

REVIEW ARTICLE

상부위장관 질환의 딥러닝 적용

방창석

한림대학교 의과대학 내과학교실

Deep Learning in Upper Gastrointestinal Disorders: Status and Future Perspectives

Chang Seok Bang

Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Korea

Artificial intelligence using deep learning has been applied to gastrointestinal disorders for the detection, classification, and delineation of various lesion images. With the accumulation of enormous medical records, the evolution of computation power with graphic processing units, and the widespread use of open-source libraries in large-scale machine learning processes, medical artificial intelligence is overcoming its traditional limitations. This paper explains the basic concepts of deep learning model establishment and summarizes previous studies on upper gastrointestinal disorders. The limitations and perspectives on future development are also discussed. (*Korean J Gastroenterol* 2020;75:120-131)

Key Words: Artificial intelligence; Neural networks, computer; Deep learning; Gastroenterology; Endoscopy

서론

통계는 모집단을 추정하고자 하는 학문이다. 임상 연구는 설정된 귀무가설을 검증하기 위하여 중심 극한 정리(central limit theorem)와 대수의 법칙(law of large number)을 가정하여 모집단에서 획득된 표본을 대상으로 한다. 이들 표본을 분석하여(주로는 대조군과 비교 분석) 그 결과인 통계량(statistic)을 통하여 모수(population parameter)를 추정(estimation)하여 가설을 검증하는 것이 임상 연구의 목적이다. 특정 질환의 정밀한 진단이나 예후 예측 등을 위하여 대규모의 데이터를 획득하여 설명력이 높은 통계학적 모델을 수립하고자 하는 시도가 이루어져 왔다. 이는 주로 회귀분석 등을 이용하였으며 그 설명력은 Nagelkerke의 결정 계수(R-square)를 주로 이용하였고, 0부터 1 사이의 값으로 그 수치가 높을수록 모델의 설명력이 높다고 해석하였다. 하지만 복잡계 이론을 따를

것으로 예상되는 실제 현상들을 통계학적인 모델로 설명해보면 모델에 사용된 변수의 종류나 사용된 표본의 숫자 또는 표본이 얼마나 현실을 반영하는 지에 따라 전체적인 모델의 설명력이 달라진다. 따라서 모델의 낮은 설명력에도 불구하고 출판되는 연구들이 존재하고 소위 과소적합(underfitting)되어 있는 모델들이 존재한다. 하지만 정밀하고 복잡한 통계학적 모델을 개발하여 현상을 예측하더라도 과대적합(overfitting)되어 있는 모델 역시 존재하며, 이 모델 또한 학습용 데이터(training data)로는 적합한 값을 예측하지만 새로운 검증용 데이터(validation data)로는 적합한 값을 도출하지 못하는 경우가 있다. 이러한 통계학적 모델을 실제 진료에 사용하기 위해서는 모델 수립에 사용되지 않은 데이터를 통한 외부 검증(external validation)을 시행한 후 개발에 사용된 결과와의 차이를 확인하는 것이 필수적이다.¹

인공신경망(artificial neural network) 분석은 1950년대에

Received February 13, 2020. Revised March 1, 2020. Accepted March 2, 2020.

© This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © 2020. Korean Society of Gastroenterology.

교신저자: 방창석, 24252, 춘천시 한림대학길 1, 한림대학교 의과대학 내과학교실

Correspondence to: Chang Seok Bang, Department of Internal Medicine, Hallym University College of Medicine, 1 Hallymdaehak-gil, Chuncheon 24252, Korea.
Tel: +82-33-240-5821, Fax: +82-33-241-8064, E-mail: csbang@hallym.ac.kr, ORCID: <https://orcid.org/0000-0003-4908-5431>

Financial support: None. Conflict of interest: None.

개발되었지만 근래에 대규모의 데이터와 고성능의 컴퓨터 graphical processing unit 및 Tensorflow, Pytorch, Keras 등의 오픈소스 라이브러리들의 활용이 가능해지면서 다시 주목 받는 머신러닝(machine learning) 기법이다. 이는 퍼셉트론(perceptron, artificial neuron with linear classifier) 기반의 간단한 논리연산 시스템에서 출발하여 입력값(feature data)이 가중치를 가지는 중간층의 노드(node, 변환함수)를 거쳐 출력값(labeling data)을 도출하는 방식이다(Fig. 1). 즉, 입력값과 출력값을 설정하고 학습을 거쳐 새로운 입력값에 대하여 출력값을 정확히 추론하고자 하는 분석 방법이다. 여러 단계의 중간층(hidden layers)을 배열하는 딥러닝(deep learning) 기법의 발전과 함께 심화된 알고리즘(deep neural network, DNN)의 구현이 가능하게 되었다(Fig. 2). 회귀분석이 독립 변수에 해당하는 입력값과 종속 변수에 해당하는 출력값 사이의 관계에 대하여 중간층의 개입 없이 직접적인 함수로 표현을 한다면, DNN은 다수의 복잡한 가중치를 가지는 중간층의 분석이 혼합되어 입력값에 대하여 출력값을 근사화(approximation)하여 도출한다. 회귀분석이 모형에 적합한 변수를 선택(variable selection)하는 통계적인 분석법이라면 DNN은 입력값의 어떤 특징을 선택(feature selection)하는

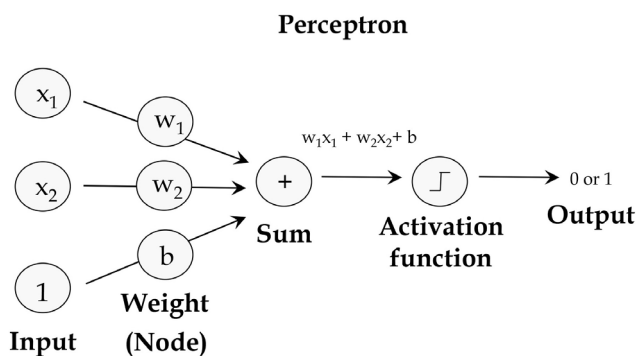


Fig. 1. Schematic view of perceptron.

