

Article

Detection Accuracy and Latency of Colorectal Lesions with Computer-Aided Detection System Based on Low-Bias Evaluation

Hiroaki Matsui ¹, Shunsuke Kamba ¹, Hideka Horiuchi ¹, Sho Takahashi ², Masako Nishikawa ², Akihiro Fukuda ³, Aya Tonouchi ³, Natsumaro Kutsuna ³, Yuki Shimahara ³, Naoto Tamai ¹ and Kazuki Sumiyama ^{1,*}

- ¹ Department of Endoscopy, The Jikei University School of Medicine, 3-25-8 Nishishimbashi, Minato-ku, Tokyo 105-8461, Japan; hiro24tth80@yahoo.co.jp (H.M.); kanba@jikei.ac.jp (S.K.); takeishihideka34@gmail.com (H.H.); tamai-naoto@jikei.ac.jp (N.T.)
- ² Clinical Research Support Center, The Jikei University School of Medicine, 3-25-8 Nishishimbashi, Minato-ku, Tokyo 105-8461, Japan; sho@jikei.ac.jp (S.T.); mnishikawa@jikei.ac.jp (M.N.)
- ³ LPixel Inc., 1-6-1 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan; fukuda@lpixel.net (A.F.); tonouchi@lpixel.net (A.T.); kutsuna@lpixel.net (N.K.); shimahara@lpixel.net (Y.S.)
- * Correspondence: kaz_sum@jikei.ac.jp; Tel.: +81-3-3433-1111



Citation: Matsui, H.; Kamba, S.; Horiuchi, H.; Takahashi, S.; Nishikawa, M.; Fukuda, A.; Tonouchi, A.; Kutsuna, N.; Shimahara, Y.; Tamai, N.; et al. Detection Accuracy and Latency of Colorectal Lesions with Computer-Aided Detection System Based on Low-Bias Evaluation. *Diagnostics* **2021**, *11*, 1922. <https://doi.org/10.3390/diagnostics11101922>

Academic Editors: Anastasios Koulaouzidis and Ervin Toth

Received: 9 August 2021

Accepted: 14 October 2021

Published: 17 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: We developed a computer-aided detection (CADE) system to detect and localize colorectal lesions by modifying You-Only-Look-Once version 3 (YOLO v3) and evaluated its performance in two different settings. The test dataset was obtained from 20 randomly selected patients who underwent endoscopic resection for 69 colorectal lesions at the Jikei University Hospital between June 2017 and February 2018. First, we evaluated the diagnostic performances using still images randomly and automatically extracted from video recordings of the entire endoscopic procedure at intervals of 5 s, without eliminating poor quality images. Second, the latency of lesion detection by the CADE system from the initial appearance of lesions was investigated by reviewing the videos. A total of 6531 images, including 662 images with a lesion, were studied in the image-based analysis. The AUC, sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 0.983, 94.6%, 95.2%, 68.8%, 99.4%, and 95.1%, respectively. The median time for detecting colorectal lesions measured in the lesion-based analysis was 0.67 s. In conclusion, we proved that the originally developed CADE system based on YOLO v3 could accurately and instantaneously detect colorectal lesions using the test dataset obtained from videos, mitigating operator selection biases.

Keywords: artificial intelligence; computer-aided detection; colonoscopy; colorectal lesions; deep learning

1. Introduction

Population-based programs for screening colorectal cancer (CRC) have been widely adopted in several countries, owing to the high prevalence of CRC [1]. Colonoscopy is the most sensitive and reliable examination for detecting CRC and precursor lesions [2]. It also facilitates the immediate prophylactic removal of early neoplastic lesions. A series of cohort and case-control studies demonstrated that colonoscopy with prophylactic lesion removal could reduce the incidence and mortality of CRCs [3–6]. It is reported that most post-colonoscopy CRCs (PCCRCs) are mainly caused by lesions missed during previous colonoscopies [7]. The insufficient quality of an examination substantially increases the risk of PCCRC and mortality of CRC [8]. The adenoma detection rate (ADR) is considered the most reliable indicator of an endoscopist's ability to detect neoplastic lesions [9]. The ADR of endoscopists is known to vary even amongst the experts, and the risk of PCCRC can increase if the examination is performed by an endoscopist with ADR < 20% [10]. Although an array of novel imaging technologies has been developed and the imaging quality of

colonic lesions has been consequently improved, their effectiveness for improving the detectability of colorectal lesions is still insufficient [11–13].

Image recognition based on deep learning (DL) has been increasingly applied in medical imaging analysis [14,15]. Various DL-based computer-aided detection (CADe) systems for colonoscopy have been developed to aid in the detection of colorectal adenomatous lesions. Pioneering studies have demonstrated that CADe could accurately detect colorectal lesions with a sensitivity of over 90% [16–19]; however, the diagnostic performance of CADe, including sensitivity and specificity, has been evaluated retrospectively in most studies, using readily available representative still images and edited video clips stored in a medical record. This involves a substantial risk of selection bias associated with the unavoidable manipulation of the rate of the presence of objects [20,21]. In addition, the accuracies could be overrated in the retrospective settings without evaluating the latency of the detection by CADe from the initial appearance of lesions or human detection. Meanwhile, the advantage of assistance from a DL-based CADe system over non-assisted human perception has already been demonstrated in several randomized controlled trials (RCTs), including our study evaluating ADR or the adenoma miss rate (AMR) as the primary outcome. The result was confirmed in a series of meta-analyses, some of which analyzed the outcomes of RCTs [22–25]. However, RCT demands a significant amount of effort and time; retesting each updated DL-based algorithm using RCT may be suboptimal as a reference to set the direction of algorithmic developments. Therefore, it is imperative to establish a simpler and more reliable evaluation standard for the unit testing to retrospectively evaluate the diagnostic performance of CADe by mitigating the risks of selection bias [26].

