Videomics: bringing deep learning to diagnostic endoscopy

Alberto Paderno[a], *F. Christopher Holsinger*[b], *and Cesare Piazza*[a]

Purpose of review
Machine learning (ML) algorithms have augmented human judgment in various fields of clinical medicine. However, little progress has been made in applying these tools to video-endoscopy. We reviewed the field of video-analysis (herein termed 'Videomics' for the first time) as applied to diagnostic endoscopy, assessing its preliminary findings, potential, as well as limitations, and consider future developments.

Recent findings
ML has been applied to diagnostic endoscopy with different aims: blind-spot detection, automatic quality control, lesion detection, classification, and characterization. The early experience in gastrointestinal endoscopy has recently been expanded to the upper aerodigestive tract, demonstrating promising results in both clinical fields. From top to bottom, multispectral imaging (such as Narrow Band Imaging) appeared to provide significant information drawn from endoscopic images.

Summary
Videomics is an emerging discipline that has the potential to significantly improve human detection and characterization of clinically significant lesions during endoscopy across medical and surgical disciplines. Research teams should focus on the standardization of data collection, identification of common targets, and optimal reporting. With such a collaborative stepwise approach, Videomics is likely to soon augment clinical endoscopy, significantly impacting cancer patient outcomes.

Keywords
deep learning, endoscopy, machine learning, Narrow Band Imaging, neural networks, Videomics

INTRODUCTION

Artificial intelligence is beginning to transform clinical medicine in specialties where large datasets of annotated images are an essential element of the clinical workflow. The use of machine learning (ML) algorithms has augmented human judgment by identifying adverse events in the operating room [1], detect diabetic retinopathy [2], and even identify skin cancer [3]. From radiology to pathology, deep learning [4] has promise in reaching a diagnostic accuracy comparable with that of human experts from automating the detection of pneumonia on chest roentgenograms [5] and CT scans [6], to identifying clinically occult nodal metastasis in breast cancer [7]. This rapid pace of innovation suggests that the development of expert systems will soon be applicable in everyday clinical settings, providing real-time assistance to the physician in a variety of diagnostic tasks, using computer technology to improve the human vision and judgment.

Notwithstanding these promises, at present, little progress has been made in applying deep learning algorithms to video-endoscopy, which plays a prominent role in otorhinolaryngology, head and neck surgery, pulmonary medicine, and gastroenterology, as well as in thoracic and abdominal surgery. For the purpose of this review, we explore how automated analysis of unstructured data obtained by video-endoscopy can provide valuable information during initial diagnosis, measuring treatment-response, and assessing prognosis. Endoscopic evaluation has always been a crucial component of head and neck oncology (HNO), since tumor

[a]Unit of Otorhinolaryngology – Head and Neck Surgery, ASST Spedali Civili of Brescia, Department of Medical and Surgical Specialties, Radiological Sciences, and Public Health, University of Brescia, Brescia, Italy and [b]Division of Head and Neck Surgery, Stanford University, Palo Alto, California, USA

Correspondence to Chris Holsinger, MD, FACS, Division of Head and Neck Surgery, Stanford University, 875 Blake Wilbur Drive, Stanford, California 94305, USA. E-mail: holsinger@stanford.edu

Curr Opin Otolaryngol Head Neck Surg 2021, 29:143–148
DOI:10.1097/MOO.0000000000000697

KEY POINTS

- The rapid pace of innovation suggests that the development of expert systems will soon be applicable in everyday clinical settings, providing real-time assistance to the physician in a variety of diagnostic tasks.
- We herein propose the term 'Videomics' to denote a burgeoning field where several methods of computer vision and deep learning are systematically used to interpret the unstructured data of video obtained during diagnostic video-endoscopy.
- As of today, efforts have been directed toward different lines of analysis: (1) blind-spot detection and automatic quality control; (2) lesion detection; (3) lesion classification; and (4) lesion characterization.
- Video-endoscopic evaluation of the upper aerodigestive tract poses even more challenges than gastrointestinal endoscopy since this anatomic region is structurally more complex, composed of a wide variety of tissues, and easily shaded.
- Research teams should focus on standardization of data collection, identification of common targets, and optimal reporting.

superficial spread assessment has a significant impact on treatment selection and may not be adequately quantified by conventional radiologic imaging. Often seen as inherently descriptive, deep learning has helped to convert subjective assessment into objective findings based on systematic evaluation of visual data seen on video, analogous to findings obtained by conventional techniques in genomics and proteomics.

