Interobserver Agreement of Confocal Laser Endomicroscopy for Bladder Cancer

Authors: Timothy C. Chang, M.S.,¹ Jen-Jane Liu, M.D.,¹ Shelly T. Hsiao, B.A.,² Ying Pan, Ph.D.,¹ Kathleen E. Mach, Ph.D.,¹ John T. Leppert, M.D.,¹,² Jesse K. McKenney, M.D.,³ Robert V. Rouse, M.D.,²,³ and Joseph C. Liao, M.D.¹,²

Institution(s)/affiliation(s) for each author:

¹ Department of Urology, Stanford University School of Medicine, Stanford, CA 94305
² Veterans Affairs Palo Alto Health Care System, Palo Alto, CA 94304
³ Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305

Corresponding author:

Joseph C. Liao, M.D.
Department of Urology, Stanford University School of Medicine
300 Pasteur Dr., Room S-287
Stanford, CA 94305-5118
jliao@stanford.edu
p: (650) 852-3284
f: (650) 849-1925

Timothy C. Chang – tcchang@stanford.edu
Jen-Jane Liu – jenliu@stanford.edu
Shelly T. Hsiao – shsiao27@gmail.com
Ying Pan – yingpan@stanford.edu
Kathleen E. Mach – kmach@stanford.edu
John T. Leppert – jleppert@stanford.edu
Jesse K. McKenney – mckennj@ccf.org
Robert Vance Rouse – rouse@stanford.edu
Joseph C. Liao – jliao@stanford.edu

Previous abstract:
Abstract:

Purpose: Emerging optical imaging technologies such as confocal laser endomicroscopy (CLE) hold promise in improving bladder cancer diagnosis. The purpose of this study was to determine the interobserver agreement of image interpretation using CLE for bladder cancer.

Methods: Experienced CLE urologists (n=2), novice CLE urologists (n=6), pathologists (n=4) and nonclinical researchers (n=5) were recruited to participate in a two-hour computer-based training consisting of a teaching and validation set of intraoperative white light cystoscopy (WLC) and CLE video sequences from patients undergoing transurethral resection of bladder tumor. Interobserver agreement was determined using the κ statistic.

Results: Of the 31 bladder regions analyzed, 19 were cancer and 12 were benign. For cancer diagnosis, experienced CLE urologists had substantial agreement for both CLE and WLC+CLE (90%, κ 0.80) compared with moderate agreement for WLC alone (74%, κ 0.46), while novice CLE urologists had moderate agreement for CLE (77%, κ 0.55), WLC (78%, κ 0.54) and WLC+CLE (80%, κ 0.59). Pathologists had substantial agreement for CLE (81%, κ 0.61), and nonclinical researchers had moderate agreement (77%, κ 0.49) in cancer diagnosis. For cancer grading, experienced CLE urologists had fair to moderate agreement for CLE (68%, κ 0.64), WLC (74%, κ 0.67) and WLC+CLE (53%, κ 0.33), as did novice CLE urologists for CLE (53%, κ 0.39), WLC (66%, κ 0.50) and WLC+CLE (61%, κ 0.49). Pathologists (65%, κ 0.55) and nonclinical researchers (61%, κ 0.56) both had moderate agreement for CLE in cancer grading.

Conclusions: CLE is an adoptable technology for cancer diagnosis in novice CLE observers following a short training with moderate interobserver agreement and diagnostic accuracy similar to WLC alone. Experienced CLE observers may be capable of achieving substantial levels of agreement for cancer diagnosis that is higher than with WLC alone.
Introduction:

Probe-based confocal laser endomicroscopy (CLE) is an emerging optical imaging technology that enables endoscopic microscopy of mucosal lesions, with dynamic, subsurface imaging of tissue microarchitecture and cellular features. The technology has been applied as an adjunct to standard white light endoscopy in the respiratory\(^1\) and gastrointestinal tracts,\(^2\)–\(^4\) and recently in the urinary tract.\(^5\)–\(^8\) Particularly for bladder cancer, a new generation of imaging technologies such as CLE may augment the diagnostic accuracy of white light cystoscopy (WLC) through improved visualization of flat lesions, differentiation of benign from neoplastic tumors, and delineation of the tumor boundaries.\(^9\)