In this study, we originally developed a CADe system to detect and localize colorectal lesions by modifying You-Only-Look-Once version 3 (YOLO v3). The aim of this study was primarily to assess the performance of the CADe system for colorectal lesions in fair testing settings, eliminating operator selection biases by automatically sampling a validation dataset from unedited videos of endoscopic observations without excluding low-quality images in the image-based analysis. Furthermore, the latency of the detection by CADe from the initial appearance of lesions was conducted by reviewing the videos of the entire colonic inspection in the same cohort.

2. Materials and Methods

2.1. DL-Based Algorithm to Assist Lesion Detection and Localization

The CADe system evaluated in this study was developed by modifying YOLO v3 to assist the detection and localization of colorectal lesions during a standard colonoscopy in real time. YOLO v3 is a DL-based image recognition algorithm that enables real-time object detection while ensuring a processing speed superior to other convolutional neural networks (CNNs). The developed system achieves a fast processing speed, which is partially improved using YOLO v3. The system is compatible with routinely used imaging techniques, including white light imaging (WLI), chromoendoscopy (CE), and narrow-band imaging (NBI).

We trained an object detection model based on YOLO v3, which contains several CNN architectures. The weights of the architectures were initialized through training on an ImageNet data corpus [27] before training on our data. Each architecture shared similar components—convolutional layers, activation functions, batch normalization [28], up-sampling, and skip connections [29,30]. We used the leaky rectified linear unit [31] as the activation function. The model did not contain a fully connected layer or a pooling layer; this enabled it to process images of all sizes.

In the final convolutional layer, the positions of the objects and objectness were output using logistic regression, and class predictions were output using independent logistic classifiers. Using the stochastic gradient descent optimizer, we optimized two types of losses during training—squared error loss for object positions and cross-entropy loss for objectness and class predictions. For implementation, the darknet software libraries (<http://pjreddie.com/darknet/> accessed on 18 October 2018) were used. Data augmentation was

applied to avoid overfitting of the training data and to generate additional data by applying random transformations to the images during training. Random mirroring, distortion, shift, and resizing were used for this purpose. Additionally, in the lesion-based analysis, to reduce false positive detection, a bounding box was set to appear on a monitor when the CADe system detects the presence of a lesion in at least 3 of any 5 sequential frames. This setting was determined via trial-and-error.

2.2. Objectives for Training and Validation

We trained the CADe system with stored still images and images extracted from videos recorded at Jikei University between July 2013 and March 2018. The total number of cases included in the dataset was 4080 (3544 still images, 536 from videos), which contained 26,462 lesions (24,860 still images, 1602 from videos). All images and video clips were obtained under a standard colonoscopy setting (endoscopes used: PCF-H290I, PCF-H290ZI, CF-HQ290I, CF-Q260AI, PCF-Q260AI, PCF-Q260AZI, CF-H260AZI, and PCF-Q260JI, Olympus Optical Co., Tokyo, Japan; video processors used: EVIS LUCERA CV-260 and VIS LUCERA ELITE system, Olympus, Tokyo, Japan). The CADe system was eventually trained with 84,550 images (79,079 images with lesions and 5471 without lesions).

Validation of the CADe system was retrospectively conducted by collecting and reviewing videos of the colonoscopies of patients who underwent endoscopic resection, excluding those who underwent endoscopic submucosal dissection (ESD) of colorectal lesions between June 2017 and February 2018 at the Jikei University Hospital. Cases that did not record the intubation or withdrawal of the scope as well as cases used for CADe training were excluded from the analysis. Then, we selected 20 cases from that period for evaluation using a random number table, as we assumed that 2000–4000 images would be necessary for validation in reference to preceding research, and 100–200 images could be obtained for each case. Subsequently, we evaluated the performance of the CADe system by performing two analyses—an image-based analysis to calculate the diagnostic performances using automatically sampled still images from videos, and a lesion-based analysis using video clips to measure the latency in the detection of lesions from their initial appearance. The study was approved by the ethics committee of the Jikei University School of Medicine, Tokyo, Japan (no. 30-173(9194)), and it conforms to the provisions of the Declaration of Helsinki. As a retrospective study, informed consent was obtained in the form of opting out on the website.

2.3. Image-Based Analysis to Calculate Diagnostic Performance Using Automatically Sampled Still Images

The still images for the validation dataset to calculate the diagnostic performance of the CADe system were automatically and blindly extracted from the video clips, including insertion to withdrawal of the scope at intervals of 5 s, without eliminating poor quality images. Then, we excluded any images with (1) the outside of the colorectum; (2) colorectal lesions and ulcers after resection; (3) imaging modes other than WLI, CE, and NBI; (4) artificial objects (endoscopic devices); (5) submucosal fluid bleb following needle injection; (6) more than two colorectal lesions.

2.3.1. Receiver Operating Characteristic (ROC) Curve Drawing to Determine the Optimal Threshold

An expert endoscopist reviewed the still images, and labeled lesions contained in the images were framed with a rectangular bounding box (ground truth bounding box).

The CADe system separately analyzed the still images, and bounded boxes (predicted bounding boxes) were drawn around detected lesions in various probability thresholds.

Thereafter, the still image dataset with ground truth bounding boxes and CADe-predicted bounding boxes were compared for calculating sensitivity and specificity in various probability thresholds based on the definitions of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), as described in Section 2.3.2. ROC curves were plotted for each imaging modality (Overall, WLI, CE, or NBI) by varying the probability

thresholds from 0.01 to 0.99 at intervals of 0.01. We defined the optimal cutoff value of the probability threshold for each modality when it was at the maximum Youden index [32].