Here, in fact, we propose the term 'Videomics' as a burgeoning field wherein several methods of computer vision and deep learning are systematically used to organize the unstructured data of video obtained during diagnostic endoscopy. Indeed, in this review of the literature, we argue that based on a growing number of publications, a new discipline is emerging within the large field of computer vision and pattern recognition. Herein, we review this promising new field, assessing preliminary findings, potential and limitations, and consider future developments.

MACHINE LEARNING IN ENDOSCOPY

Because of the higher caseload compared with HNO, gastrointestinal endoscopy was the first field in which ML was effectively applied. For this reason, it is useful to first analyze the progress in this branch of endoscopy to identify potential advances that are applicable to evaluation of the upper aerodigestive tract (UADT). Even in this broader research field, reports assessing the role of ML are scarce. As of today, efforts have been directed toward different lines of analysis, in particular: (1) blind-spot detection and automatic quality control; (2) lesion detection; (3) lesion classification; and (4) lesion characterization. This approach is strictly related to the perceived needs in digestive system endoscopy, and has given promising results and potential real-life applications.

Concerning blind-spot detection, Wu *et al.* [8] developed a convolutional neural network (CNN)-based system aimed at detecting early gastric cancer although avoiding blind-spots during esophagogastroduodenoscopy. The algorithm was trained to identify the different subsites of the esophagus and stomach to ensure the complete visualization of the entire gastroesophageal mucosa. Furthermore, CNN was trained to distinguish between normal mucosa and early gastric cancer. In both tasks, the accuracy was remarkable (>90%). The same authors [9▪] validated the efficacy of this blind-spot detection system in a randomized controlled trial, showing a significantly lower blind-spot rate in CNN-assisted endoscopy vs. a control group (5.9% vs. 22.5%).

In the same perspective, Su *et al.* [10] developed an automatic quality control system aimed at improving diagnostic accuracy during colonoscopy. The system was based on CNN models for timing the withdrawal phase, supervise withdrawal stability, evaluate bowel preparation, and detect colorectal polyps. A randomized controlled trial showed that this CNN-based quality control system significantly increased lesion detection (adenomas and polyps) during colonoscopy compared to that without CNN assistance.

Lesion detection and characterization remain the main objective of ML-based strategies in gastrointestinal endoscopy. Texture analysis has shown good preliminary results in detecting mucosal abnormalities (e.g., colon polyps) [11], and CNNs proved to be a key instrument in this field. In fact, the vast majority of recent reports on automatic lesion detection and classification have taken advantage of this algorithm architecture. Different authors have described its significant potential in the detection and diagnosis of gastric, esophageal, and small bowel cancers, as well as gastrointestinal polyps [12–16]. Furthermore, CNNs are also useful in classification tasks, distinguishing between normal and inflamed mucosa (gastritis), and identifying early gastric cancer using magnifying endoscopy [17–19].

Interestingly, although some authors employed conventional white light (WL) endoscopy, most

studies have applied ML evaluation to Narrow Band Imaging (NBI) pictures or videos. In this view, multispectral imaging may have the potential to further improve detection and characterization of mucosal lesions in the field of automatic analysis, adding more definition to tumor margins and highlighting features of submucosal vascularization that are not evident during WL endoscopy. In 2012, Takemura *et al.* [20] demonstrated the value of NBI in the classification of colonoscopy magnified images using support vector machines. This aspect was explicitly investigated by Horie *et al.* [14], who reported that NBI had a higher sensitivity compared with conventional WL endoscopy (although not reaching a statistically significant difference).