Based to the well-established principle of confocal microscopy\(^10,11\), CLE is performed using a fiberoptic, sterilizable probe that fits within the working channel of a standard cystoscope. Optical sectioning of the tissue of interest with micron-scale resolution is achieved using a 488 nm laser as the light source and fluorescein, a FDA-approved drug that may be administered intravesically or intravenously, as the contrast agent. Previously, we demonstrated the feasibility of using CLE to obtain real-time \textit{in vivo} images of bladder tumors\(^6\) and developed diagnostic criteria for grading bladder tumors using CLE.\(^5\)

Interobserver agreement studies are useful in determining the subjective variation in interpretation amongst observers in analyzing images.\(^12\) Methods that demonstrate higher levels of agreement between observers are deemed more reliable.\(^12\) In disciplines that require subjective interpretation of diagnostic imaging such as pathology\(^13,14\) and radiology,\(^15\) interobserver studies are commonly applied to determine reproducibility of the results from one observer to another. Interobserver agreement studies for CLE have been studied in the
gastrointestinal tract, which have ranged from “moderate” to “good” agreement in the diagnosis of colorectal cancer to “substantial” and “almost perfect” agreement in Barrett’s esophagus. Interobserver agreement of CLE in the urinary tract has not been previously examined, which can assess the reliability of the technology between observers and evaluate the adoptability of the CLE in novice users. The aim of this study was to determine the interobserver agreement and diagnostic accuracy of CLE for bladder cancer.

Methods:

Observer Recruitment

The study was approved by the Stanford University Institutional Review Board and the VA Palo Alto Health Care System (VAPAHCS) Research and Development Committee. The study consisted of four observer groups. The first was an experienced CLE urologist group consisting of a board certified urologist and a urology chief resident involved in the protocol development of CLE for the urinary tract. The remaining three groups were novice CLE observers with no prior experience with CLE, consisting of six board certified urologists, four pathologists, and five nonclinical researchers. The novice CLE urologists were expert WLC users, but the pathologists and nonclinical researchers had no experience with WLC. The nonclinical researchers ranged from an undergraduate student to Ph.D. scientists and engineers. All groups participated in identical training sessions.

Teaching and Validation Set

The observers participated in a two-hour, computer-based training that consisted of separate teaching and validation sets that featured intraoperative WLC and CLE videos from selected patients undergoing transurethral resection of bladder tumor at the VAPAHCS from 2008-2011.
The computer-based, interactive training was created in a website format compatible with common internet browsers for easy access and future scalability. The observers were first introduced to background information on bladder cancer and CLE technology (Figure 1A). Using still images and video sequences, the observers were instructed to identify three microarchitectural features (flat vs. papillary, tissue organization, and vascularity) and three cellular features (morphology, cohesiveness, and borders) of benign and pathological urothelium using CLE. Diagnostic criteria (Table 1) were developed that associated the six features to benign and cancerous urothelium, and the observers were instructed to categorize each sequence as benign, low grade (LG) or high grade (HG) cancer. Fifteen CLE video sequences were provided in a teaching set for the observers to iteratively practice diagnosing the sequences (Figure 1B).