2.3.2. Definitions of True Positive, True Negative, False Positive, and False Negative

In the image-based analysis, we primarily evaluated if the CADe system could demonstrate lesion-positive frames with bounding boxes in order to calculate the diagnostic performances, that is, AUC, sensitivity, specificity, PPV, NPV, and accuracy as binary variables in a 2×2 table. Therefore, TP, TN, FP, and FN in this study were defined as follows: Cases in which both ground truth bounding boxes and the CADe bounding boxes were in the same image were defined as TP, regardless of their overlapping. Cases without either ground truth bounding boxes or the CADe bounding boxes were solely defined as TN. When the CADe bounding boxes existed in an image without ground truth bounding boxes, the case was categorized as FP, regardless of the number of the CADe bounding boxes. Cases were defined as FN when the CADe system failed to place bounding boxes on an image with ground truth bounding boxes (Figures 1 and 2).

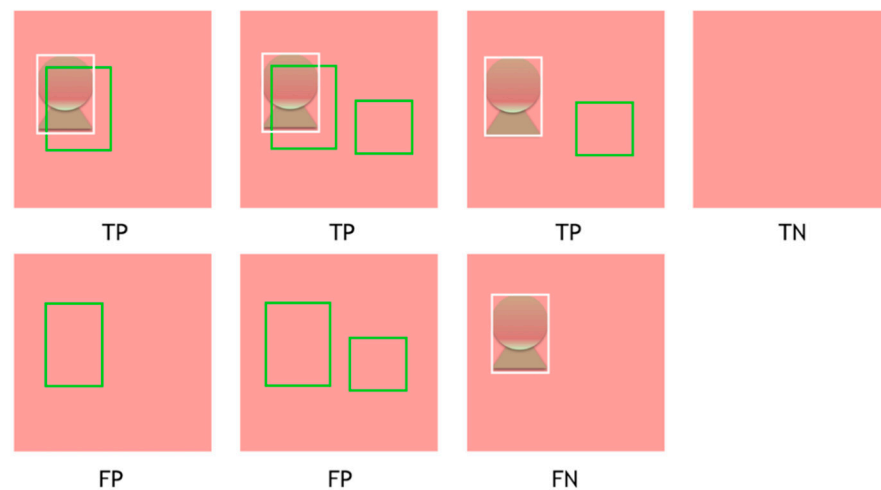


Figure 1. Schematic presentation of the definition of TP, TN, FP, and FN in the image-based analysis. White bounding box: ground truth bounding box provided by expert endoscopist. Green bounding box: predicted bounding box provided by CADe. TP, true positive; TN, true negative; FP, false positive.

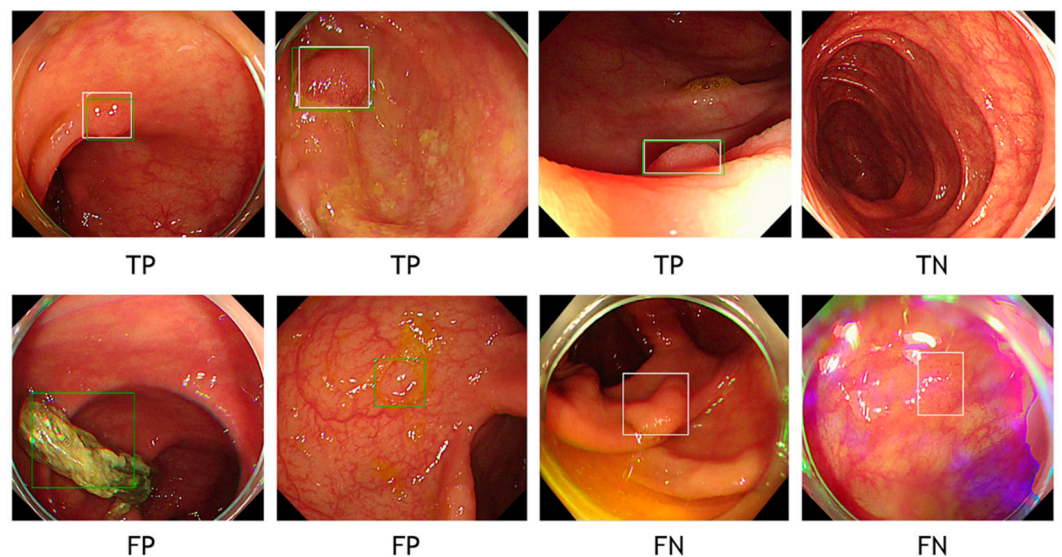


Figure 2. Ground truth and predicted bounding boxes overlaid on endoscopic images. White box: ground truth bounding box provided by expert endoscopist. Green box: predicted bounding box provided by CADe.

Subsequently, sensitivity, specificity, PPV, NPV, and accuracy were calculated using the following expressions: $\text{Sensitivity} = n(\text{TP}) / (n(\text{TP}) + n(\text{FN}))$

$$\text{Specificity} = n(\text{TN}) / (n(\text{TN}) + n(\text{FP}))$$

$$\text{Accuracy} = (n(\text{TP}) + n(\text{TN})) / (n(\text{TP}) + n(\text{FN}) + n(\text{FP}) + n(\text{TN}))$$

$$\text{PPV} = n(\text{TP}) / (n(\text{TP}) + n(\text{FP}))$$

$$\text{NPV} = n(\text{TN}) / (n(\text{TN}) + n(\text{FN})),$$

where $n(\text{TP})$, $n(\text{FN})$, $n(\text{TN})$, and $n(\text{FP})$ denote the number of images of TP, FN, TN, and FP, respectively.