분석법이다. 따라서 회귀분석처럼 정리된 수식이나 위험도(risk ratio), 오즈비(OR) 등으로는 설명이 불가능하지만 전체적인 모델의 정확도는 더욱 향상되는 결과를 흔히 도출한다. 즉, DNN 분석은 전체적인 모델의 설명력을 강조한 응용 통계 기법 중 하나이다.

이들 알고리즘 중에서도 이미지를 학습하고 분류(classification)하는데 가장 많이 사용되고 있는 것은 convolutional neural network (CNN, ConvNet)이다. 이는 이미지를 특정 filter (kernel)를 통하여 수학적 연산을 시행하여 특징을 추출하고 pooling 과정(여러 convolution layer에서 한 layer 씩 뽑아 사이즈를 축소하는데, 예를 들면 4x4 pixel의 이미지를 2x2 사이즈로 줄이는 것이고, filter를 이용한 convolution 연산[합성곱 연산]의 값 중 가장 큰 값을 고르는 것은 max pooling)을 통하여 차원(dimension)을 축소하여 학습한다. Convolution 연산과 pooling 과정을 통하여 추출된 activation map이라는 이미지의 특징을 나타내는 수치가 CNN 모델의 입력값으로 적용되어 softmax regression 등의 함수 연산을 통하여 이미지 분류의 역할을 수행한다. DNN과의 차이는 convolution과 max pooling 과정으로 최근에는 이미지 분류 이외에도 다양한 분석에 기본으로 사용되고 있다(Fig. 3). 본고에서는 인공지능 알고리즘 개발의 일반적인 내용에 대하여 소개하고, 최근의 방대한 인공지능 관련 연구의 증가를 고려하여 상부위장관 질환에 국한하여 인공지능의 의료 적용 현황과 향후 발전 방향에 대하여 논하고자 한다.

본 론

빅데이터는 흔히 전체 데이터의 양이 많은 것을 뜻하는 개념이었지만 데이터의 생산이 빠르고 대량화되면서 변수가 많은 데이터로 개념이 바뀌고 있다. 상부위장관 질환 중에서는 특히 내시경 이미지에 CNN이 대부분 적용이 되었는데 이미

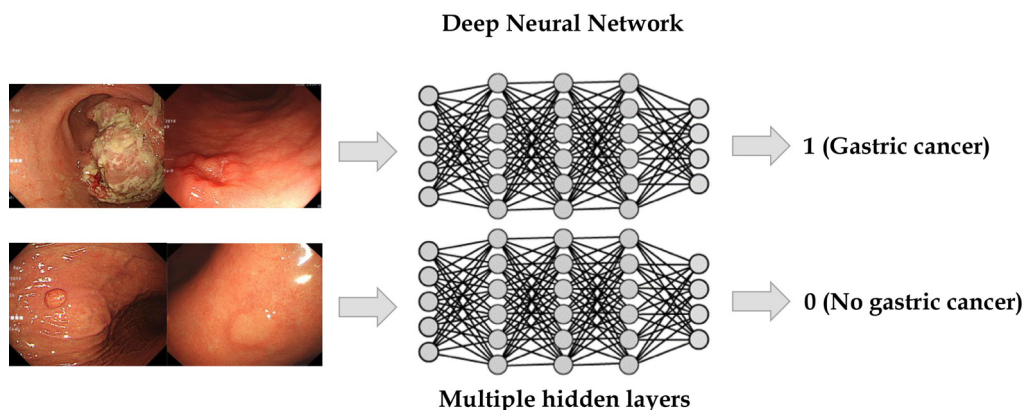


Fig. 2. Schematic view of a deep neural network.

지 분류(image classification) 분야에 주로 적용이 되었다. 이외에도 Region-based CNN이나 You Only Look Once (YOLO) 등의 네트워크를 이용한 병변의 발견(lesion detection and classification with localization) 그리고 fully convolutional networks를 이용한 병변의 범위를 지정(lesion segmentation, delineation)해주는 방식으로 적용이 시도되었다. 이들 기능을 혼합한 방식으로 작동하는 모델 또한 개발이 되었으며, 이외에도 임상 결정에 도움을 주는 의사 결정 보조(decision support)의 기능도 수행이 가능하고, 이미지 이외에도 비디오 영상에 위의 기능들(real-time diagnosis)을 구현하고자 하였다.^{1,2} 인공지능의 임상 적용에 여러 제한점이 극복되었는데, 그중 언어 또한 제한사항 중 하나였다. 하지만 순환신경망(recurrent neural network)을 이용한 자연어 처리(natural language processing)와 음성 인식(speech recognition) 기술의 발전으로 언어를 기계가 인식하고 처리하는 것이 가능하게 되었다. 이는 언어를 이해하고 분석하는데 그치지 않고 생성하는 것 또한 가능하기 때문에 기계와 사람 사이의 의사소통 기술로 발전이 되었다. 비정형 텍스트에서 의미가 있는 정보를 찾기도 하는데 최근에는 대장 내시경 기록지와 병리 결과 기록지를 자연어 처리 기술로 분석하여 비교적 정확한 대장선종 발견율(adenoma detection rate)을 자동으로 분석하는 연구가 발표되어 실제 임상에 바로 적용이 가능하리라 생각된다.³ 영상 또한 마찬가지로 위의 기능들을 적용할 수 있으며, generative adversarial network와 같은 모델이 이미지나 영상을 분석하는데 그치지 않고 존재하지 않는 새로운 이미지를 생성하는 것 또한 가능하다. 즉, 이는 어떤 분류를 위한 판별모델(discriminative model)의 특징만을 가지는 것이 아니고 생성모델(generative model)의 특징 또한 가지고 있다.

