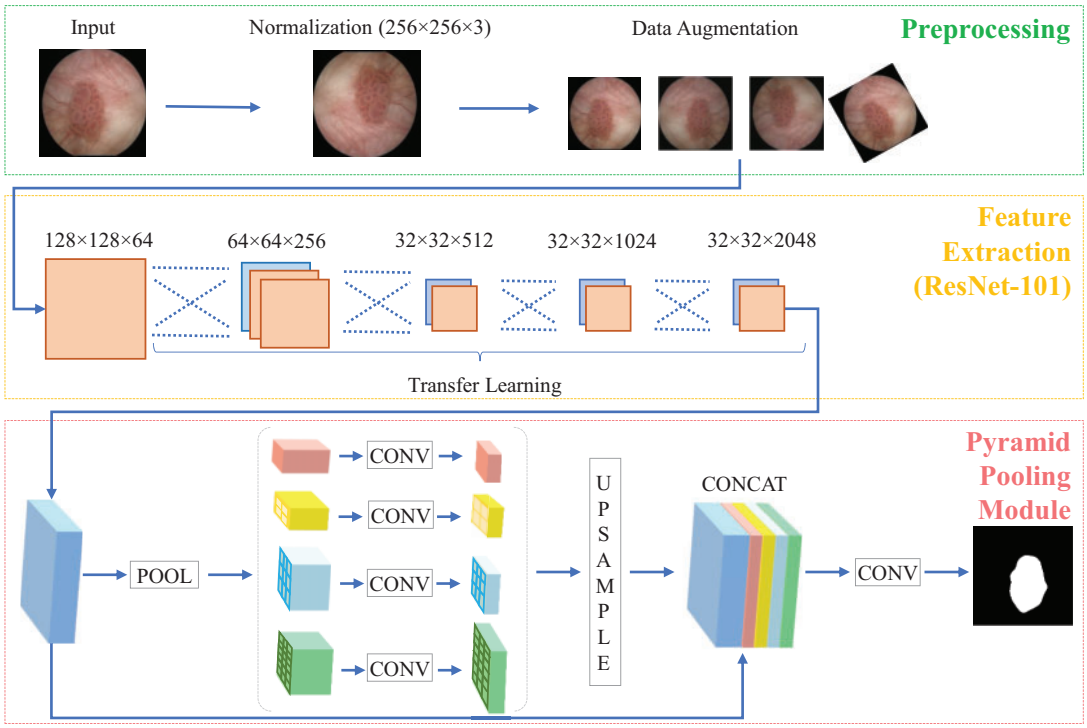

Figure 2. Framework of the Cystoscopy Artificial Intelligence Diagnostic System. An overview of the proposed framework. Images were captured under white light from different angles and were saved in jpg format. Preprocessing: normalizes images to meet the requirements of the model and augments each image into 4 transformed versions. Feature extraction: extracts feature maps through a ResNet-101 model pretrained on ImageNet. Pyramid pooling module: collects more information by concatenating feature maps at different scales in parallel through a 4-level pyramid. CONV = convolutional layer; CONCAT = concat layer.

9302 patients were included. For the external validation sets, 48 tumor images and 639 control images from 211 patients were captured at SZSH, 175 tumor images and 1364 control images from 472 patients from STCH were included, 96 tumor images and 349 control images from 109 patients were captured at AMUFH, and 328 tumor images and 2753 control images from 635 patients were obtained at NJFH. Every positive image had confirmed pathological results, and every negative image was confirmed by biopsy or 6-months follow-up. Detailed staging information for each dataset is listed in Table 1.

After 50 epochs of iteration, the CAIDS was developed. There was a high degree of agreement between the region predicted by the CAIDS in segmentation and the region labeled by the urologists, with a satisfactory median intersection over union of 0.770 (interquartile range = 0.361-0.903) in the internal validation set (Supplementary Figure 1, available online). Several example cases are shown in Supplementary Figure 2 (available online). The accuracy was 0.977 (95% CI = 0.974 to 0.979). The sensitivity, specificity, and NPV were all higher than 0.975. The PPV was 0.819 (95% CI = 0.799 to 0.838), and the most common causes of false-positives were elevation of the bladder wall and inflammation.