2.3.3. Localization Accuracy for Predicted Bounding Boxes Using Intersection over Union (IoU)

Intersection over union (IoU) (the area of overlap between the predicted bounding box and the ground truth bounding box) was evaluated for CADe-predicted bounding boxes overlapping with ground truth bounding boxes. In order to identify if the CADe system could accurately predict the localization of lesions or opportunistically place bonding boxes on images with the ground truth bounding boxes, the IoU > 50 ratio for the overlapping boxes was analyzed.

2.4. Lesion-Based Analysis by Reviewing Videos for Evaluating the Latency of CADe

Initially, an expert endoscopist identified and defined the initial appearance of all 69 lesions clinically requiring resection by meticulously reviewing the videos during the scope withdrawal from the scene of the cecal bottom or the anastomotic site with the ileum to the scene of the anal canal, without CADe assistance for all 20 cases. The median length of the videos reviewed was 1589 s (751–3592 s). Then, the CADe system evaluated clipped video-recorded scenes around the initial lesion appearance with a median length of 15 s (7–48 s). The lesion detection time of the CADe system was assessed for each lesion by counting frames between the frame with the initial appearance of a lesion and the time when the same lesion was initially detected by the CADe system. When a lesion had not been detected by the CADe system for 5 s from the initial appearance, we considered the lesion detection by the CADe system as failed.

2.5. Statistical Analysis

Medians with range are presented for continuous variables. Frequencies and proportions are provided for categorical variables. The sensitivity, specificity, PPV, NPV, and accuracy of the CADe system for lesion detection were calculated using the formulas given in Section 2.3.2, and the corresponding two-sided 95% exact confidence intervals (CIs) based on the binomial distribution were estimated. The ROC curve of the CADe system was plotted, and the AUC was calculated using the trapezoidal rule, and the corresponding two-sided 95% exact CI based on the binomial distribution was estimated. The Youden index for each probability threshold of the system was calculated, and the value with the maximum Youden index was defined as the cutoff value. All statistical analyses were performed using Stata 14.2 (StataCorp LP, College Station, TX, USA).

3. Results

A total of 1062 cases underwent endoscopic resection besides endoscopic submucosal resection between June 2017 and February 2018 at the Jikei University Hospital (Figure 3). A total of 738 cases (103 not recorded, 635 used for training) were excluded. From the remaining 324 cases, 20 cases (16 males and 4 females, median age of 63 (36–80) and one case after a right side hemicolectomy) were randomly selected and analyzed.

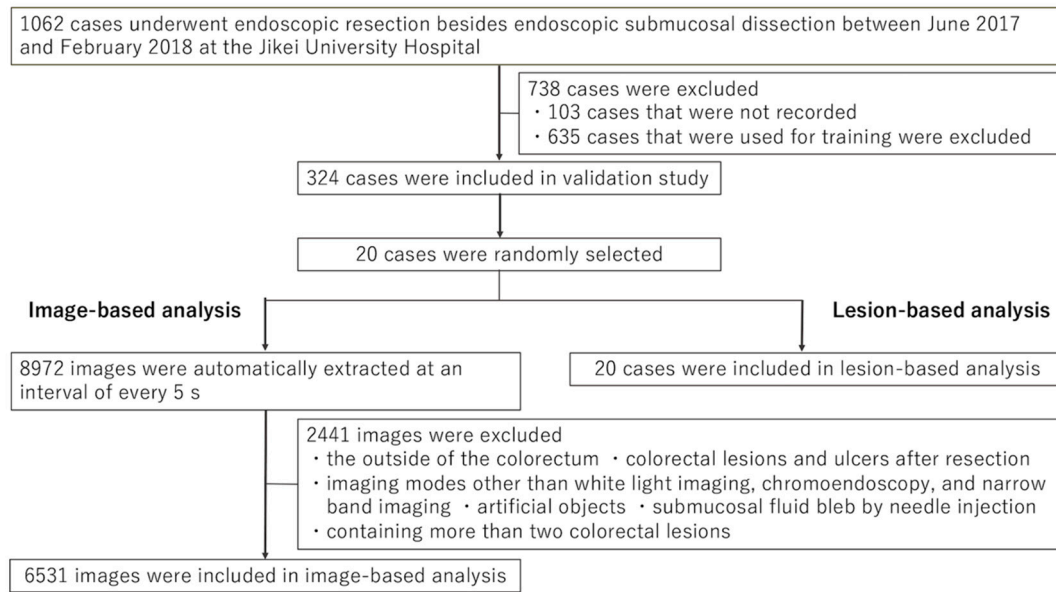


Figure 3. Flowchart for case enrollment and image selection for validation.

3.1. Image-Based Analysis Using Automatically Sampled Still Images

Overall, videos were recorded for 20 cases for a total length of 12 h 28 min 9 s, and the number of still images extracted was 8972. The total number of images analyzed was 6531, after excluding 2441 images according to the exclusion criteria (Figure 3). The number of images with ground truth bounding boxes was 662 (662/6531, 10.1%). The analyzed images contained 5527 WLI images, 824 CE images, and 180 NBI images. The AUC of lesion detection with the CADe system was 0.983 overall, 0.982 in WLI, 0.986 in CE, and 0.986 in NBI (Figure 4). The probability thresholds regarding the maximum Youden index for overall, WLI, CE, and NBI were 0.27, 0.22, 0.43, and 0.05–0.06, respectively (Table 1). When probability thresholds of 0.27 were applied, the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 94.6%, 95.2%, 68.8%, 99.4%, and 95.1%, respectively.