1. 인공지능망 학습모델 개발

일반적으로 의료 인공지능 모델을 개발하기 위하여 우선은

가설을 수학적으로 설정하고 데이터를 학습용 데이터(training dataset)와 시험용 데이터(test dataset)로 나누어 분석한다. 인공지능망 학습 가설의 성능을 측정할 수 있는 지표로 손실 함수(loss function) 또는 비용 함수(cost function)라는 것을 설정하여 이것을 최소화할 수 있는 최적의 매개변수(parameter) 가중치 값을 찾는 것(최적화 기법, optimization)이 모델을 수립하는 과정이다. 일반적인 인공지능망 학습모델의 개발은 입력값과 결과값을 가지고 이들 사이의 논리적인 관계에 대한 함수적인 표현을 모르는 상태에서 컴퓨터 알고리즘이 데이터로부터 이러한 논리적인 관계를 학습하는 데이터 의존적인 방식(end-to-end learning)이다(Fig. 4).

대표적인 손실 함수로 연속형 변수에서 사용하는 mean of squared error나 분류에서 사용이 가능한 cross entropy error와 같은 함수가 있다. 신경망에서 가중치 연산(입력값에 가중치를 곱하고 bias를 더하는 연산)을 하는 일반 노드와는 달리 활성화 함수(activation function)는 신경망 모델에 비선형성(non-linearity)을 제공하는 역할로 입력값이 일정량 이상이 되면 활성화되어 결과값을 생성한다. Sigmoid 함수, rectified linear unit, softmax regression, hyperbolic tangent 등이 있다. 활성화 함수를 통하여 도출된 예측 수치는 실제 수치와의 차이(error term)를 계산하여 각각의 신경망

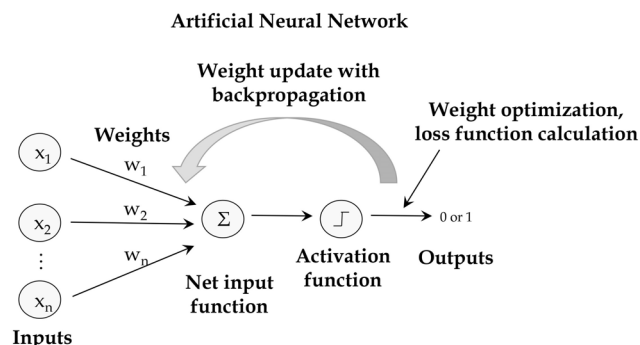


Fig. 4. Mechanistic scheme of an artificial neural network.

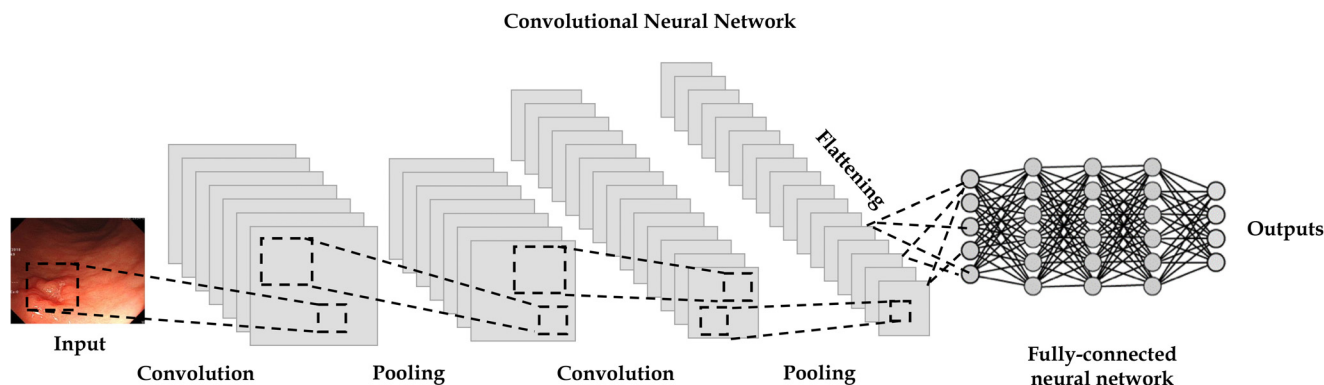


Fig. 3. Schematic view of a convolutional neural network.

Table 1. Summary of Clinical Studies Using Machine Learning in the Barrett's Esophagus