Following the teaching set, the observers proceeded to the validation set to evaluate 32 CLE video sequences consisting of 12 benign, 9 LG, 11 HG images (Figure 1C). The benign sequences were from biopsy-confirmed normal mucosa, inflammation, and papilloma, LG sequences were from LG papillary tumors, and HG sequences were from HG papillary tumors and carcinoma-in-situ. The observers were able to pause and review each video clip frame-by-frame. All observers were blinded to patient history and final pathology and were asked to diagnose and grade each clip as benign, LG or HG. Upon completing the 32 CLE sequences, the four pathologists and five nonclinical researchers concluded the study. The experienced CLE urologists and novice CLE urologists were asked to further diagnose an additional set of 32 corresponding WLC images in different order, followed by a third and final set of 32 images where the CLE and WLC were shown together. The data were gathered to determine the interobserver agreement and diagnostic accuracy.
From the 32 sequences reviewed by the observers, one of the original HG sequences was excluded due to a discrepancy in the final pathologic diagnosis. The resulting 31 responses of benign, LG or HG from each of the observers for the 31 images with correlating histopathological information were used to generate the interobserver agreement and diagnostic accuracy data analyses. For cancer diagnosis, all 31 responses were used. However, for cancer grading, a subset of the 19 cancer (9 LG and 10 HG) sequences found on histopathology were used.

In addition to diagnosing and grading each of the sequences as the other groups did, the experienced CLE urologist group was asked to also identify the six key CLE features (Table 1) for each sequence.

**Statistical Analysis**

Interobserver agreement was assessed using the Fleiss κ statistic. The description for κ statistic developed by Landis and Koch with 0.00-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial and 0.81-1.00 as almost perfect levels of agreement was used in this study.  

To determine diagnostic accuracy, a dedicated pathologist (R.V.R.) who was blinded to the clinical history and not included in the pathologist group reviewed all the histopathological slides corresponding to each bladder lesion. Sensitivity and specificity was calculated using the histopathological results as the standard.

**Results:**

Table 2 shows the percent agreement and κ statistic for the two experienced CLE urologists for each of the six features. Tissue organization, vascular features, cellular morphology and
cellular borders had substantial levels of agreement. The experienced CLE urologists had moderate agreement in the ability to determine flat vs. papillary using CLE, while they had a fair level of agreement in characterizing cellular cohesiveness.

Table 3 shows the interobserver agreement and diagnostic accuracy for cancer diagnosis. Experienced CLE urologists had substantial to almost perfect levels of agreement for CLE and WLC+CLE, which were both greater than WLC alone. The novice CLE urologists had moderate levels of agreement for all groups. The sensitivity and specificity for all groups were between 73 and 89%.

Table 4 shows the interobserver agreement and diagnostic accuracy for cancer grading. All groups had decreased levels of agreement compared with cancer diagnosis except for nonclinical researchers who had a slightly increased κ compared with cancer diagnosis (0.56 vs. 0.49). There was also an unexpected decrease in κ for experienced CLE urologists from WLC to WLC+CLE (0.67 to 0.33). For low grade cancer, experienced CLE urologists had increased sensitivity and specificity with the addition of CLE to WLC (sensitivity 64%, specificity 81%) compared with standard WLC alone (sensitivity 55%, specificity 50%). Novice CLE urologists showed an increase in specificity with the addition of CLE to WLC as well (specificity 65%) compared to WLC alone (specificity 54%). For low grade, the novice groups all had similar diagnostic accuracy using CLE alone (sensitivity 49-52%, specificity 66-75%). For high grade cancer, experienced CLE urologists showed an increase in sensitivity for WLC+CLE (sensitivity 69%) compared with WLC (sensitivity 50%) while maintaining specificity (specificity 73% for both) with the addition of CLE to standard WLC. For novice CLE urologists, there was an increase in sensitivity but a decrease in specificity with the addition of CLE to standard WLC.

Discussion:
Our single-center study indicates that CLE image interpretation of bladder cancer is adoptable by novice observers through a two-hour computer-based interactive training. We created a training session centered on guiding observers to identify six key CLE features in analyzing bladder cancer microarchitecture and cellular morphology. There were substantial levels of agreement for four of the six features, suggesting that these features can be identified reliably with proper training. The moderate level of agreement for the flat vs. papillary feature was not unexpected as it is a feature more easily identified on a macroscopic level by WLC than on a microarchitectural level. The cohesiveness feature had a fair level of agreement, suggesting that further refinement of the criteria may be necessary to improve the reliability of this feature.