Table 1. Diagnostic performance of computer-aided detection system in image-based analysis with automatically sampled still images.

Type of Imaging Mode (No. of Lesions/No. of Images (Prevalence (%)))	Youden Index	Probability Threshold	Sensitivity (95% CI)(%)	Specificity (95% CI)(%)	PPV (95% CI)(%)	NPV (95% CI)(%)	Accuracy (95% CI)(%)	IoU ≥ 0.5 (TP)(%)
Overall (662/6531 (10.1%))	0.897	0.27	94.6 (92.6–96.2)	95.2 (94.6–95.7)	68.8 (65.7–71.8)	99.4 (99.1–99.6)	95.1 (94.5–95.6)	97.3 (609/626)
WLI (334/5527 (6.0%))	0.9	0.22	95.5 (92.7–97.5)	94.5 (93.8–95.1)	52.6 (48.5–56.6)	99.7 (99.5–99.8)	94.5 (93.9–95.1)	96.9 (309/319)
	0.899	0.27	94.6 (91.6–96.8)	95.3 (94.7–95.8)	56.3 (52.1–60.5)	99.6 (99.4–99.8)	95.2 (94.6–95.8)	97.2 (307/316)
CE (182/824 (22.1%))	0.912	0.43	95.1 (90.8–97.7)	96.1 (94.3–97.5)	87.4 (81.9–91.7)	98.6 (97.3–99.3)	95.9 (94.3–97.1)	97.1 (168/173)
	0.909	0.27	96.7 (93.0–98.8)	94.2 (92.1–95.9)	82.6 (76.9–87.5)	99.0 (97.9–99.6)	94.8 (93.0–96.2)	97.2 (307/316)
NBI (146/180 (81.1%))	0.956	0.05-0.06	95.9 (91.3–98.5)	94.1 (80.3–99.3)	98.6 (95.0–99.8)	84.2 (68.7–94.0)	95.6 (91.4–98.1)	96.4 (135/140)
	0.859	0.27	91.8 (86.1–95.7)	94.1 (80.3–99.3)	98.5 (94.8–99.8)	72.7 (57.2–85.0)	92.2 (87.3–95.7)	97.0 (130/134)

PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; TP, true positive; IoU, intersection over union; WLI, white light imaging; CE, chromoendoscopy; NBI, narrow band imaging.

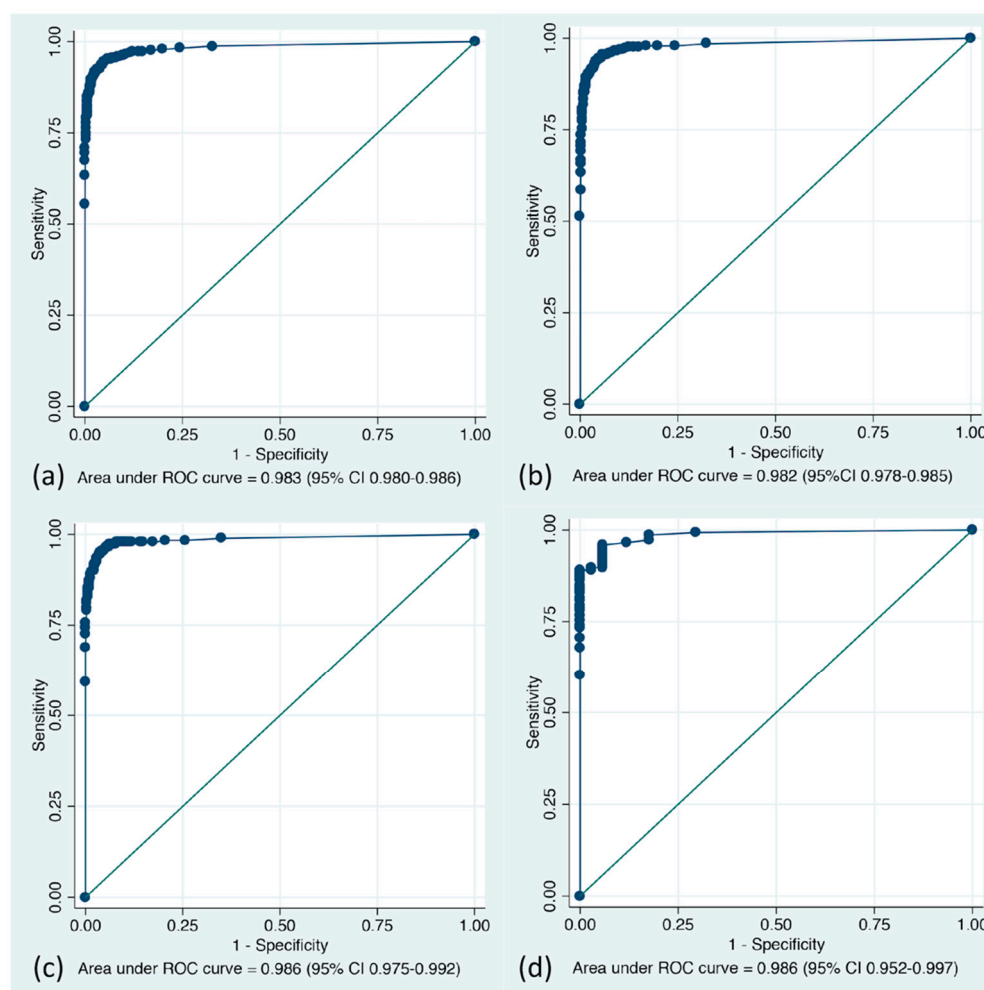


Figure 4. Receiver operating characteristic curves of the CAde system for overall, white light imaging, chromoendoscopy, and narrow-band imaging images in the image-based analysis. (a) Overall ($n = 6531$); (b) White light imaging ($n = 5527$); (c) Chromoendoscopy ($n = 824$); (d) Narrow band imaging ($n = 180$); ROC, Receiver operating characteristic.