Study	Aim of study	Design of study	Number of subjects	Type of artificial intelligence	Modality	Outcomes
de Groof et al. (2019) ⁴	Endoscopic classification and segmentation of early neoplasia in patients with Barrett's esophagus	Retrospective	Pretraining: 494,364 images. Refinement training: 1,247 images. Refinement training and internal validation: 297 images. External validation: 160 images	CNN (hybrid ResNet/U-Net)	White-light endoscopy	Accuracy in the test dataset: 89%. Accuracy in the external validation: 88% (vs. 73% in the endoscopists). The algorithm identified the optimal site for biopsy of detected neoplasia in 92-97% of cases.
de Groof et al. (2019) ⁵	Endoscopic detection and localization of early neoplasia in patients with Barrett's esophagus	Prospective	40 neoplastic Barrett's lesions and 20 non-dysplastic Barrett's esophagus	SVM	White-light endoscopy	Accuracy: 92%
Struyvenberg et al. (2019) ¹⁰	Automatic data extraction followed by computer-aided diagnosis using a multiframe approach for detection of Barrett's esophagus neoplasia	Prospective	3,060 volumetric laser endomicroscopy frames	Eight predictive models (e.g., SVM, random forest, and naive bayes)	Ex vivo volumetric laser endomicroscopy	Maximum AUC 0.94, median 0.91
Ebigbo et al. (2019) ⁶	Endoscopic classification and segmentation of early neoplasia in patients with Barrett's esophagus	Retrospective	Training: 129 images, validation: 36 of early esophageal adenocarcinoma and 26 of normal Barrett's esophagus	CNN (ResNet with DeepLab V.3+ encoder-decoder neural network)	Capture system of images from the real-time endoscopic camera livestream	Accuracy in the validation dataset: 89.9%
Sehgal et al. (2018) ⁷	Decision tree algorithm generated by expert endoscopists could be used to improve dysplasia detection in non-expert endoscopists	Retrospective	Videos from 40 patients	Decision tree algorithm	Video recordings of high-definition endoscopy with i-Scan enhancement	Accuracy: 92% (vs. 88% of experts' average accuracy for dysplasia prediction)
Hong et al. (2017) ¹¹	To classify three conditions of intestinal metaplasia, gastric metaplasia and neoplasms based on endomicroscopy images	Retrospective	Training: 262 endomicroscopy images, testing: 26 test images	CNN	Endomicroscopy images	Accuracy: 80.77%
Swager et al. (2017) ⁹	Identification of early Barrett's esophagus neoplasia on ex vivo volumetric laser endomicroscopy images	Retrospective	60 volumetric laser endomicroscopy images (training and validation with leave-one-out cross-validation)	Ensemble method (SVM, discriminant analysis, AdaBoost, random forest, etc.)	Ex vivo volumetric laser endomicroscopy	AUC 0.95 (vs. 0.81 for the volumetric laser endomicroscopy experts)
van der Sommen et al. (2016) ⁸	Discrimination of early neoplastic lesions in Barrett's esophagus	Retrospective	100 endoscopic images from 44 patients (leave-one-out cross-validation on a per-patient basis)	SVM	White-light endoscopy	Sensitivity: 83%, specificity: 83% (per-image analysis)

CNN, convolutional neural network; SVM, support vector machine; AUC, area under the curve.

노드들에 backpropagation (역전파) 방식으로 전달되어 각 노드의 손실 함수의 가중치를 업데이트하고 다음 학습을 진행한다. 이외에 최적의 가중치를 빠르게 찾기 위하여 optimizer (i.e., stochastic gradient descent, adaptive gradient algorithm, root mean square propagation, adaptive moment estimation 등)를 사용한다. 인공지능 모델은 학습을 통하여 손실 함수의 최적의 가중치를 조정하고 새로운 데이터가 주어졌을 때 추론이 가능하도록 개발된다(Fig. 4).

2. 인공지능의 상부위장관 질환 적용 사례

대표적인 상부위장관 질환의 인공지능 적용 예로 바렛식도 (Barrett's esophagus)에서 신생물의 발생을 내시경 사진에서 발견하거나 구분하고자 하는 연구들이 발표되었다.⁴⁻⁹ 기존의 한 가지 인공지능 알고리즘을 사용하여 분석모델을 개발하는데 그치지 않고 여러 모델을 혼합하여 분석하는 알고리즘이 개발되었고,⁹ 실제 임상에 사용하기 위하여 내시경 검사 도중에 시술자가 이미지를 캡처하지 않아도 모니터에서 이미지를 캡처하는 시스템을 도입하여 실시간 분석에 적용하려는 시도

가 이루어지고 있다.⁶ Optical coherence tomography 기술을 이용한 volumetric laser endomicroscopy로 바렛식도에서 조기 신생물의 발생을 발견하고자 하는 연구들 또한 발표되었다.⁹⁻¹¹ 전반적으로 90% 이상의 정확도(accuracy)나 area under the curve 값을 보이지만 전향적인 검증 연구가 부족한 단점이 있다(Table 1).

식도암의 경우 내시경의 narrow-band imaging 데이터를 이용하여 암 전구 병변(pre-cancerous lesion)을 진단하거나 백색광 내시경(white-light endoscopy) 이미지를 이용하여 편평세포암(squamous cell carcinoma)을 진단,¹² 발견,¹³ 침윤 깊이를 구별하거나^{14,15} 또는 선암(adenocarcinoma)을 발견하는 모델이 발표되었다.¹⁶ 백색광 내시경이나 narrow-band imaging 이미지에 관계없이 또는 편평세포암이나 선암에 관계없이 식도암을 구별하는 데 높은 정확도를 보이는 모델 또한 발표되었다.¹⁷ 모델 개발에 모두 CNN을 사용하였고 전반적으로 높은 정확도를 보이고 있지만 전향적인 검증은 역시 부족하다(Table 2).

Helicobacter pylori 감염을 인공지능 알고리즘을 이용하

Table 2. Summary of Clinical Studies Using Machine Learning in the Diagnosis of Esophageal Cancers

Study	Aim of study	Design of study	Number of subjects	Type of artificial intelligence	Modality	Outcomes
Guo et al. (2020) ¹²	Diagnosis of precancerous lesions and early esophageal squamous cell carcinomas	Retrospective	Training: 6,473 narrow-band imaging images, validation with four datasets including images and video clips	CNN (SegNet)	Narrow-band imaging	AUC: 0.989
Tokai et al. (2020) ¹⁵	Detection and classification for invasion depth of esophageal squamous cell carcinoma	Retrospective	Training: 1,751 images, testing: 291 images	CNN (GoogLeNet)	White light endoscopy	Accuracy: 80.9%
Cai et al. (2019) ¹³	Detection of esophageal squamous cell carcinoma	Retrospective	Training and testing: 2,428 images, validation: 187 images	CNN	White light endoscopy	Average accuracy of endoscopists were increased with CNN from 81.7% to 91.1%
Nakagawa et al. (2019) ¹⁴	Classification for invasion depth of esophageal squamous cell carcinoma	Retrospective	Training: 8,660 non-ME and 5,678 ME images, validation: 405 non-ME images and 509 ME images	CNN	Non-ME and ME images of white-light endoscopy	Accuracy: 91% (mucosa, SM1 vs. SM2, SM3)
Ghatwary et al. (2019) ¹⁶	Detection of esophageal adenocarcinoma	Retrospective	100 images from 39 patients	CNNs	High-definition white light endoscopy	F-measure: 0.94
Horie et al. (2019) ¹⁷	Classification and detection of esophageal cancers including squamous cell carcinoma and adenocarcinoma	Retrospective	Training: 8,428, testing: 1,118 images	CNN	White-light endoscopy images and narrow-band imaging images	Accuracy: 98%

CNN, convolutional neural network; AUC, area under the curve; ME, magnifying endoscopy; SM, submucosa.