The CAIDS also reached a high level of accuracy for all external validation sets. Specifically, the diagnostic accuracies were 0.990 (95% CI = 0.979 to 0.996) for SZSH, 0.982 (95% CI = 0.974 to 0.988) for STCH, 0.978 (95% CI = 0.959 to 0.989) for AMUFH, and 0.991 (95% CI = 0.987 to 0.994) for NJFH.

Near-ideal areas under the ROC curves (AUCs) were also observed in all validation sets, ranging from 0.989 (95% CI = 0.988 to 0.991) to 0.998 (95% CI = 0.996 to 1.000) (Figure 3, A). The sensitivity, specificity, and NPV of the CAIDS were all higher than 0.95 in all validation sets, with the exception of the sensitivity for SZSH (0.875, 95% CI = 0.748 to 0.953). The PPV ranged from 0.819 (95% CI = 0.799 to 0.838) to 0.977 (95% CI = 0.877 to 0.999) (Table 2). In our dataset, negative cases were more numerous than positive cases. To address this imbalance, we also plotted the PR curves to evaluate the performance of the CAIDS. In the external validation sets, the areas under the PR curves ranged from 0.949 (95% CI = 0.919 to 0.993) to 0.986 (95% CI = 0.980 to 1.000) (Figure 3, B), supporting the potential of the CAIDS to assist urologists in decreasing the misdiagnosis rate during cystoscopy. The F1 scores were 0.923 (95% CI = 0.873 to 0.989), 0.924 (95% CI = 0.899 to 0.955), 0.948 (95% CI = 0.920 to 0.983), and 0.957 (95% CI = 0.942 to 0.974) for SZSH, STCH, AMUFH, and NJFH, respectively.

In the CAIDS vs urologists comparison (n = 260), the sensitivities of the trainees, competent urologists, and expert urologists were 0.508 (95% CI = 0.419 to 0.596), 0.562 (95% CI = 0.472 to 0.648), and 0.754 (95% CI = 0.671 to 0.825), respectively. The sensitivity of the CAIDS was 0.954 (95% CI = 0.902 to 0.983), much higher than any group of urologist. The AUC was 0.963 (95% CI = 0.940 to 0.987), and the areas under the PR curve was 0.960 (95% CI = 0.935 to 0.996), indicating that the CAIDS also performs well on complex cases (Figure 3, C and D). The F1 score on these images was 0.939 (95% CI = 0.917 to 0.976). Then, we assessed the performance of the urologists with CAIDS guidance. The sensitivities of the trainee, competent, and expert urologists improved to 0.731 (95% CI = 0.646 to 0.805), 0.846 (95% CI = 0.772 to 0.903), and 0.908 (95% CI = 0.844 to 0.951), respectively (Figure 3, E, and Table 2). On average, it took the trainee, competent, and expert urologists 45 minutes, 38 minutes, and 35 minutes, respectively, to evaluate the 260 images, whereas the CAIDS

Table 1. Baseline clinicopathological characteristics of the enrolled patients[a]

Characteristic	SYSMH (n = 3820)		RJH (n = 5482)		SZSH (n = 211)		STCH (n = 472)		AMUFH (n = 109)		NJFH (n = 635)	
	Control	Tumor	Control	Tumor	Control	Tumor	Control	Tumor	Control	Tumor	Control	Tumor
Total No.	3161	659	4605	877	204	7	436	36	62	47	527	108
Age, median (IQR), y	57 (45-66)	64 (57-71)	65 (58-72)	66 (59-74)	44 (35-52)	61.5 (42-65)	57 (48-64)	68 (62-72)	62 (48-67)	61.5 (52-70)	64 (53-73)	65 (57-76)
Sex, No. (%)												
Male	1925 (60.9)	557 (84.5)	3615 (78.5)	708 (80.7)	144 (70.2)	6 (85.7)	273 (62.6)	30 (83.3)	48 (77.4)	37 (78.7)	398 (75.5)	89 (82.4)
Female	1236 (39.1)	102 (15.5)	990 (21.5)	169 (19.3)	60 (29.8)	1 (14.3)	163 (37.4)	6 (13.7)	14 (22.6)	10 (21.3)	129 (24.5)	19 (17.6)
T stage, No. (%)												
Tis		43 (6.5)		81 (9.2)		0		1 (2.9)		2 (4.3)		7 (6.5)
Ta		253 (38.4)		378 (43.1)		2 (28.6)		26 (72.2)		20 (42.6)		47 (43.5)
T1		173 (26.3)		331 (37.7)		4 (57.1)		7 (19.4)		14 (29.8)		23 (21.3)
T2		94 (14.3)		63 (7.2)		1 (14.3)		2 (5.5)		8 (17.0)		29 (26.9)
T3		72 (10.9)		18 (2.1)		0		0		3 (6.3)		2 (1.8)
T4		24 (3.6)		6 (0.7)		0		0		0		0
Grade, No. (%)												
Low grade		284 (43.1)		442 (50.4)		6 (85.7)		26 (72.2)		28 (59.6)		62 (57.4)
High grade		375 (56.9)		435 (49.6)		1 (14.3)		10 (27.8)		19 (40.4)		46 (42.6)