The IoU > 50 ratio for the overlapping boxes was 97.3%. The results indicate that the CAde system could accurately predict the localization of detected lesions.

3.2. Lesion-Based Analysis with Video Reviewing

A total of 69 lesions were endoscopically removed in 20 cases. Of those, 57 lesions were morphologically classified as elevated lesions and 12 were classified as flat lesions (morphology of lesions in the Paris classification: 1 lesion of 0-Ip, 56 lesions of 0-Is, and 12 flat lesions (0-IIa)). The median size of the lesions was 5 (2–20) mm (Table S1). The CAde system found 98.6% of the treated lesions. One diminutive adenomatous lesion that was 3 mm in size was not identified (Case #14, Lesion #2, Table S1). The median time between the initial appearance of the lesion and the CAde detection was 0.67 s (0.13–4.53) (Video S1).

4. Discussion

Preceding retrospective studies on DL-based CAde for colorectal lesions have shown excellent diagnostic performances with a sensitivity of 90–97.3% and specificity of 63.3–99% [16–19]. However, the study designs greatly varied, and the terms “sensitivity” and “specificity” used in these studies may be misleading, implying that they were used as statistical measures of a binary classification test on a fixed dataset, as is widely used in studies on endoscopic differential diagnosis. They could be optimal for demonstrating the

diagnostic performance of CADe in the analysis of stored still images because it is clinically acceptable to perform the characterization and differentiation of lesions by freezing the endoscopic motion and selecting the best quality images with a latency time. Meanwhile, the analysis and feedback of CADe need to be performed in real time to enable endoscopists to make timely decisions. Therefore, we believe that the binary classification test of selected stored datasets would be suboptimal to evaluate the diagnostic performance of CADe. Theoretically, the diagnostic performance of CADe could be analyzed per image, per lesion, per boundary box, or per pixel at any chronological phase during a colonoscopy. Different procedures were adopted in previous studies; some studies evaluated specificity using images that did not contain a lesion separately from the analysis of sensitivity with images of lesions. The negative clinical impact of false positives by CADe with low PPV tends to be underestimated in cases where sensitivity and specificity cannot be addressed equally. In this study, we primarily evaluated whether CADe could demonstrate lesion-positive frames configuring bounding boxes to calculate the diagnostic performances as binary variables in a 2×2 table. The trade-off between sensitivity and specificity calculated in a 2×2 table enabled an ROC curve to be plotted, and it allowed us to set the reasonable cutoff value of the probability threshold for each imaging mode at the Youden index. Even though the validation dataset used in this study was blindly sampled from video clips of the entire colonoscopy procedure, including poor quality images obtained during scope intubation, the CADe system still achieved an excellent AUC (0.983 in overall, 0.982 in WLI, 0.986 in CE, and 0.986 in NBI) and desirable diagnostic performances for every imaging mode. In addition, we found that the comparative diagnostic performances could be achieved with the same cutoff value of the probability threshold set with the ROC curve in overall images regardless of the imaging mode. It warranted the usage of the cutoff value of 0.27 for the lesion-based analysis using videos. However, there was a discrepancy between WLI and other imaging modes in the PPV. We surmised that it might be explained by the difference in the situations under which the imaging modes were used. CE and NBI were mostly used during the withdrawal of the endoscope and more often after the initial detection of the suspicious area with WLI. The results imply that the PPV could be a highly sensitive reference for further improving CADe, although the impact of the prevalence of the test data needs to be considered [33]. The definitions of TP and FP for the binary data analysis made it impossible to assess if the bounding boxes predicted by CADe were overlapped on lesions, but the low NPV and high localization accuracy demonstrated using the IoU analysis indicated that the CADe system could point out the location of the polyp.

As far as we know, there is no previous study measuring the latency from the initial appearance of lesions, as we conducted in this study. Hassan et al. evaluated the time gap in detecting colorectal lesions between CADe and endoscopists and demonstrated CADe detected lesions faster than endoscopists by an average of 1.27 s [34]. In this study, the latency of the detection by the CADe system from the initial appearance of lesions had a median of only 0.67 s, although the measures taken to decrease false positives required the analysis of five sequential frames to create each bounding box. The results prove that the CADe system would detect lesions much faster than humans.

There are several limitations to this study. The diagnostic performance in this study was only tested retrospectively in a small number of therapeutic cases with a high prevalence of the disease; screening and surveillance performance should be evaluated using a testing dataset comprising a different number of lesions. It was also difficult to evaluate whether the CADe system could find more lesions than those identified by the endoscopist in this retrospective study. However, the validation data automatically sampled from the whole endoscopic inspection images during the scope intubation consequently contained a lot of low-quality images, and we confirmed that the CADe system could detect lesions even in difficult locations in the still image analysis and immediately after the appearance in the movie analysis. We recognize the need for a prospective controlled study to clinically analyze validated quality indicators to demonstrate the advantage of

CADe over non-assisted human diagnosis. Meanwhile, the aim of this study was purely to conduct performance testing for the CADe system by using readily available stored videos of colonoscopies, analyzing outcomes, and mitigating operator selection biases. We believe that such validation methods would be more preferable in a developmental phase to decide the appropriate direction of research and also to reveal the strengths and weaknesses of CADe. In fact, a prospective multi-center randomized trial evaluating AMR as the primary outcome in a tandem study design, in which the CADe-assisted group achieved better AMR than the non-assisted group, was warranted from the desirable results of this retrospective trial [23].

