Introduction

Outbreaks of duodenoscope-related infections represent a global problem, with infections occurring despite reported adherence to reprocessing guidelines. This has prompted enhanced scrutiny of all potential contributing factors, including endoscope reprocessing. Reprocessing of flexible endoscopes is a multistep process that includes point-of-use pre-cleaning, manual cleaning, and high-level disinfection (HLD) followed by alcohol flushes and drying. Undetected damage within the inner channels of endoscopes, such as deep grooves and scratches, may serve as sanctuaries for bacteria, impede adequate manual cleaning and HLD, and potentially promote infection transmission [1–3]. Accordingly, some have recommended that duodenoscopes be returned to the endoscope manufacturer at least once per year for inspection and servicing [4]. However, because damage to internal channels of endoscopes may occur in the course of any given procedure, an annual inspection, while being a useful step forward, is not sufficient [5–7].

Initial application of deep learning to borescope detection of endoscope working channel damage and residue

ABSTRACT

Background and study aims Outbreaks of endoscopy-related infections have prompted evaluation for potential contributing factors. We and others have demonstrated the utility of borescope inspection of endoscope working channels to identify occult damage that may impact the adequacy of endoscope reprocessing. The time investment and training necessary for borescope inspection have been cited as barriers preventing implementation. We investigated the utility of artificial intelligence (AI) for streamlining and enhancing the value of borescope inspection of endoscope working channels.

Methods We applied a deep learning AI approach to borescope inspection videos of the working channels of 20 endoscopes in use at our academic institution. We evaluated the sensitivity, accuracy, and reliability of this software for detection of endoscope working channel findings.

Results Overall sensitivity for AI-based detection of borescope inspection findings identified by gold standard endoscopist inspection was 91.4%. Labels were accurate for 67% of these working channel findings and accuracy varied by endoscope segment. Read-to-read variability was noted to be minimal, with test-retest correlation value of 0.986. Endoscope type did not predict accuracy of the AI system (P=0.26).

Conclusions Harnessing the power of AI for detection of endoscope working channel damage and residue could enable sterile processing department technicians to feasibly assess endoscopes for working channel damage and perform endoscope reprocessing surveillance. Endoscopes that accumulate an unacceptable level of damage may be flagged for further manual evaluation and consideration for manufacturer evaluation/repair.
Small-caliber borescopes are utilized at manufacturers’ repair facilities to detect damage within endoscope working channels. A recent technological advance has been the development of “consumer” borescopes, which we and others have utilized to visualize the inner working channels of endoscopes [6, 8, 9]. These studies have detected damage, debris, and persistent residue despite HLD within endoscope working channels [6, 8, 9]. The extent of damage and residue reported within these channels has varied by institution and practice setting [6, 8, 9]. These studies suggest that borescope evaluation of endoscope working channels at the institutional level may potentially help guide the need and timing for return of endoscopes to manufacturers for repair [6, 8, 9]. Visual inspection of endoscopes is a required step in the reprocessing guidelines of several associations, including the European Society of Gastrointestinal Endoscopy, European Society of Gastroenterology Nurses and Associates, Association of Perioperative Registered Nurses, Association for the Advancement of Medical Instrumentation and the International Association of Healthcare Central Service Material Management, and borescope inspection may comprise a facet of this visual inspection step [10–12].

Despite the availability and proven utility of “consumer” borescopes, and tentative endorsements by professional societies for incorporating routine borescope inspection into endoscope reprocessing, adoption by endoscopy units has been very limited, given the existing barriers to implementation.

Considerable time investment and training would be necessary to bring technicians to competency in interpreting borescope video feeds. Interobserver variability in identification and rating the degree of working channel damage will remain an issue. An additional inevitable issue will be human failure and inattention during endoscope reprocessing, which remain problematic [13, 14]. Increased automation of processes, therefore, is preferable [13, 14].

Artificial intelligence (AI) has become ubiquitous and indispensable for solving complex problems in many sciences. It consistently outperforms human observation in many domains and is most valuable when integrated with human intelligence [15, 16]. Visual diagnosis in radiology and pathology has been enhanced by application of deep learning algorithms, which have decreased missed lesions or findings [17–19]. Within the realm of clinical medicine, AI is being integrated with human intelligence to increase diagnostic yield in cardiology imaging [20, 21], and in gastrointestinal endoscopy [22, 23].