여 내시경 사진만으로 예측하려는 시도가 이루어졌다.¹⁸⁻²³ 대부분 CNN 모델을 사용하였고 백색광 내시경 이미지만을 사용한 연구가 대부분이며, 분류의 정확도는 80%에서 90% 이상까지 다양하게 보고되었다(Table 3).

위신생물 및 위암의 경우 위신생물 자체를 비신생물(또는 위암을 분류)과 분류하여 진단하는 모델²⁴과 조기위암의 침윤 깊이를 진단하여 내시경 치료의 대상이 될 가능성이 있는지 확인하는 모델이 개발되었다.²⁵⁻²⁷ 또한 위암의 발견^{28,29}과 범위를 지정하는 연구²⁹ 및 궤양형 병변 중에서 위암을 구별하는 연구³⁰에 인공지능 알고리즘이 활용되었으며 CNN이나 support vector machine 모델이 적용되었다(Table 4).

이미지에서 병변을 발견하거나 진단하는 것과 더불어 검사 자체의 질을 높이려는 시도가 이루어지고 있다. Wu 등³¹은 상부위장관 내시경 검사 도중 실시간으로 맹점(blind spot)을

모니터링해주는 인공지능 알고리즘을 개발하여 무작위 대조군 연구를 통하여 알고리즘을 검사에 사용할 경우 맹점을 줄일 수 있음을 발표하였다. 또한 같은 연구 그룹에서 의식화 진정 내시경에 인공지능 알고리즘을 사용하였을 경우 비진정 상태의 검사보다 맹점을 더욱 줄일 수 있음을 발표하였다.³²

이미지 분석뿐만 아니라 예후 예측 모델의 정확도를 향상시키고자 하는 시도가 이루어지고 있다. 상부위장관 출혈의 경우 재출혈이나 사망률 등의 예후 예측을 위한 점수 체계(scoring system)를 임상 진료에 적용하여 왔는데, 인공지능 알고리즘이 이들 모델보다 높은 예측력을 보여주고 있다(Table 5).³³⁻³⁶ 이외에도 호산구성 식도염(eosinophilic esophagitis) 환자에서 조직 검사를 통하여 얻은 검체의 inflammatory transcript profile을 이용하여 머신러닝 알고리즘을 통하여 보다 정밀한 진단을 가능하게 하고, 치료에 따른

Table 3. Summary of Clinical Studies Using Machine Learning in the Diagnosis of *Helicobacter pylori* (*H. pylori*) Infection in Endoscopic Images

Study	Aim of study	Design of study	Number of subjects	Type of artificial intelligence	Modality	Outcomes
Zheng et al. (2019) ¹⁸	Diagnosis of <i>H. pylori</i> infection	Retrospective pilot	Training: 11,729 images, testing: 3,755 images	CNN	White-light endoscopy	AUC: 0.93. Accuracy: 84.5% in a single image diagnosis
Shichijo et al. (2019) ¹⁹	Diagnosis of <i>H. pylori</i> infection	Retrospective	Training set: 98,564 images, testing: 23,699 images	CNN	White-light endoscopy	Accuracy: 80% (465/582) of negative diagnoses, 84% (147/174) eradicated, and 48% (44/91) positive were accurate. The time needed to diagnose 23,699 images was 261 seconds
Nakashima et al. (2018) ²⁰	Diagnosis of <i>H. pylori</i> infection	Prospective pilot	222 patients (training: 162, testing: 60)	CNN	White-light endoscopy and image-enhanced endoscopy, such as blue laser imaging and linked color imaging	AUC: 0.96 (blue laser imaging), 0.95 (linked color imaging)
Itoh et al. (2018) ²¹	Diagnosis of <i>H. pylori</i> infection	Prospective	Training: 149 images (596 images through data augmentation), testing: 30 images	CNN	White-light endoscopy	AUC: 0.956
Shichijo et al. (2017) ²²	Diagnosis of <i>H. pylori</i> Infection	Retrospective	Training: 32,208 images, testing: 11,481 images	CNN	White-light endoscopy	Accuracy: 83.1%
Huang et al. (2004) ²³	Diagnosis of <i>H. pylori</i> infection	Prospective	Training: 30 patients, testing: 74 patients	Refined feature selection with neural network	White-light endoscopy	Accuracy over 80 % in predicting the presence of gastric atrophy, intestinal metaplasia and the severity of <i>H. pylori</i> -related gastric inflammation

CNN, convolutional neural network; AUC, area under the curve.