In conclusion, the developed CADe system for assisting in lesion detection and localization achieved high sensitivity and specificity in fair testing settings, mitigating operator selection biases by automatically sampling a test data set from unedited videos of endoscopic observations, without eliminating low-quality images. Further, by reviewing the video clips of lesions, starting with the initial appearance during the colonoscopic inspection, it was demonstrated that the CADe system could detect most lesions that required removal in real time. Although the advantage of CADe should be validated in direct comparison with a human analyzing clinically approved quality indicators of the examination, such as ADR and AMR, we believe that the retrospective study under the strictly controlled environment, mitigating biases generally associated with research for computer-aided diagnosis, is valuable and should be refocused in this rapidly progressing research field.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/diagnostics11101922/s1>, Table S1: Latency of lesion detection using the computer-aided detection system in lesion-based analysis, Video S1: Colorectal lesion detection by the computer-aided detection system using video clips.

Author Contributions: Conceptualization, H.M., S.K., and K.S.; methodology, H.M., S.K., and K.S.; software, A.F., A.T., N.K., and Y.S.; validation, H.M. and K.S.; formal analysis, S.T. and M.N.; data curation, H.M., H.H., S.K., and N.T.; writing—original draft preparation, H.M.; writing—review and editing, N.T. and S.K.; supervision, K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by AMED under grant number JP17ck0106272.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the ethics committee of the Jikei University School of Medicine, Tokyo, Japan (no. 30-173(9194), 15 August 2019).

Informed Consent Statement: As a retrospective study, informed consent was obtained in the form of opting out on the website.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors are grateful to the engineers at LPixel Inc. (Tokyo, Japan) for their cooperation in developing the CADe system.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schreuders, E.H.; Ruco, A.; Rabeneck, L.; Schoen, R.E.; Sung, J.J.Y.; Young, G.; Kuipers, E.J. Colorectal cancer screening: A global overview of existing programmes. *Gut* **2015**, *64*, 1637–1649. [[CrossRef](#)] [[PubMed](#)]
2. Ladabaum, U.; Dominitz, J.A.; Kahi, C.; Schoen, R.E. Strategies for colorectal cancer screening. *Gastroenterology* **2020**, *158*, 418–432. [[CrossRef](#)]
3. Kahi, C.J.; Pohl, H.; Myers, L.J.; Mobarek, D.; Robertson, D.J.; Imperiale, T.F. Colonoscopy and colorectal cancer mortality in the veterans affairs health care system: A case-control study. *Ann. Intern. Med.* **2018**, *168*, 481–488. [[CrossRef](#)] [[PubMed](#)]
4. Nishihara, R.; Wu, K.; Lochhead, P.; Morikawa, T.; Liao, X.; Qian, Z.R.; Inamura, K.; Kim, S.A.; Kuchiba, A.; Yamauchi, M.; et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N. Engl. J. Med.* **2013**, *369*, 1095–1105. [[CrossRef](#)] [[PubMed](#)]