Finally, ML has shown promise in the in-depth characterization of known lesions of the gastrointestinal tract and may also provide risk stratification for malignant transformation of nonneoplastic mucosa. Specifically, recent studies [21,22] have demonstrated that CNNs can differentiate between early and deeply infiltrating gastric cancer. This result shows the potential of Videomics approaches to go beyond simple diagnosis and extract more extensive information on the lesion itself. Nakahira *et al.* [22] further confirmed this potential by showing that CNN is able of correctly stratify the risk of gastric tumor development by analyzing the nonneoplastic mucosa at video-endoscopy.

MACHINE LEARNING APPLICATIONS IN UPPER AERO-DIGESTIVE TRACT ENDOSCOPY

Video-endoscopic evaluation of the UADT poses even more challenges than gastrointestinal endoscopy [23]. This anatomic region is, in fact, structurally more complex, composed of a wide variety of tissues [24,25], and easily shaded. Furthermore, deglutition, gag, and cough reflexes often come into play, interrupting or limiting the observation. The oral cavity and oropharynx are the most accessible sites; however, their video-endoscopic evaluation is not standardized and may be performed using rigid or flexible endoscopes, or even external cameras. Therefore, this factor adds an adjunctive layer of complexity to image analysis since data collection should be ideally standardized and characterized by low variance.

Oral cavity and oropharynx

Different authors [26,27] have recognized the value of ML in the evaluation and screening of oral cancer and potentially malignant lesions. Song *et al.* [28] developed a smartphone-based automatic image classification system for oral dysplasia and malignancy employing CNNs. The system aimed to screen high-risk populations in middle- and low-income countries and took advantage of dual-modality images (WL and autofluorescence). The authors demonstrated the potential of dual-modal image analysis, which showed better diagnostic performance than single-modal images. The final model reached an accuracy of 87%, sensitivity of 85%, and specificity of 89%.

Mascharak *et al.* [29▪] were the first to use ML to better identify oropharyngeal tumor margins using a simple naïve Bayesian classifier (color and texture). Interestingly, the diagnostic performance was significantly enhanced by multispectral NBI compared with conventional WL video-endoscopy. Five-fold cross-validation yielded an area under the curve (AUC) above 80% for NBI models and below 55% for WL endoscopy models ($P < 0.001$).

Finally, Paderno *et al.* [30▪] published preliminary data showing that it is possible to obtain real-time oral and oropharyngeal tumors segmentation using different fully CNNs applied to NBI endoscopic images, identifying potential confounding factors and technical drawbacks.

Larynx and hypopharynx

In general, laryngo-pharyngeal lesions are those more frequently investigated when assessing the role of automatic analysis by ML. This is due to use of a standardized endoscopic approach through trans-nasal or transoral video-endoscopy and the relative similarity with gastrointestinal subsites. In 2014, Huang *et al.* [31] proposed an automatic system aimed at recognizing images of the glottis and classifying different vocal fold disorders. The technique was based on a support vector machine classifier and reached an accuracy of 99%. However, the patterns to be classified were limited to 'normal vocal fold,' 'vocal fold paralysis,' 'vocal fold polyp,' and 'vocal fold cyst,' and did not include dysplasia or malignancy.

A preliminary attempt at automatic detection and classification of laryngeal tumors has been described by Barbalata *et al.* [32]. The authors used anisotropic filtering to analyze the submucosal vasculature of normal and neoplastic laryngeal mucosa during NBI video-endoscopic examination, obtaining an overall classification accuracy of 83%. Although not employing adaptive algorithms, the study confirmed the value of NBI in maximizing feature extraction in endoscopic images.

Subsequent studies, focusing on the diagnosis and classification of pharyngo-laryngeal lesions at video-endoscopy, extensively employed CNNs and

demonstrated remarkable results. A work by Laves *et al.* [33] used CNNs to segment a novel 7-class (void, vocal folds, other tissue, glottal space, pathology, surgical tools, and tracheal tube) dataset of the human larynx during transoral laser microsurgery. The dataset, consisting of 536 manually segmented endoscopic images, was tested to monitor the morphological changes and autonomously detect pathologies. Different CNN architectures were investigated, and a weighted average ensemble network of UNet and ErfNet (two of the most used CNNs in the current literature on this topic) was the best suited for laryngeal segmentation with a mean Intersection-over-Union (IoU) evaluation metric of 84.7%.