[a]AMUFH = The First Hospital Affiliated to Army Medical University; IQR = interquartile range; NJFH = The First Affiliated Hospital of Nanjing Medical University; RJH = Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine; STCH = Shantou Central Hospital; SYSMH = Sun Yat-sen University Memorial Hospital; SZSH = Shenzhen Second People's Hospital.

needed only 12 seconds to produce more accurate diagnoses, suggesting that the CAIDS can efficiently reduce the operation time for cystoscopy while improving diagnostic performance by reducing misdiagnosis.

Additionally, we performed a subgroup analysis to assess the performance of CAIDS on carcinomas of different T stages (Supplementary Figure 1, available online). To better understand what the CAIDS finds important, we visualized the predictions of the model through heat maps (Supplementary Figure 3, available online).

To further validate the CAIDS, we collected an independent consecutive cohort of patients who underwent TURBT in SYSMH from January 1, 2021, to March 1, 2021. A total of 80 patients (468 positive images, 222 negative images) were enrolled. The sensitivity was 0.996 (95% CI = 0.958 to 1.000); the specificity was 0.770 (95% CI = 0.709 to 0.824); the PPV was 0.901 (95% CI = 0.872 to 0.926), and the NPV was 0.988 (95% CI = 0.959 to 0.999). The AUC was 0.976 (95% CI = 0.963 to 0.989). The AU PRC was 0.985(95% CI = 0.976 to 0.997). F1 score was 0.946 (95% CI = 0.932 to 0.962). The CAIDS performs well in this independent validation set, supporting that the CAIDS has great potential to be applied into clinical practice.

Discussion

In this study, using 69 204 cystoscopic images from 10 729 patients, we developed a cystoscopic AI-based diagnostic system known as the CAIDS, the use of which made valuable contributions to the diagnosis process.

To some degree, cystoscopic analysis depends on the skill and experience of urologists. TURBt by less experienced urologists predicted a higher early recurrence rate (22). Trained on a large number of cystoscopic images labeled by expert urologists, our CAIDS shows promising diagnostic performance in identifying BCa. Thus, the CAIDS is a practical means of bridging the diagnosis gap between national hospitals and primary care hospitals and the gap between experienced urologists and less experienced urologists.

Moreover, even for expert urologists, the limitations of cystoscopy in identifying flat cancerous tissue (CIS) and very small tumors must be considered (23). In the subgroup analysis for complex lesions in our study (n = 260), the performance of the CAIDS was superior to that of expert urologists. Of tumor lesions, 24.6% were misdiagnosed by the expert urologists, but when the CAIDS was applied, 62.5% of misdiagnosed cases could be avoided. Therefore, the CAIDS has the potential to improve the diagnostic accuracy of detecting flat cancerous tissue (CIS) and very small tumors.

There are several reasons why the CAIDS was able to outperform human urologists in this study. First, our system is based on 69 204 cystoscopic images from 10 729 patients. The larger the training dataset is, the more accurate the system will be. Moreover, we used pathological proof as the gold standard to overcome the limitations of human expert opinions, especially for complex lesions.

In our study, the NPV ranged from 0.989 to 0.999, and the PPV ranged from 0.819 to 0.977. The high NPV indicates that the CAIDS has the potential to achieve a low negative misdiagnosis rate, and the acceptable PPV suggests that the use of the CAIDS would not result in a large increase in the number of unnecessary biopsies performed to avoid a positive misdiagnosis. Notably, for most diagnostic models, the PPV and NPV based on the general population are more suitable for evaluation.