5. Zauber, A.; Winawer, S.; O'Brien, M.J.; Lansdorp-Vogelaar, I.; Van Ballegooijen, M.; Hankey, G.; Shi, W.; Bond, J.; Schapiro, J.; Panish, J.; et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N. Engl. J. Med.* **2012**, *366*, 687–696. [[CrossRef](#)] [[PubMed](#)]
6. Brenner, H.; Chang-Claude, J.; Seiler, C.M.; Rickert, A.; Hoffmeister, M. Protection from colorectal cancer after colonoscopy: A population-based, case-control study. *Ann. Intern. Med.* **2011**, *154*, 22–30. [[CrossRef](#)]
7. Robertson, D.J.; Lieberman, D.A.; Winawer, S.J.; Ahnen, D.J.; Baron, J.A.; Schatzkin, A.; Cross, A.J.; Zauber, A.G.; Church, T.R.; Lance, P.; et al. Colorectal cancers soon after colonoscopy: A pooled multicohort analysis. *Gut* **2014**, *63*, 949–956. [[CrossRef](#)] [[PubMed](#)]
8. Kaminski, M.; Wieszczy, P.; Rupinski, M.; Wojciechowska, U.; Didkowska, J.; Kraszewska, E.; Kobiela, J.; Franczyk, R.; Rupinska, M.; Kocot, B.; et al. Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death. *Gastroenterology* **2017**, *153*, 98–105. [[CrossRef](#)]
9. Corley, D.A.; Jensen, C.D.; Marks, A.; Zhao, W.K.; Lee, J.K.; Doubeni, C.; Zauber, A.G.; De Boer, J.; Fireman, B.H.; Schottinger, J.E.; et al. Adenoma detection rate and risk of colorectal cancer and death. *N. Engl. J. Med.* **2014**, *370*, 1298–1306. [[CrossRef](#)]
10. Kaminski, M.; Regula, J.; Kraszewska, E.; Polkowski, M.; Wojciechowska, U.; Didkowska, J.; Zwierko, M.; Rupinski, M.; Nowacki, M.P.; Butruk, E. Quality indicators for colonoscopy and the risk of interval cancer. *N. Engl. J. Med.* **2010**, *362*, 1795–1803. [[CrossRef](#)]
11. Atkinson, N.S.; Ket, S.; Bassett, P.; Aponte, D.; De Aguiar, S.; Gupta, N.; Horimatsu, T.; Ikematsu, H.; Inoue, T.; Kaltenbach, T.; et al. Narrow-band imaging for detection of neoplasia at colonoscopy: A meta-analysis of data from individual patients in randomized controlled trials. *Gastroenterology* **2019**, *157*, 462–471. [[CrossRef](#)]
12. Tziatzios, G.; Gkolfakis, P.; Lazaridis, L.D.; Facciorusso, A.; Antonelli, G.; Hassan, C.; Repici, A.; Sharma, P.; Rex, D.K.; Triantafyllou, K. High-definition colonoscopy for improving adenoma detection: A systematic review and meta-analysis of randomized controlled studies. *Gastrointest. Endosc.* **2020**, *91*, 1027–1036.e9. [[CrossRef](#)]
13. Subramanian, V.; Mannath, J.; Hawkey, C.J.; Ragnath, K. High definition colonoscopy vs. standard video endoscopy for the detection of colonic polyps: A meta-analysis. *Endoscopy* **2011**, *43*, 499–505. [[CrossRef](#)]
14. Wang, X.; Yang, W.; Weinreb, J.; Han, J.; Li, Q.; Kong, X.; Yan, Y.; Ke, Z.; Luo, B.; Liu, T.; et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: Deep learning versus non-deep learning. *Sci. Rep.* **2017**, *7*, 15415. [[CrossRef](#)]
15. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **2016**, *316*, 2402–2410. [[CrossRef](#)] [[PubMed](#)]
16. Misawa, M.; Kudo, S.-E.; Mori, Y.; Cho, T.; Kataoka, S.; Yamauchi, A.; Ogawa, Y.; Maeda, Y.; Takeda, K.; Ichimasa, K.; et al. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. *Gastroenterology* **2018**, *154*, 2027–2029.e3. [[CrossRef](#)] [[PubMed](#)]
17. Urban, G.; Tripathi, P.; Alkayali, T.; Mittal, M.; Jalali, F.; Karnes, W.; Baldi, P. Deep Learning Localizes and Identifies Polyps in Real Time with 96% Accuracy in Screening Colonoscopy. *Gastroenterology* **2018**, *155*, 1069–1078.e8. [[CrossRef](#)] [[PubMed](#)]
18. Wang, P.; Xiao, X.; Brown, J.R.G.; Berzin, T.M.; Tu, M.; Xiong, F.; Hu, X.; Liu, P.; Song, Y.; Zhang, D.; et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nat. Biomed. Eng.* **2018**, *2*, 741–748. [[CrossRef](#)]
19. Yamada, M.; Saito, Y.; Imaoka, H.; Saiko, M.; Yamada, S.; Kondo, H.; Takamaru, H.; Sakamoto, T.; Sese, J.; Kuchiba, A.; et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci. Rep.* **2019**, *9*, 14465. [[CrossRef](#)] [[PubMed](#)]
20. Vinsard, D.G.; Mori, Y.; Misawa, M.; Kudo, S.-E.; Rastogi, A.; Bagci, U.; Rex, D.K.; Wallace, M.B. Quality assurance of computer-aided detection and diagnosis in colonoscopy. *Gastrointest. Endosc.* **2019**, *90*, 55–63. [[CrossRef](#)]
21. Lui, T.K.L.; Guo, C.G.; Leung, W.K. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: A systematic review and meta-analysis. *Gastrointest. Endosc.* **2020**, *92*, 11–22.e6. [[CrossRef](#)]
22. Wang, P.; Berzin, T.M.; Brown, J.R.G.; Bharadwaj, S.; Becq, A.; Xiao, X.; Liu, P.; Li, L.; Song, Y.; Zhang, D.; et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study. *Gut* **2019**, *68*, 1813–1819. [[CrossRef](#)]
23. Kamba, S.; Tamai, N.; Saitoh, I.; Matsui, H.; Horiuchi, H.; Kobayashi, M.; Sakamoto, T.; Ego, M.; Fukuda, A.; Tonouchi, A. Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: A multicenter randomized controlled trial. *J. Gastroenterol.* **2021**, *56*, 746–757. [[CrossRef](#)]
24. Hassan, C.; Spadaccini, M.; Iannone, A.; Maselli, R.; Jovani, M.; Chandrasekar, V.T.; Antonelli, G.; Yu, H.; Areia, M.; Dinis-Ribeiro, M.; et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: A systematic review and meta-analysis. *Gastrointest. Endosc.* **2021**, *93*, 77–85.e6. [[CrossRef](#)] [[PubMed](#)]
25. Ashat, M.; Klair, J.S.; Singh, D.; Murali, A.R.; Krishnamoorthi, R. Impact of real-time use of artificial intelligence in improving adenoma detection during colonoscopy: A systematic review and meta-analysis. *Endosc. Int. Open.* **2021**, *9*, E513–E521. [[PubMed](#)]
26. Sumiyama, K.; Futakuchi, T.; Kamba, S.; Matsui, H.; Tamai, N. Artificial intelligence in endoscopy: Present and future perspectives. *Dig. Endosc.* **2021**, *33*, 218–230. [[CrossRef](#)] [[PubMed](#)]
27. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]

28. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
31. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. *Proc. ICML* **2013**, *30*, 3.
32. Perkins, N.J.; Schisterman, E.F. The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am. J. Epidemiol.* **2006**, *163*, 670–675. [[CrossRef](#)]
33. Pepe, M.S. *The Statistical Evaluation of Medical Tests for Classification and Prediction*; Oxford University Press: Oxford, UK, 2003.
34. Hassan, C.; Wallace, M.B.; Sharma, P.; Maselli, R.; Craviotto, V.; Spadaccini, M.; Repici, A. New artificial intelligence system: First validation study versus experienced endoscopists for colorectal polyp detection. *Gut* **2020**, *69*, 799–800. [[CrossRef](#)] [[PubMed](#)]