Once the observers were trained to identify these features, they were introduced to a table correlating the six features to bladder cancer diagnoses. This table was developed based on the 2004 WHO Classification of Bladder Tumours, and further refinement of the table is expected as the technology matures and is adopted by additional end users. Since most of the observer groups with the exception of the pathologist group do not diagnose bladder cancer routinely using microscopy, the table served as a useful guide for novice observers, particularly nonclinical researchers, to diagnose bladder cancer.

The interobserver agreement and diagnostic accuracy results were reported separately for CLE, WLC, and WLC+CLE. Data on CLE alone was useful in assessing the technology itself and observing its standalone performance without the bias of WLC. The WLC information provided baseline data as it is the standard imaging modality for bladder cancer. However, the practical use of CLE in the clinical setting would involve WLC for the initial survey of the bladder and guiding of the CLE probe to the area of interest. Thus, the WLC+CLE data provides the most clinically relevant information. As pathologists and nonclinical researchers had no prior training in WLC, they were only asked to review CLE alone.
The ability of novice observers to learn CLE image interpretation was demonstrated by the performance of the various groups after a two-hour training session. For cancer diagnosis (Table 3), novice CLE urologists had similar moderate levels of agreement and diagnostic accuracy for CLE, WLC and WLC+CLE, indicating that the two-hour training was sufficient for CLE to reliably diagnose cancer with diagnostic accuracy comparable to standard WLC. Interestingly, nonclinical researchers who had no prior CLE or clinical experience also had a moderate level of agreement for cancer diagnosis, similar to the other novice CLE observer groups, while also maintaining similar levels of diagnostic accuracy. Thus, both the novice CLE urologists demonstrated comparable agreement for WLC and CLE, while concurrently, nonclinical researchers with no prior clinical experience demonstrated similar levels of agreement for CLE as other clinically trained novice groups. These results provide evidence on the relative ease in training the novice observers to interpret CLE images. It is notable that these results are in line with the moderate levels of interobserver agreement reported for CLE imaging of colorectal neoplasia.\textsuperscript{16}

Experienced CLE urologists obtained substantial to nearly perfect levels of agreement for CLE and WLC+CLE with $\kappa$ of 0.80 while also maintaining the level of sensitivity and specificity compared to WLC. CLE outperformed WLC for the experienced CLE urologists group with respect to interobserver agreement. The result suggests that for cancer diagnosis, CLE may provide added value compared to WLC since it has greater interobserver reliability without sacrificing diagnostic accuracy. In short, the results indicate that a two-hour training session may enable novice users to achieve similar levels of interobserver reliability and diagnostic accuracy as WLC for cancer diagnosis, while greater reliability is seen for CLE compared with WLC in experienced CLE urologists.
In regards to cancer grading, the interobserver agreement for cancer grading was in the fair to moderate range for CLE, WLC and WLC+CLE, which was generally lower than for cancer diagnosis. No clear patterns were noted on diagnostic accuracy. The results suggest that CLE may not be as reliable for cancer grading when compared with cancer diagnosis. The interobserver agreement for cancer grading using CLE is similar to pathology literature using the 2004 WHO classification system. May et al.\textsuperscript{14} reported $\kappa$ of 0.30-0.52 for interobserver agreement and van Rhijn et al.\textsuperscript{13} reported $\kappa$ of 0.14-0.58 and 0.55-0.81 for interobserver and intraobserver agreement, respectively. Moreover, a comparison of the cancer grading results in our study derived from the electronic medical records (by multiple pathologists) to the results of a single pathologist (R.V.R.) with expertise in bladder cancer showed a $\kappa$ of 0.58. These findings illustrate the inherent challenges of bladder cancer grading with CLE or standard pathology.