Xiong *et al.* [34] developed a CNN-based diagnostic system trained using 13,721 laryngoscopic images of cancer, premalignant lesions, benign alterations, and normal tissue collecting exams across several centers in China. The CNN distinguished malignant/premalignant lesions from benign ones and normal tissues with an accuracy of 87% (sensitivity 73%, specificity 92%, and AUC 92%). Ren *et al.* [35] described a similar approach, training the CNN with a total of 24,667 laryngoscopy images (normal, vocal nodule, polyps, leukoplakia, and malignancy), and achieving an overall accuracy of 96%. Strikingly, the CNN-based classifier outperformed physicians in the evaluation of the abovementioned conditions.

Further detection and classification attempts have mainly taken advantage of NBI images, which yielded superior results in terms of diagnostic performance, as previously demonstrated by Mascharak *et al.* [29▪] in the oropharyngeal site, and confirmed by Tamashiro *et al.* [36▪]. These studies [36▪–38▪] were performed in the setting of transoral esophagogastroduodenoscopy and were aimed at detecting incidental laryngo-pharyngeal cancer during the procedure. However, direct comparisons between the different studies may be misleading because of the heterogeneous definition of 'correct diagnosis.'

Tamashiro *et al.* [36▪] focused on pharyngeal cancer and reported an accuracy, sensitivity, and specificity of 67%, 80%, and 57%, respectively. These results were slightly improved when limiting the analysis to NBI frames only. The authors trained a 'Single Shot MultiBox Detector' with a total of 5,403 images. Adequate detection was considered as frames including less than 80% of the area with noncancerous sites. Kono *et al.* [37▪] showed similar results in pharyngeal cancer detection by using a mask region-based CNN trained with 4,559 images. Each frame was judged as cancer when its probability score was $\geq$0.60, and its dimensions overlapped with the cancer area by a factor of $\geq$0.20. Accuracy, sensitivity, and specificity were 66%, 92%, and 47%, respectively.

Finally, Inaba *et al.* [38▪] trained a CNN-based algorithm (RetinaNet) with sequential sets of images until reaching 400 frames of superficial laryngopharyngeal cancer and 800 frames of normal mucosa. The diagnostic accuracy gradually improved with the sequential addition of training images until reaching an accuracy, sensitivity, and specificity of 97%, 95%, and 98%, respectively. The definition of correct diagnosis was set with an IoU parameter $>$0.4.

FUTURE PERSPECTIVES AND CONCLUSIONS

Videomics is an emerging discipline that has the potential to significantly improve human detection of clinically significant lesions during video-endoscopy across medical and surgical disciplines. Preliminary reports have shown promising diagnostic potential and demonstrated the ability of ML algorithms to provide adjunctive information on tumor characteristics, such as depth of infiltration and, hence, infer important tumor-related issues such as extra-visceral extension, submucosal spread, and risk of regional/distant metastases. However, as early 'proof-of-concept' studies are published, it is important to note that these efforts are not yet part of routine endoscopic examination. In this view, further advances may allow obtaining an ever-growing amount of data from video-endoscopic sequences, thus assisting in tumor staging, margin recognition, treatment planning, and prognostic assessment. Furthermore, features extracted from video-endoscopy may be integrated into broader '-omic' models (including radiomics, genomics, proteomics, salivaomics, etc.), thus creating a precise representation of a given tumor and/or fine-tune the assessment of a specific patient. This is a crucial step in the perspective of tailoring treatment and personalized medicine.

For Videomics to flourish and to deliver practical tools for clinicians in daily practice, it is imperative to create large-scale image and video repositories. Currently, many ongoing efforts are fragmented and highly variable in their approach: the anatomical regions investigated (upper or lower digestive tracts), quality of images (definition, focus, illumination, and color balance), type of spectral filters (WL, NBI, autofluorescence, others), and setting (office-based, intraoperative) vary widely. Nonetheless, the current bottleneck for the development of ML-based video-analysis techniques is represented by the need to manually annotate training images.