반응이 다를 것으로 예측되는 하위 그룹을 분류하는 연구가 발표되었다.³⁷

3. 인공지능 모델의 정확도 및 필요성

현재의 인공지능 알고리즘의 정확도는 임상 연구에서 일반적으로 사용되는 통계학적인 모델보다는 높지만 그 이유에 대해서는 구체적으로 설명하기가 어렵다. 설정된 입력값과 출력값에 대하여 가중치를 가지는 여러 함수층의 분석을 통하여 최적의 가중치를 찾아가는 것이 인공신경망 분석의 목표로, 수많은 함수의 연결과 관계를 단순히 이해하기는 어렵고 또한 입력값 중에 어떤 특징적인 인자가 출력값에 영향을 미쳤는지도 파악하기가 어렵다. 즉, 설명이 어려운 알고리즘(black-box nature)으로 통상적으로 사용하는 방식으로 진단/구별 능력을 향상시키기 어려운 분야에 적용이 가능하다. 인공지능 알

고리즘의 작동 방식에서 입력값과 결과값과의 관계를 논리적으로 따져보는 작업은 일부 연구에서 다루고 있으나 인공신경망의 작동 방식이나 판단 근거를 단순한 수식으로 함축시키기는 어렵기 때문에 이를 위해서는(독립 변수와 종속 변수 간의 관계, 영향) 기존의 다른 통계학적 방법이 유리하다. 기전을 설명하기 어려운 인공지능 알고리즘의 판단이 적절한지 확인하는 방법은 모델 검증이며, 최근 대규모의 검증 논문들이 등장하고 있다.³⁸

전향적 검증이 없는 연구들의 정확도는 과대적합 문제를 고려하여 해석해야 한다. 과대적합의 경우 모델이 학습용 데이터에 과도하게 최적화되어 noise (학습 과정에서 발생하는 무작위 오류)까지 수용하는 것이다. 따라서, 학습 시에는 예측력이 높지만 모델 개발에 사용되지 않은 검증용 데이터(external validation dataset)를 사용할 경우에는 예측력이

Table 4. Summary of Clinical Studies Using Machine Learning in the Gastric Neoplasms

Study	Aim of study	Design of study	Number of subjects	Type of artificial intelligence	Modality	Outcomes
Cho et al. (2019) ²⁴	Diagnosis of gastric neoplasms	Retrospective model establishment and prospective validation	Training and testing: 5,017 images, validation: 200 images	CNN	White-light endoscopy	AUCs of classifying gastric cancer: 0.877, gastric neoplasm: 0.927
Yoon et al. (2019) ²⁵	Classification of endoscopic images as early gastric cancer (T1a or T1b) or non-cancer	Retrospective	11,539 endoscopic images (896 T1a-, 809 T1b-, and 9834 non-early gastric cancer)	CNN	White-light endoscopy	AUC of early gastric cancer detection: 0.981, depth prediction: 0.851
Zhu et al. (2019) ²⁶	Diagnosis of depth of invasion in gastric cancer (mucosa/SM1/deeper than SM1)	Retrospective	Training: 790 images, testing: 203 images	CNN	White-light endoscopy	Accuracy: 89.2%, AUC: 0.94
Hirasawa et al. (2018) ²⁸	Detection of gastric cancers	Retrospective	Training: 13,584 images, testing: 2,296 images	CNN	White-light endoscopy, chromoendoscopy, narrow-band imaging	Accurate detection rate with a diameter of 6 mm or more: 98.6%
Kanesaka et al. (2018) ²⁹	Diagnosis and delineation of early gastric cancer using magnifying narrow-band imaging images	Retrospective	Training: 126 images, testing: 81 images	SVM	Magnifying narrow-band imaging	Accuracy: 96.3%
Kubota et al. (2012) ²⁷	Diagnosis of depth of invasion in gastric cancer	Retrospective	902 images	ANN	White-light endoscopy	Accuracy: 77.2%, 49.1%, 51.0%, and 55.3% for T1-4 staging, respectively
Lee et al. (2020) ³⁰	Classification of normal, benign ulcer, and gastric cancer	Retrospective	200 normal, 367 cancer, and 220 ulcer cases	CNN	White-light endoscopy	Accuracy: normal vs. ulcer/normal vs. cancer: above 90%; ulcer vs. cancer: 77.1%

CNN, convolutional neural network; AUC, area under the curve; SM, submucosa; SVM, support vector machine; ANN, artificial neural network.

떨어지는 현상이다. 이를 해결하기 위하여 모델 개발 단계에서 drop out, rectified linear unit 함수의 적용이나 weight decay (regularization), batch normalization, data augmentation 등의 기술적인 방식이 적용되고 있다.

이미 개발이 완료된 인공지능 알고리즘이 다수 존재함에도 불구하고 우리는 이러한 모델들이 정말 필요한가를 고민해야 한다. 질병 진단의 정확도가 인공지능 알고리즘의 적용으로 상승한다는 것은 여러 임상 분야의 연구에서 공통으로 보이는 현상이다. 하지만 정확도만이 의료에서 항상 중요한 지표는 아니고 검사의 역할에 따라 민감도나 특이도, 양성 예측도 등의 지표가 임상적으로 더욱 중요한 분야가 있다. 따라서 모델을 수립하기 전에 이들이 어디에 사용될 것인지를 고려하는 것이 중요하다. 이러한 인공지능 모델의 적용이 적합한 분야는 매개 변수로 결과값을 예측하는 방식에 논리적인 설명이 필요한 분야는 아닐 가능성이 높다. 인공지능 알고리즘이 학습용 데이터의 어떤 특징을 학습하여 특정 방식의 분류를 하였는지는 이해하기 어렵더라도 정확도, 민감도, 특이도 또는 양성 예측도 등의 진단 지표를 높여야 할 필요성이 있는 분야나 또는 이미 충분한 과학적인 논의가 이루어져 작동 방식에 대한 논리적인 설명이 더 이상 필요하지 않지만 임상 의사의 피로도를 고려해야 하는 분야가 적합할 것으로 생각한다.