Our study has several limitations. First, our interobserver agreement for cancer grading analysis was a subset analysis of the confirmed cancers from the original data set. When reviewing the sequences, the observers were given the option to grade the sequences as benign, LG, or HG, rather than simply LG or HG. The additional choice may have contributed to the overall lower interobserver agreement compared to cancer diagnosis. Nevertheless, the occurrences were few, and as the data analysis reflects a more clinically relevant and practical scenario of clinicians grading an unknown lesion, the study was designed accordingly. Second, there may be selection bias of the CLE video sequences, which were edited offline and chosen non-randomly by a member of our team with the subjective criteria of image quality (fair to good) and roughly equal distribution of benign, LG, and HG lesions. This individual did not participate as an observer for the study. Third, there may be recall bias from the experienced CLE urologists who acquired the original CLE sequences. The use of an independent member from our team to select the images and video sequences mitigates, but does not eliminate, this
potential bias. Fourth, an inherent limitation of using CLE for bladder cancer diagnosis is the reliance on microarchitectural and cellular features, whereas pathology utilizes additionally nuclear morphology (e.g. size, mitotic figures). Nuclear features are not routinely seen under CLE, as fluorescein is used as the contrast agent, which stains the extracellular matrix nonspecifically.\textsuperscript{5,6}

Overall, novice CLE observers demonstrated the ability to use CLE as an adjunct to WLC to diagnose bladder cancer following a brief training. Nonclinical researchers with no clinical training were able to diagnose bladder cancer to a comparable level as clinically trained novice CLE observers, highlighting the adoptability and translatability of the CLE technology to a wide range of novice users. Our results indicate that further studies are warranted that refines the CLE features used in diagnosing and grading bladder cancer as well as multi-center studies that validate the translatability of this study.

Future directions include prospective multi-center studies to investigate the overall clinical utility of CLE, as well as cost-benefit analyses, which will be necessary for widespread adoption of CLE for bladder cancer diagnosis and grading. In addition, CLE, as a microscopic imaging modality, may be combined with other new macroscopic imaging technologies (i.e. photodynamic diagnosis and narrow band imaging) already in clinical use to improve the overall optical diagnosis of bladder cancer.\textsuperscript{9}

**Conclusion:**

CLE is an adoptable technology for novice CLE observers following a training session with moderate interobserver agreement and diagnostic accuracy similar to WLC alone. Experienced CLE observers may be capable of achieving substantial levels of agreement for cancer diagnosis that is markedly higher than with WLC alone. Fair to moderate levels of agreement
are achieved for cancer grading, although literature suggests the variability may in part be attributable to the grading classification system.

Tables and Figures:

FIG 1. Computer-based, interactive training module for confocal laser endomicroscopy of the bladder (A) CLE Training: Observers were trained to identify six key CLE features. (B) Teaching set: Teaching set consisting of fifteen video sequences to iteratively practice diagnosing and grading CLE sequences. (C) Validation set: All observer groups reviewed and diagnosed CLE sequences. The experienced CLE and novice CLE urologist groups continued on to diagnose two additional sets of WLC and WLC+CLE images.

TABLE 1. DIAGNOSIS TABLE.

TABLE 2. INTEROBSERVER AGREEMENT OF FEATURES OBSERVED ON CLE.

TABLE 3. INTEROBSERVER AGREEMENT AND DIAGNOSTIC ACCURACY FOR CANCER DIAGNOSIS.

TABLE 4. INTEROBSERVER AGREEMENT AND DIAGNOSTIC ACCURACY FOR CANCER GRADING.

Acknowledgement:

The authors would like to acknowledge our colleagues who participated in the study (H.G., J.D.B., C.V.C., M.E., E.S., M.S., D.B., R.M., C.Z., H.R., J.H., S.O., and A.S.). We also thank Mauna Kea Technologies for technical support and helpful discussions. This work was supported in part by the US National Institutes of Health (NIH) R01 CA160986 (J.C.L.).

Disclosure Statement:

No competing financial interests exist.