In this view, the development of self-supervised learning techniques using unlabeled data for CNN pretraining and training may significantly and progressively improve algorithms without the need for human intervention [39].

Finally, study objectives (detection, classification, or segmentation) are often not clearly stated or distinguished in each study. Last but not least, the statistical definition of correct and incorrect diagnosis is subjectively determined by each author, leading to significant variation in diagnostic performance metrics (e.g., accuracy, sensitivity, specificity, and AUC). For this reason, at this early stage in the field, research teams should focus on standardization of data collection, identification of common targets, and optimal reporting. With such a collaborative stepwise approach, Videomics is likely soon augment human detection during endoscopy and improve cancer treatment and subsequent outcomes.

Acknowledgements

None.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest

1. Gordon L, Austin P, Rudzicz F, Grantcharov T. MySurgeryRisk and machine learning: a promising start to real-time clinical decision support. Ann Surg 2019; 269:e14–e15.
2. Gulshan V, Peng L, Coram M, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016; 316:2402–2410.
3. Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542:115–118.
4. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521:436–444.
5. Rajpurkar P, Irvin J, Ball RL, *et al.* Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018; 15:e1002686.
6. Harmon SA, Sanford TH, Xu S, *et al.* Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. Nat Commun 2020; 11:4080.
7. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017; 318:2199–2210.
8. Wu L, Zhou W, Wan X, *et al.* A deep neural network improves endoscopic detection of early gastric cancer without blind spots. Endoscopy 2019; 51:522–531.
9. ■ Wu L, Zhang J, Zhou W, *et al.* Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut 2019; 68:2161–2169.

Randomized clinical trial demonstrating for the first time the clinical efficacy of CNN-assisted endoscopy in blind-spot detection in a real-life setting.

10. Su JR, Li Z, Shao XJ, *et al.* Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointest Endosc 2020; 91:415–424.
11. Geetha K, Rajan C. Automatic colorectal polyp detection in colonoscopy video frames. Asian Pac J Cancer Prev 2016; 17:4869–4873.
12. Barbosa DC, Roupar DB, Ramos JC, *et al.* Automatic small bowel tumor diagnosis by using multiscale wavelet-based analysis in wireless capsule endoscopy images. Biomed Eng Online 2012; 11:3.
13. Billah M, Waheed S, Rahman MM. An automatic gastrointestinal polyp detection system in video endoscopy using fusion of color wavelet and convolutional neural network features. Int J Biomed Imaging 2017; 2017:Article ID 9545920.
14. Horie Y, Yoshio T, Aoyama K, *et al.* Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. Gastrointest Endosc 2019; 89:25–32.
15. Hirasawa T, Aoyama K, Tanimoto T, *et al.* Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018; 21:653–660.
16. Yoon HJ, Kim JH. Lesion-based convolutional neural network in diagnosis of early gastric cancer. Clin Endosc 2020; 53:127–131.
17. Ueyama H, Kato Y, Akazawa Y, *et al.* Application of artificial intelligence using a convolutional neural network for diagnosis of early gastric cancer based on magnifying endoscopy with narrow-band imaging. J Gastroenterol Hepatol 2020. doi: 10.1111/jgh.15190. Online ahead of print.
18. Li L, Chen Y, Shen Z, *et al.* Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. Gastric Cancer 2020; 23:126–132.
19. Horiuchi Y, Aoyama K, Tokai Y, *et al.* Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging. Dig Dis Sci 2020; 65:1355–1363.
20. Takemura Y, Yoshida S, Tanaka S, *et al.* Computer-aided system for predicting the histology of colorectal tumors by using narrow-band imaging magnifying colonoscopy (with video). Gastrointest Endosc 2019; 8:1310.
21. Yoon HJ, Kim S, Kim JH, *et al.* A Lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer. J Clin Med 2019; 8:1310.
22. Nakahira H, Ishihara R, Aoyama K, *et al.* Stratification of gastric cancer risk using a deep neural network. JGH Open 2020; 4:466–471.
23. Abe S, Oda I. Real-time pharyngeal cancer detection utilizing artificial intelligence: Journey from the proof of concept to the clinical use. Dig Endosc 2020. doi: 10.1111/den.13833. Online ahead of print.
24. Lin YC, Wang WH, Lee KF, *et al.* Value of narrow band imaging endoscopy in early mucosal head and neck cancer. Head Neck 2012; 34:1574–1579.
25. Piazza C, Del Bon F, Paderno A, *et al.* The diagnostic value of narrow band imaging in different oral and oropharyngeal subsites. Eur Arch Otorhinolaryngol 2016; 273:3347–3353.
26. Yoshida K. Future prospective of light-based detection system for oral cancer and oral potentially malignant disorders by artificial intelligence using convolutional neural networks. Photobiomodul Photomed Laser Surg 2019; 37:195–196.
27. Kar A, Wreesmann VB, Shwetha V, *et al.* Improvement of oral cancer screening quality and reach: the promise of artificial intelligence. J Oral Pathol Med 2020; 49:727–730.
28. Song B, Sunny S, Uthoff RD, *et al.* Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning. Biomed Opt Express 2018; 9:5318–5329.
29. ■ Mascharak S, Baird BJ, Holsinger FC. Detecting oropharyngeal carcinoma using multispectral, narrow-band imaging and machine learning. Laryngoscope 2018; 128:2514–2520.