또한 인공지능 알고리즘을 적용해볼 수 있는 분야의 진단 지표가 낮은가를 고려해야 하고, 만약 이런 인공지능 알고리즘이 실제로 사용된다면 이들 모델의 질 평가 및 관리(quality assurance) 또한 준비가 되어 있어야 한다.³⁹ 최근에 논문으로 출판된 대부분의 알고리즘이 DNN 중에서 CNN을 사용하

여 주제와 관계없이 90% 이상의 정확도를 보이고 있지만 전향적 검증이 없는 연구가 대부분이기 때문에 실제 바로 임상에 적용하기에는 무리가 있다. 인공지능 알고리즘의 잘못된 진단이나 분류가 미치는 윤리적인 문제 등을 고려하면 주제에 따라서는 더 높은 진단 지표가 보장되어야 하는 분야가 있고 임상 현장에 사용되더라도 진료에 보조적인 수단이 될 가능성이 높다.¹

임상 연구를 통하여 기존의 진료에 부가적인 목적으로 인공지능 알고리즘을 사용하여 실제 진단이나 분류의 정확도가 상승하거나¹³ 무작위 대조군 연구를 통하여 직접 환자의 진단이나 예후에 영향을 미친다는 보고도 필요하다.³² 의사의 눈으로 인지하지 못하는 인자를 인공지능 알고리즘이 파악할 수 있다면 임상적인 유용성이 높을 것으로 판단되지만 의료기기의 검증은 다른 분야에 비하여 더욱 신중해야 함을 고려할 때 여전히 해석 가능성(interpretability) 문제는 해결해야 할 숙제이다. 책임 소재(liability)의 문제로 인공지능 알고리즘을 진료에 사용하지 않고 기존의 표준 진료지침만을 따르는 경우가 발생할 수 있고, 반대로 일부 분야에서는 인공지능 알고리즘의 적용 자체가 미래의 표준 진료지침이 될 가능성 또한 배제할 수 없다.⁴⁰

4. 인공지능 모델의 한계점

머신러닝 알고리즘의 정확성과 설명력은 trade-off 관계를 가지고 있다.¹ 앞서 예를 들었던 회귀분석은 신경망 분석에 비하여 설명력은 높지만 정확도는 낮다. 인공신경망 모델은 회귀분석에 비하여 정확도는 높지만 설명력은 낮은 모델이다.

Table 5. Summary of Clinical Studies Using Machine Learning in the Upper Gastrointestinal Hemorrhage

Study	Aim of study	Design of study	Number of subjects	Type of artificial intelligence	Outcomes
Shung et al. (2020) ³³	Develop a model to calculate the risk of hospital-based intervention or death in patients with upper gastrointestinal hemorrhage	Prospective	Training and internal validation set: 1,958 patients, external validation: 399 patients	Gradient Boosting Algorithm	AUC: 0.91 (internal validation) AUC: 0.90 (external validation)
Rotondano et al. (2011) ³⁴	Develop a model to predict any death occurring within 30 days of the index bleeding episode	Prospective	Training and testing: 2,380 patients	ANN	Accuracy: 96.8%, AUC: 0.95
Das et al. (2008) ³⁵	Develop a model to predict stigmata of recent hemorrhage and need for endoscopic therapy	Prospective	Training: 194 patients, testing: 193 patients, external validation: 200 patients	ANN	Accuracy: 77% (predict stigmata of recent hemorrhage), 61% (need for endoscopic therapy) in external validation
Grossi et al. (2008) ³⁶	Develop a model to predict the risk of death in patients with nonvariceal upper gastrointestinal bleeding	Prospective	Training and testing: 807 patients	ANN	Accuracy: 89%

AUC, area under the curve; ANN, artificial neural network.

CNN 모델은 주로 가중치로만 이루어진 설명력이 부족한 모델임을 고려하면 최종 판단 근거가 잘못되었을 경우에 알고리즘에 오류가 있기 때문인지 확인이 어렵기 때문에 여전히 불완전한 모델이다. 이러한 인공지능의 신뢰와 안전성 문제를 해결하기 위하여 설명이 가능한 인공지능(explainable artificial intelligence)을 개발하려는 시도가 이루어지고 있다. 간단한 예로 CNN 학습 이후에 이미지의 어떤 부분을 근거로 출력값을 도출하였는지 확인하기 위하여 class activation map과 같은 방식이 적용이 되고 있다. 하지만 앞서 언급한 대로 정확도를 높이는 것이 필요한 분야가 있고 설명력을 높이는 것이 필요한 분야가 따로 존재하기 때문에 한계점을 이해하고 적용 분야를 선택하는 것이 필요하다. 역설적이지만 한 연구에서 설명이 가능한 간단한 모델(transparent model, white-box or glass-box model)이 오히려 오류를 발견하여 수정하는 것이 어렵다는 결론을 내렸으며,⁴¹ CNN이 noise에 비교적 강한 모델(noise-robust)임은 잘 알려져 있는 사실이다. 인간이 사고하는 방식을 아직 자세히 이해하기는 어렵듯이 DNN 알고리즘이 높은 성능을 보이는 이유도 복잡한 인간의 신경망을 모사한 단순히 설명하기 어려운 알고리즘이기 때문인지 않을까 추측된다.³⁹

진전을 설명하기 어려운 인공지능 알고리즘의 판단을 확인하는 방법은 모델 검증이다. 이는 의료 인공지능 알고리즘(특히 CNN)이 데이터에 의존적인 모델이기 때문이다. 따라서 인공지능을 학습시키기 위한 데이터의 양과 질이 중요하다. 사람은 직관이라는 특징을 가지고 있어 데이터의 양이 적어도 학습하는데 문제가 없을 수 있지만 인공지능 알고리즘은 데이터로부터 특징을 발견하여 학습하기 때문에 이런 특징을 가지고 있는 데이터의 양이 충분해야 한다. 최근에는 84,424명으로부터 획득된 1,036,496개의 내시경 사진을 이용하여 상부위장관 암을 진단하는 모델이 개발되었으며 내부, 외부 검증에서 모두 90%가 넘는 진단 정확도를 보여주었다.³⁸ 데이터의 용량만이 중요한 것은 아니고 현실을 반영하는 균형 있는 데이터를 확보하는 것 또한 중요하다. 인공지능 알고리즘이 구별해야 하는 비정상 병변이 데이터에 얼마나 포함되어 있는지에 따라, 즉 데이터의 불균형(class imbalance) 문제로 진단 능력이 떨어지는 경우가 다수 있기 때문에 균형 있는 데이터가 중요하다.¹ 소화기 영역에 적용되고 있는 CNN의 성격이 주로 이미지의 특징을 발견하여 학습하는 것임을 고려하면 실제로는 빅데이터라 하더라도 이러한 다양한 특징을 가진 이미지가 학습용 데이터에 충분히 포함되어야 한다.