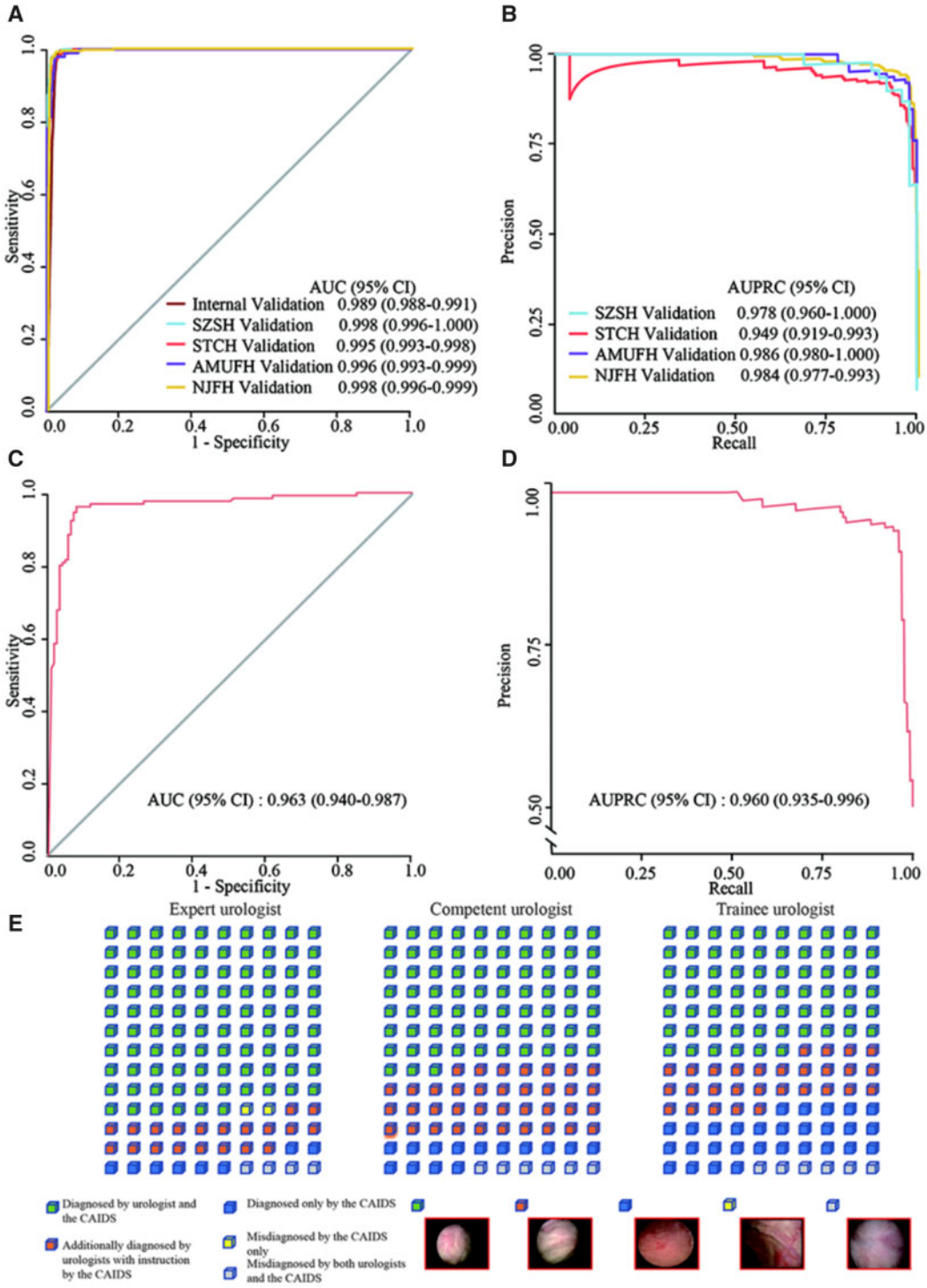

Figure 3. Diagnostic performance of the CAIDS in different validation sets. The ROC curves were generated by plotting sensitivity against specificity. The PR curves were generated by plotting recall (also known as sensitivity) against precision (also known as the positive predictive rate). **A)** ROC curves and their 95% confidence intervals (CIs) in the validation sets; **B)** PR curves and their 95% CIs in the validation sets; **C)** ROC curve and its 95% CIs in the validation set of complex cases; **D)** PR curve and its 95% CIs in the validation set of complex cases; **E)** performance of the CAIDS compared with urologists with different levels of experience (expert urologists, competent urologists, and trainees; with and without the guidance of the CAIDS) in identifying BCa in a subgroup analysis of complex lesions. AMUFH = The First Hospital Affiliated to Army Medical University; AUC = area under the ROC curve; AUPRC = area under the PR curve; BCa = bladder cancer; CAIDS = Cystoscopy Artificial Intelligence Diagnostic System; NJFH = The First Affiliated Hospital of Nanjing Medical University; PR = precision recall curve; RJH = Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine; ROC = receiver operating characteristic; STCH = Shantou Central Hospital; SYSMH = Sun Yat-sen University Memorial Hospital; SYSMH = Sun Yat-sen University Memorial Hospital; SZSH = Shenzhen Second People's Hospital.

Table 2. Performance of the CAIDS in different validation datasets

Dataset	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
In internal and external validation sets[a]					
SYSMH and RJH	0.977 (0.974 to 0.979)	0.987 (0.979 to 0.993)	0.975 (0.972 to 0.978)	0.819 (0.799 to 0.838)	0.999 (0.998 to 0.999)
SZSH	0.990 (0.979 to 0.996)	0.875 (0.748 to 0.953)	0.998 (0.991 to 1.000)	0.977 (0.877 to 0.999)	0.991 (0.980 to 0.997)