First study to demonstrate the value of maching learning to identify tumor-normal interface of oropharyngeal tumors including an 'unknown' primary. Narrow-band imaging appeared to enhance the ability of deep learning to identify clinically significant differences, based not only upon color, but also by using measures of texture or 'entropy.'

30. ■ Paderno P, Piazza C, Del Bon F, *et al.* Deep learning for automatic segmentation of oral and oropharyngeal cancer using Narrow Band Imaging: Preliminary experience in a clinical perspective. Front Oncol 2020. (in press).

First study ever showing preliminary data on the possibility to obtain real-time oral and oropharyngeal tumors segmentation using different fully CNNs applied to NBI video-endoscopic images, thus identifying potential confounding factors and basic technical drawbacks.

31. Huang CC, Leu YS, Kuo CF, *et al.* Automatic recognizing of vocal fold disorders from glottis images. Proc Inst Mech Eng H 2014; 228:952–961.
32. Barbalata C, Mattos LS. Laryngeal tumor detection and classification in endoscopic video. IEEE J Biomed Health Inform 2016; 20:322–332.
33. Laves MH, Bicker J, Kahrs LA, Ortmaier T. A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation. Int J Comput Assist Radiol Surg 2019; 14:483–492.
34. Xiong H, Lin P, Yu JG, *et al.* Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. EBioMedicine 2019; 48:92–99.
35. Ren J, Jing X, Wang J, *et al.* Automatic recognition of laryngoscopic images using a deep-learning technique. Laryngoscope 2020; 130:E686–E693.

36. Tamashiro A, Yoshio T, Ishiyama A, *et al.* Artificial intelligence-based detection ■ of pharyngeal cancer using convolutional neural networks. Dig Endosc 2020; 32:1057–1065.
Study assessing the potential of CNNs in hypopharyngeal cancer detection, using WL and NBI images.
37. Kono M, Ishihara R, Kato Y, *et al.* Diagnosis of pharyngeal cancer on ■ endoscopic video images by Mask region-based convolutional neural network. Dig Endosc 2020; doi: 10.1111/den.13800. [Online ahead of print]
Study confirming the value of CNNs in hypopharyngeal cancer detection with a wide number of images.
38. Inaba A, Hori K, Yoda Y, *et al.* Artificial intelligence system for detecting ■ superficial laryngopharyngeal cancer with high efficiency of deep learning. Head Neck 2020; 42:2581–2592.
Research demonstrating remarkable diagnostic results in the automatic detection of hypopharyngeal cancer using CNNs (under WL and NBI). The authors demonstrated a significant improvement of results with the incremental addition of more training images.
39. Ross T, Zimmerer D, Vemuri A, *et al.* Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. Int J Comput Assist Radiol Surg 2018; 13:925–933.