References:


Abbreviations Used

CLE = confocal laser endomicroscopy

HG = high grade

LG = low grade

\( p_a \) = percent agreement

\( \text{Sn} \) = sensitivity

\( \text{Sp} \) = specificity

VAPAHCS = Veterans Affairs Palo Alto Health Care Systems

WLC = white light cystoscopy
<table>
<thead>
<tr>
<th>Architectural</th>
<th>Benign</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat vs. papillary</td>
<td>Flat</td>
<td>Papillary</td>
</tr>
<tr>
<td>Organization</td>
<td>Organized, normal thickness</td>
<td>Loose cells in LP</td>
</tr>
<tr>
<td>Vascular features</td>
<td>Capillary network in LP</td>
<td>Fibrovascular stalk</td>
</tr>
<tr>
<td></td>
<td>n/a</td>
<td>Tortuous vessels in</td>
</tr>
<tr>
<td>Cellular</td>
<td>Monomorphic</td>
<td>Monomorphic</td>
</tr>
<tr>
<td></td>
<td>Cohesive</td>
<td>Cohesive</td>
</tr>
<tr>
<td></td>
<td>Distinct</td>
<td>Distinct</td>
</tr>
</tbody>
</table>

TABLE 1. DIAGNOSIS TABLE.
### Table 2. Interobserver Agreement of Features Observed on CLE

<table>
<thead>
<tr>
<th></th>
<th>$p_a$</th>
<th>$\kappa$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microarchitectural</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flat vs. papillary</td>
<td>77%</td>
<td>0.54 (0.19 – 0.89)</td>
</tr>
<tr>
<td>Tissue organization</td>
<td>87%</td>
<td>0.70 (0.35 – 1.00)</td>
</tr>
<tr>
<td>Vascular features</td>
<td>81%</td>
<td>0.74 (0.49 – 0.99)</td>
</tr>
<tr>
<td><strong>Cellular</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphology</td>
<td>84%</td>
<td>0.66 (0.31 – 1.00)</td>
</tr>
<tr>
<td>Cohesiveness</td>
<td>74%</td>
<td>0.33 (-0.03 – 0.68)</td>
</tr>
<tr>
<td>Borders</td>
<td>87%</td>
<td>0.74 (0.38 – 1.00)</td>
</tr>
</tbody>
</table>

$p_a =$ percent agreement.
### TABLE 3. INTEROBSERVER AGREEMENT AND DIAGNOSTIC ACCURACY FOR CANCER DIAGNOSIS

<table>
<thead>
<tr>
<th></th>
<th>CLE</th>
<th>WLC</th>
<th>WLC + CLE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interobserver Agreement</strong></td>
<td>$p_a$</td>
<td>$\kappa$ (95% CI)</td>
<td>$p_a$</td>
</tr>
<tr>
<td>Experienced CLE urologists</td>
<td>90%</td>
<td>0.80 (0.45 – 1.00)</td>
<td>74%</td>
</tr>
<tr>
<td>Novice CLE urologists</td>
<td>77%</td>
<td>0.55 (0.46 – 0.64)</td>
<td>78%</td>
</tr>
<tr>
<td>Pathologists</td>
<td>81%</td>
<td>0.61 (0.47 – 0.76)</td>
<td>-</td>
</tr>
<tr>
<td>Nonclinical researchers</td>
<td>77%</td>
<td>0.49 (0.38 – 0.60)</td>
<td>-</td>
</tr>
<tr>
<td><strong>Diagnostic Accuracy</strong></td>
<td>Sn</td>
<td>Sp</td>
<td>Sn</td>
</tr>
<tr>
<td>Experienced CLE urologists</td>
<td>84%</td>
<td>88%</td>
<td>89%</td>
</tr>
<tr>
<td>Novice CLE urologists</td>
<td>75%</td>
<td>81%</td>
<td>86%</td>
</tr>
<tr>
<td>Pathologists</td>
<td>84%</td>
<td>81%</td>
<td>-</td>
</tr>
<tr>
<td>Nonclinical researchers</td>
<td>89%</td>
<td>73%</td>
<td>-</td>
</tr>
</tbody>
</table>