STCH	0.982 (0.974 to 0.988)	0.977 (0.943 to 0.994)	0.982 (0.974 to 0.989)	0.877 (0.822 to 0.920)	0.997 (0.992 to 0.999)
AMUFH	0.978 (0.959 to 0.989)	0.958 (0.897 to 0.989)	0.983 (0.963 to 0.994)	0.939 (0.871 to 0.977)	0.989 (0.971 to 0.997)
NJFH	0.991 (0.987 to 0.994)	0.951 (0.922 to 0.972)	0.996 (0.992 to 0.998)	0.963 (0.936 to 0.981)	0.999 (0.996 to 1.000)
In a select subset of complex cases[b]					
CAIDS	0.939 (0.902 to 0.964)	0.954 (0.902 to 0.983)	0.923 (0.863 to 0.963)	0.925 (0.867 to 0.960)	0.952 (0.899 to 0.982)
Expert	0.842 (0.792 to 0.884)	0.754 (0.671 to 0.825)	0.931 (0.873 to 0.968)	0.916 (0.846 to 0.961)	0.791 (0.718 to 0.852)
Competent	0.692 (0.632 to 0.748)	0.562 (0.472 to 0.648)	0.823 (0.747 to 0.884)	0.760 (0.663 to 0.842)	0.652 (0.574 to 0.725)
Trainee	0.608 (0.546 to 0.667)	0.508 (0.419 to 0.596)	0.708 (0.622 to 0.784)	0.635 (0.535 to 0.727)	0.590 (0.508 to 0.668)
Expert + CAIDS	0.927 (0.888 to 0.955)	0.908 (0.844 to 0.951)	0.946 (0.892 to 0.978)	0.944 (0.888 to 0.977)	0.911 (0.850 to 0.953)
Competent + CAIDS	0.861 (0.814 to 0.901)	0.846 (0.772 to 0.903)	0.877 (0.808 to 0.928)	0.873 (0.802 to 0.923)	0.851 (0.779 to 0.906)
Trainee + CAIDS	0.819 (0.767 to 0.876)	0.731 (0.646 to 0.805)	0.831 (0.755 to 0.891)	0.812 (0.729 to 0.878)	0.755 (0.676 to 0.823)

[a]Internal validation sets included SYSMH and RJH; external validation sets included SZSH, STCH, AMUFH and NJFH. AMUFH = The First Hospital Affiliated to Army Medical University; CAIDS = Cystoscopy Artificial Intelligence Diagnostic System; CI = confidence interval; NJFH = The First Affiliated Hospital of Nanjing Medical University; NPV = negative predictive value; PPV = positive predictive value; RJH = Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine; STCH = Shantou Central Hospital; SYSMH = Sun Yat-sen University Memorial Hospital; SZSH = Shenzhen Second People's Hospital.
[b]The select subset of complex cases consisted of 130 carcinomas in situs and small tumor images vs 130 inflammation images.