A potential solution to the existing barriers to implementation of borescope evaluation of endoscopes following every use is application of AI. This proof-of-concept study characterizes for the first time, application of the power of AI in streamlining and facilitating borescope evaluation of endoscope working channels, for the detection of damage and residue. A simplified AI-assisted process may potentially allow institution of a policy of “borescope evaluation of every endoscope following every procedure.”

**Methods**

**Institution**

This study was conducted in a high-volume tertiary-care academic medical center endoscopy unit (>50 endoscopic procedures performed daily). The study was approved by the Stanford University Medical Center Institutional Review Board (40603). No human subjects were involved, and patient-specific data were not collected.

**Endoscope evaluation**

We utilized an ultra-slim flexible inspection borescope (SteriCam, Sanovas Inc., San Rafael, California, United States) to inspect endoscope working channels for retained residual fluid and damage following reprocessing, as previously described [9]. Previously recorded videos of endoscope working channel borescope inspections were utilized for this study [5]. Video recordings of borescope inspections of five standard/diagnostic gastroscopes, five adult/standard colonoscopes, five linear echoendoscopes and five duodenoscopes manufactured by both Olympus (Olympus America, Center Valley, Pennsylvania, United States) and Pentax (Pentax of America, Montvale, New Jersey, United States) were included in this study.

As we have previously described, typical endoscopes have three distinct working channel segments: the inlet region where the biopsy port in the endoscope handpiece joins the working channel within the endoscope shaft, the uniform cylindrical channel segment within the endoscope shaft, characterized by a white or gray-green lining with a metallic sheen, and the channel segment at the distal bending tip, characterized by circular rings [5].

Borescope examination videos were reviewed in a blinded fashion by three endoscopists, who scored working channel damage, fluid residue, and/or debris in accordance with our previously described scoring system [5]. Videos of each borescope inspection were used to compare evaluations by multiple endoscopists and AI technology, which would be extremely challenging to accomplish with live borescope inspections.

**Scoring system for borescope evaluation of endoscope working channel findings**

A scoring system was developed at Stanford by consensus between investigators based on the entire range of possible endoscope working channel damage detected at both our institution and at an endoscope manufacturer’s national service center [5]. Working channel damage was identified and labeled (from least to most severe) as: superficial scratch, adherent peel, deep scratch, burn, channel buckling, stain, and perforation. The severity of this working channel damage was rated on a scale of 0–3 (0 = none, 1 = Mild, 2 = Moderate, 3 = Severe) [5].

**Artificial intelligence technology**

The newly developed deep learning system (WatchDog) utilized in this study is a proprietary multi-layer convolutional neural network-based algorithm with elements of its code written in C++ and Python. This algorithm was trained on 30 independent
learning sample videos of endoscope working channel borescope inspections, of which 1000 frames included working channel damage, debris and residue appropriate for training of the algorithm to detect significant working channel findings. This independent learning sample was not utilized for the test portion of this study.

The deep learning algorithm conducts a frame-by-frame analysis and identifies, highlights, and details each working channel abnormality. In addition, it annotates each finding with the probability/certainty that the finding is consistent with the assigned label (percentage value, with higher percentage indicating increased certainty). Each finding and the level of certainty of each finding (if over 90% certainty) is exported to an Excel spreadsheet to enable quantification and support documentation. Findings with <90% certainty of label accuracy are depicted in the analysis, but not exported to the Excel spreadsheet.

Evaluation of artificial intelligence technology
A second independent “test” set of borescope videos of 20 endoscope working channel inspections was used to evaluate the AI system’s performance in identification of endoscope working channel findings and accurate labeling of these findings within endoscope working channels. These 20 borescope inspection videos had previously been consensus rated by three endoscopists. These endoscopist ratings were based on each specific working channel finding in accordance with our endoscope working channel damage and residue rating scale and included a score reflecting the overall extent and severity of endoscope working channel damage.

The deep learning algorithm was applied to each borescope video of endoscope working channel inspection and findings detected by the AI algorithm were directly compared with endoscopist reviewer detection of each finding. Endoscopist detection and label assignment for each finding (triplicate independent review followed by consensus review) was established as the gold standard. For the sensitivity analysis, the focus was on detection of a finding rather than label assigned. Sensitivity was calculated as percentage of endoscopist-detected findings that were also detected by the AI deep learning algorithm. Accuracy analysis evaluated assignment of the correct label to each finding, with the correct label again based on the endoscopist-assigned label and the AI deep learning algorithm label evaluated relative to the endoscopist label. This study design was conceptualized to assess the extent to which this AI software could accurately identify and label endoscope working channel findings in comparison with manual endoscopist evaluation of endoscope working channel borescope inspection.