$p_a$ = percent agreement; $Sn$ = sensitivity; $Sp$ = specificity
TABLE 4. INTEROBSERVER AGREEMENT AND DIAGNOSTIC ACCURACY FOR CANCER GRADING

<table>
<thead>
<tr>
<th></th>
<th>CLE</th>
<th>WLC</th>
<th>WLC + CLE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interobserver Agreement</strong></td>
<td>$p_a$</td>
<td>$\kappa$ (95% CI)</td>
<td>$p_a$</td>
</tr>
<tr>
<td>Experienced CLE urologists</td>
<td>68%</td>
<td>0.64 (0.41 – 0.87)</td>
<td>74%</td>
</tr>
<tr>
<td>Novice CLE urologists</td>
<td>53%</td>
<td>0.39 (0.31 – 0.47)</td>
<td>66%</td>
</tr>
<tr>
<td>Pathologists</td>
<td>65%</td>
<td>0.55 (0.43 – 0.68)</td>
<td>-</td>
</tr>
<tr>
<td>Nonclinical researchers</td>
<td>63%</td>
<td>0.56 (0.47 – 0.65)</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Diagnostic Accuracy</strong></th>
<th>Sn</th>
<th>Sp</th>
<th>Sn</th>
<th>Sp</th>
<th>Sn</th>
<th>Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experienced CLE urologists</td>
<td>50%</td>
<td>94%</td>
<td>55%</td>
<td>50%</td>
<td>64%</td>
<td>81%</td>
</tr>
<tr>
<td>Novice CLE urologists</td>
<td>52%</td>
<td>75%</td>
<td>59%</td>
<td>54%</td>
<td>55%</td>
<td>65%</td>
</tr>
<tr>
<td>Pathologists</td>
<td>50%</td>
<td>66%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nonclinical researchers</td>
<td>49%</td>
<td>73%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>High Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experienced CLE urologists</td>
<td>75%</td>
<td>64%</td>
<td>50%</td>
<td>73%</td>
<td>69%</td>
<td>73%</td>
</tr>
<tr>
<td>Novice CLE urologists</td>
<td>46%</td>
<td>74%</td>
<td>44%</td>
<td>76%</td>
<td>54%</td>
<td>67%</td>
</tr>
<tr>
<td>Pathologists</td>
<td>50%</td>
<td>66%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nonclinical researchers</td>
<td>63%</td>
<td>60%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$p_a$ = percent agreement;  
Sn = sensitivity;  
Sp = specificity
FIG 1. Computer-based, interactive training modules for confocal laser endomicroscopy of the bladder. (A) CLE Training: Observers were trained to identify six key CLE features. (B) Teaching set: Teaching set consisting of fifteen video sequences to iteratively practice diagnosing and grading CLE sequences. (C) Validation set: All observer groups reviewed and diagnosed CLE sequences. The experienced CLE and novice CLE urologist groups continued on to diagnose two additional sets of WLC and WLC+CLE images.
Tables and Figures:

FIG 1. Computer-based, interactive training module for confocal laser endomicroscopy of the bladder (A) CLE Training: Observers were trained to identify six key CLE features. (B) Teaching set: Teaching set consisting of fifteen video sequences to iteratively practice diagnosing and grading CLE sequences. (C) Validation set: All observer groups reviewed and diagnosed CLE sequences. The experienced CLE and novice CLE urologist groups continued on to diagnose two additional sets of WLC and WLC+CLE images.

TABLE 1. Diagnosis Table.

TABLE 2. Interobserver Agreement of Features Observed on CLE.

TABLE 3. Interobserver Agreement and Diagnostic Accuracy for Cancer Diagnosis.

TABLE 4. Interobserver Agreement and Diagnostic Accuracy for Cancer Grading.