However, cystoscopy is an invasive examination and is conducted only for those who are with a strong suspicion to have BCa rather than as a screening examination. Therefore, the PPV and NPV based on the population requiring cystoscopic examination are more suitable to evaluate the system presented in this study.

Several previous studies have investigated the application of AI to improve the performance of cystoscopy in detecting BCa. Shkolyar et al. (17) developed an AI system using videos of 95 patients. The per-frame sensitivity and specificity were 90.9% and 98.6%, respectively. Ikeda et al. (24) reported another system based on 2102 cystoscopic images; the sensitivity of this system was 89.7%, and the specificity was 94.0%. However, the applications of these systems are limited because of the small sample sizes, the incomplete inclusion of tumor stages, and the relatively low sensitivities. Compared with the systems developed in these previous studies, our system is based on a much larger dataset, has been validated on data from multiple centers, and includes all tumor stages. The accuracy of our system is also higher than that of the existing systems, with a pooled sensitivity of 97.5% and a pooled specificity of 98.3% on the validation sets. Moreover, with a latency of less than 50 milliseconds per image, our system is much more efficient.

In fact, we have implemented our system in practice and deployed it in the operating room at SYSMH. The system can capture the video stream from a cystoscope through various types of video data cables, such as DVI and S-Video, which output video frames in real time and are supported by modern cystoscopic devices. Each video frame is converted into a normalized image and fed to the proposed framework. Then, a mask of the lesion region is predicted by the model. The original frame is overlapped with the mask to form a new frame, which is displayed on the monitor of our system. Thanks to the high performance of graphics processing units, our system is capable of processing live images generated by a cystoscope in real time, which can be processed by the CAIDS for real-time diagnosis.

Prospective studies will be needed to integrate the CAIDS into clinical practice. It is challenging to incorporate AI into widespread clinical use because of safety concerns. In addition, any technology must have minimal impact on the urologist's workflow, and its onscreen presence must be nondistracting. The short latency of the CAIDS (<50 milliseconds for per-image assessment) makes it efficient and practical.

Our study had some limitations. First, the CAIDS was developed using a large Chinese cohort dataset; thus, its performance in other populations remains unclear. Second, prospective studies need to be conducted to test the clinical applicability of the CAIDS.

Using 69 204 cystoscopic images collected from 10 729 patients at 6 hospitals of different tiers, we developed and tested a cystoscopic AI-based diagnostic system known as the CAIDS. It achieved near-perfect accuracy, sensitivity, and specificity in all internal and external validation sets. For the identification of complex bladder carcinomas such as flat cancerous tissue (CIS) and very small tumors (n = 260), the CAIDS performed better than expert urologists. The CAIDS has the potential to assist urologists in clinical practice to improve the diagnosis of BCa during cystoscopy and the resection rate during TURBt, thus reducing misdiagnosis and the early recurrence rate.

Funding

This work was supported by the Science and Technology Planning Project of Guangdong Province (2018B010109006), the National Natural Science Foundation of China (82003151, 81770967 and 81822010), Medical Research Funding of Guangdong Province (A2019101), the Key Areas Research and Development Program of Guangdong (810229511112), the National Key R&D Program of China (2018YFA0902803 and 2018YFC0116500), the Science and Technology Planning Project of Guangdong Province (2017B020227007), Guangdong Science and Technology Department (2020B1212060018), and the Project Supported by Guangdong Province Higher Vocational Colleges & Schools Pearl River Scholar Funded Scheme (for TXL).

ARTICLE

Notes

Role of the funders: The funding sources had no role in the study design; data collection, analysis, or interpretation; writing of the article; or decision to submit it for publication.

Disclosures: None.

Author contributions: SX-W, X-C, HG-C, J-H, and TX-L played roles in conceptualization, data curation, methodology. SX-W and X-C prepared the original draft. All authors reviewed and edited the report. HG-C, J-H, and TX-L played roles in funding acquisition and project administration. G-Q and H-C played roles in methodology and developed the software. All authors participated in the formal analysis and investigation.