Statistical analysis
Analyses were conducted using SAS Enterprise Guide version 7.11 HF3 (SAS Institute Inc., Cary, North Carolina, United States) and Microsoft Excel. Regression analysis was performed using generalized linear models.

Results
Algorithm output
The AI algorithm successfully detected endoscope working channel abnormalities and applied a label to each abnormality (Fig. 1). A sample output frame from the AI system is depicted in Fig. 1. The algorithm’s real-time performance within an endoscope working channel is depicted in a video (Video 1).

Sensitivity and accuracy
The sensitivity and accuracy of the frame-based analysis were evaluated using endoscopist borescope inspection and identification of working channel findings as the gold standard benchmark.

Overall sensitivity of AI for detection of the presence of any working channel finding was 91.4%. When a finding was identified within an endoscope working channel, accurate labeling of that finding was accomplished 67% of the time. Accuracy of endoscope working channel findings varied by finding type (Table 1). When labels were inaccurate, the most common inaccuracies included assignment of a “peeling” or “debris” label to findings endoscopists labeled as “scratch” (41%) and mis-assignment of a “scratch” label to findings endoscopists labeled as “debris,” “droplet,” or “peeling” (33%).

The most common basis for a false-positive AI finding was glare/reflection on the inner working channel surface associated with the borescope light source. Borescope glare was associated with 67% of false-positive AI working channel findings. The glare was most commonly mislabeled by the AI algorithm as scratch (41%), droplet (26%), or peeling (21%).

Due to the frame-by-frame analysis approach, which enhances sensitivity of detection of working channel findings, duplicate working channel findings were detected in consecutive frames for a subset of findings and these duplicate findings required manual review to ensure accurate reporting of findings.

Accuracy by endoscope working channel segment
Accuracy of AI analysis varied within the three distinct endoscope working channel segments (inlet region, shaft, and distal bending tip). The highest accuracy of detection of endoscope working channel abnormalities was evident in the endoscope shaft region (78%) and lower accuracy was noted in the inlet region (44%) and distal bending tip (41%) (Table 2).

Accuracy of AI algorithm in different endoscopes
Endoscope type (colonoscope vs. gastroscope vs. echoendoscope vs. duodenoscope) did not predict accuracy of the AI system in detection of working channel findings (P=0.19) in our regression analysis.

Accuracy based on burden of working channel damage
Regression analysis revealed that the overall working channel damage rating of a given endoscope did not predict accuracy of working channel finding detection by the AI algorithm (P=0.26).
Reliability of AI algorithm

The AI algorithm was applied to a 50% subset of borescope working channel inspection recordings in duplicate to measure variability from read to read. Read-to-read variability was noted to be minimal, with an overall test-retest correlation value of 0.986.

Discussion

In this high-tech era of voice activated digital assistants and self-driving automobiles, it is evident that deep convolutional neural network learning technology has the capability to mimic the human brain. In many applications, both medical and non-medical, performance is enhanced by integration of AI with human intelligence [16, 24]. Our data represent the first application of AI to enhance borescope evaluation of endoscope working channel damage and residue and address the global challenge of infection transmission associated with endoscopic procedures. The integration of AI into borescope evaluation has the potential to transform this process and may facilitate more widespread adoption.

Although the calls by some for borescope evaluation of endoscope working channels after every procedure may be considered extreme, there is some rationale to this viewpoint, as devices inserted through endoscope working channels may potentially cause damage during any given endoscopic procedure. The annual evaluation of endoscopes by manufacturers, currently recommended by manufacturers and the FDA [25], although a significant step forward, may therefore not be effective in detecting internal damage in a timely manner. A tempered approach with more frequent scheduled borescope evaluations of endoscopes following a specified number of usages or specified time intervals might be a reasonable middle ground. Ideally, such evaluation would identify the type and degree of damage, if present, and determine when the damage is significant enough to require removal of the endoscope from in-service.
clinical use for repair by the manufacturer. If endoscope working channel borescope inspection is implemented on a rotational surveillance basis as we recommend (e.g. every endoscope is inspected on a weekly or monthly basis), it makes sense to perform this borescope inspection after complete reprocessing and drying of "patient ready" endoscopes. If endoscope units implement inspection of every scope after every procedure, endoscope working channel inspection could potentially be considered after manual cleaning to conserve costs, but the logistics of underwater inspection of a wet scope and adequate cleaning of borescopes after this inspection would need to be addressed. Ultimately endoscope units will need to decide how they wish to implement endoscope working channel inspections.