Data Availability

All data used or analyzed in the study are available from the corresponding author upon reasonable request.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, 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(6):394–424.
2. Burger M, Catto JW, Dalbagni G, et al. Epidemiology and risk factors of urothelial bladder cancer. *Eur Urol*. 2013;63(2):234–241.
3. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013;49(6):1374–1403.
4. Babjuk M, Burger M, Comperat EM, et al. European association of urology guidelines on non-muscle-invasive bladder cancer (TaT1 and carcinoma in situ)–2019 update. *Eur Urol*. 2019;76(5):639–657.
5. Brausi M, Collette L, Kurth K, et al.; for the EORTC Genito-Urinary Tract Cancer Collaborative Group. Variability in the recurrence rate at first follow-up cystoscopy after TUR in stage Ta T1 transitional cell carcinoma of the bladder: a combined analysis of seven EORTC studies. *Eur Urol*. 2002;41(5):523–531.
6. Mowatt G, N'Dow J, Vale L, et al.; for the Aberdeen Technology Assessment Review (TAR) Group. Photodynamic diagnosis of bladder cancer compared with white light cystoscopy: systematic review and meta-analysis. *Int J Technol Assess Health Care*. 2011;27(1):3–10.
7. Isfoss BL. The sensitivity of fluorescent-light cystoscopy for the detection of carcinoma in situ (CIS) of the bladder: a meta-analysis with comments on gold standard. *BJU Int*. 2011;108(11):1703–1707.
8. Mariappan P, Johnston A, Padovani L, et al.; for members of the Scottish Bladder Cancer QPI Research Collaborative. Enhanced quality and effectiveness of transurethral resection of bladder tumour in non-muscle-invasive bladder cancer: a multicentre real-world experience from Scotland's Quality Performance Indicators programme. *Eur Urol*. 2020;78(4):520–530.
9. Cumberbatch MGK, Foerster B, Catto JWF, et al. Repeat transurethral resection in non-muscle-invasive bladder cancer: a systematic review. *Eur Urol*. 2018;73(6):925–933.
10. Witjes JA. Fluorescence cystoscopy: the end of biopsies for carcinoma in situ detection? *BJU Int*. 2015;116(1):2–3.
11. Drejer D, Béji S, Munk Nielsen A, Høyer S, Wrist Lam G, Jensen JB. Clinical relevance of narrow-band imaging in flexible cystoscopy: the DaBlaCa-7 study. *Scand J Urol*. 2017;51(2):120–123.
12. Li X, Zhang S, Zhang Q, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet Oncol*. 2019;20(2):193–201.
13. Huynh E, Hosny A, Guthier C, et al. Artificial intelligence in radiation oncology. *Nat Rev Clin Oncol*. 2020;17(12):771–781.
14. Trister AD, Buist DSM, Lee CI. Will machine learning tip the balance in breast cancer screening? *JAMA Oncol*. 2017;3(11):1463–1464.
15. Luo H, Xu G, Li C, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol*. 2019;20(12):1645–1654.
16. Kriegmair MC, Rother J, Grychtol B, et al. Multiparametric cystoscopy for detection of bladder cancer using real-time multispectral imaging. *Eur Urol*. 2020;77(2):251–259.
17. Shkolyar E, Jia X, Chang TC, et al. Augmented bladder tumor detection using deep learning. *Eur Urol*. 2019;76(6):714–718.
18. Wittekind C, Oberschmid B. TNM classification of malignant tumors 2010: general aspects and amendments in the general section. *Pathologe*. 2010;31(5):333–334, 336–338.
19. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO classification of tumours of the urinary system and male genital organs-part B: prostate and bladder tumours. *Eur Urol* . 2016;70(1):106–119.
20. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE; 2017:6230–6239.
21. Grau J, Grosse I, Keilwagen J. PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics*. 2015;31(15):2595–2597.
22. Mariappan P, Zachou A, Grigor KM; for the Edinburgh Uro-Oncology Group. Detrusor muscle in the first, apparently complete transurethral resection of bladder tumour specimen is a surrogate marker of resection quality, predicts risk of early recurrence, and is dependent on operator experience. *Eur Urol*. 2010;57(5):843–849.
23. Witjes JA, Redorta JP, Jacqmin D, et al. Hexaminolevulinate-guided fluorescence cystoscopy in the diagnosis and follow-up of patients with non-muscle-invasive bladder cancer: review of the evidence and recommendations. *Eur Urol*. 2010;57(4):607–614.
24. Ikeda A, Nosato H, Kochi Y, et al. Support system of cystoscopic diagnosis for bladder cancer based on artificial intelligence. *J Endourol*. 2020;34(3):352–358.