In previous borescope studies, interpretation of findings was performed by highly motivated endoscopists and researchers [6, 8, 9]. To operationalize widespread adoption of frequent borescope evaluations, these would ideally be integrated into endoscope reprocessing and performed by sterile processing department (SPD) technicians. However, SPD technicians have relatively low levels of training and many competing demands for their time. An endoscope reprocessing technician could reasonably advance the borescope at a steady rate within the endoscope working channel. Endoscope working channel findings could then be interpreted by the algorithm to enforce quality control for these borescope inspections of endoscope working channels and to avoid dedicating physician time to something that is within domain of AI. AI could enforce quality control of endoscope working channel inspections in this setting, as described in the text of the manuscript. Moreover, studies of their performance during the manual steps of reprocessing have highlighted issues related to operator lapses and inattention [26]. Introduction of AI into the process of borescope inspection can potentially help resolve all of these problematic issues.

For AI to replace a well-trained human performing borescope inspections, high performance on three key AI algorithm metrics would be essential: (1) sensitivity in detecting working channel damage; (2) accuracy and specificity in characterizing working channel damage; and (3) reliability and consistency of detection of working channel findings over repeat inspections. Accuracy in characterizing borescope inspection findings would include optimal filtering of artifactual visual phenomena such as glare. Rapidity of interpretation of endoscope working channel findings is desirable with a final report ready prior to completion of the remaining steps of the reprocessing. This would allow the endoscope to be pulled immediately from clinical use should significant damage be detected. Finally, the AI program should be compatible with all major borescope platforms and all endoscope types from all major manufacturers.

Reassuringly, this initial application of deep learning algorithms to borescope inspection is notable for high sensitivity (> 90%) for detection of endoscope working channel damage. Accuracy of deep learning-based labeling of these working channel findings was also relatively high and influenced by endoscope segment, with the endoscope shaft having the highest accuracy for correct identification of working channel damage and residue. We suspect that this higher accuracy for identification of working channel damage and residue within the endoscope shaft relates to the characteristics of the endoscope shaft region, as this is the most uniform segment. It is reassuring that endoscope type did not impact the sensitivity or accuracy of this AI algorithm, as channel diameters and visual characteristics vary by endoscope type. Furthermore, the test-retest reliability of this AI algorithm was excellent.

Some enhancements to the current state of AI for this application are currently in progress and will facilitate the ultimate goal of unsupervised AI overseen borescope evaluations. Enhancement of the training set and algorithm will refine frame-by-frame detection of borescope findings so that AI may discern the initial, ongoing and final components of a given finding (e.g. scratch) rather than detecting the same scratch as multiple independent findings. Second, erroneous detection of glare/reflection as endoscope working channel damage and residue will be addressed by additional training of the algorithm in endoscope segments most affected by glare/reflection so that the algorithm will detect and disregard these artifactual visual phenomena within endoscope working channels. These enhancements to the AI algorithm should enable not only detection and identification of working channel irregularities, but also grading of the severity of these findings and assignment of a global damage rating [5] which reflects both the overall extent and severity of damage.

This deep learning endoscope inspection technology brings us one step closer to the ideal scenario to minimize transmission of infection by endoscopes, in which, as a part of standard workflow, an SPD technician would advance a borescope through each endoscope working channel on a scheduled basis.
Conclusions

Our data represent the first demonstration of the application and feasibility of AI for borescope detection of endoscope working channel damage. This transformative innovation has the potential to decrease the risk of endoscopy-related transmission of infection. Utilizing AI for this application is consistent with the principle that automation is preferable to manual processes, when standardized and reliable outcomes are desired. Inclusion of AI algorithms could conceivably automate borescope evaluations, thereby facilitating widespread adoption and integration of this process into standard endoscope reprocessing protocols.

Acknowledgements

The SteriCam device and WatchDog software were loaned to the investigators for this study; however, the study was entirely investigator initiated, with study design, data collection, analysis, and interpretation performed independently by the investigators.

Competing interests

The authors declare that they have no conflict of interest.

References