아직까지 임상에 적용할 만한 인공지능 알고리즘은 지도 학습(supervised learning)의 형태를 가진 학습용 데이터에 의존적인 모델이다. 잘못 labeling된 학습용 데이터로 만들어진 인공지능 알고리즘이 오류로 이어지고 그것이 임상 진료에

잘못 적용되었을 때 확대될 수 있는 문제 등을 고려하면 전문가들이 신중하게 학습용 데이터를 준비하여 알고리즘을 개발하는 것이 중요하다. 의료 데이터는 정상적인 데이터는 많지만 질환을 가지고 있는 비정상 데이터는 부족한 것이 일반적이다. 이럴 경우 병변에 labeling (annotation)이 되면 CNN 모델의 분류 능력이 상승하는 경우를 확인할 수 있다.³⁹ 하지만 소규모 기관별 후향적 데이터는 질환을 가지고 있는 비정상 데이터가 부족한 경우가 많고 얼마나 현실을 반영하는 데이터인지 확인하기가 어렵기 때문에 학습이 어려운 경우가 있다. Anomaly detection with generative adversarial networks의 경우 이런 상황에서 유용할 것으로 추측된다. 이는 이상 부위를 탐지하는 알고리즘(anomaly detection)으로 정상적인 데이터로 모델의 학습을 시행하고 비정상 병변이 포함된 이미지에서 모델이 인공적으로 생성한 이미지와의 비교를 통하여 이상이 있는 부위를 찾는 방식이기 때문이다. 따라서 병변에 labeling이 없이 학습을 시행한다는 장점이 있지만 아직 의료 분야의 연구 결과가 많지 않다.⁴²

5. 인공지능의 발전 방향

현재까지 소화기 분야의 인공지능 알고리즘의 개발은 진단의 정확도를 높이는 방향으로 주로 진행이 되었지만 향후에는 진단 이외에도 치료에 접목될 가능성이 높다. 이미 micro-endoscopy, robot-assisted treatment, decision support system-based treatment modality, 디지털 치료제(digital therapeutics, 의약품이 아닌 application, game, augmented reality, virtual reality, mixed reality 등의 소프트웨어로 질병을 치료하는 것) 등의 기술을 활용한 치료법의 개발이 인공지능 알고리즘을 바탕으로 이루어지고 있다. 하지만 한편으로는 여전히 기술적인 문제와 사회적인 합의, 환자 안전, 윤리 문제, 비용-효과 문제 등 논의사항이 많은 것도 사실이다.

다수의 연구들이 인공지능 알고리즘이 의사의 진단 능력을 추월하였다는 보고를 하고 있기 때문에 교육 분야에서도 인공지능 알고리즘을 도입할 것이라는 추측이 있다. 하지만 앞서 기술한 인공지능 알고리즘의 현실적인 한계들을 고려할 때 역할의 침범(의사 vs. 인공지능)의 구도가 아닌 교육이나 진료에 도움을 주는 부가적인 도구로서의 상승 효과(의사+인공지능 vs. 의사)를 탐색하는 것이 적합한 흐름이다. 아직 의료 인공지능 알고리즘이 사용자의 제작 의도대로만 작동하는 불완전한 모델임을 고려하면 의료 교육이나 수련을 대체할 수는 없다.³⁹ 따라서 앞으로의 교육 방향 역시 의료술기 분야에서의 도제식 교육은 지속적으로 필요할 것으로 생각한다. 다만 관찰자 간의 차이(inter-observer variability)가 많고 표준화하기 어려운 육안 진단 분야에서는 일부 도움을 받을 수 있을

것으로 예상된다.

현재는 주어진 문제를 해결하는 위주의 단일 기능 또는 두 가지 정도의 기능을 복합한 인공지능 모델이 다수이고 제작자의 의도대로 활용이 가능한 약인공지능의 시대이다.³⁹ 향후 일부 사회과학자들이 예상하는 인간과 같은 사고가 가능한 강인공지능의 시대나 인류를 뛰어넘는 초인공지능의 시대가 도래하여 미래학자인 Kurzweil⁴³이 저술한대로 인공지능과 인간이 각자의 한계점을 뛰어넘는 특이점(singularity)이 올지는 알 수 없다. 인공지능의 개발 역사에서 초창기의 퍼셉트론은 선형 함수(분류)를 구현하는 인간의 단일 neuron을 인공화한 형태였는데, 이는 비트 연산(bitwise operation)의 비선형의 논리 연산인 배타적 논리합(exclusive OR, XOR) 문제를 해결하지 못하여 침체기를 맞이하였다(Fig. 5). 추후 다층 퍼셉트론(multi-layer perceptron)과 backpropagation 알고리즘의 개발로 이러한 비선형 문제를 해결하면서 다시 인공지능이 발전하였다. 하지만 최근 사람은 대뇌피질의 단일 neuron만으로도 배타적 논리합 추론을 해결한다는 연구 결과가 발표되어 인공지능 알고리즘보다 더욱 고차원의 연산이 가능한 것으로 추측된다.⁴⁴

인공지능의 의료 적용은 먼 미래가 아니다. 이미 국내에서도 대장 용종과 위암을 발견하여 segmentation box로 내시경 화면에 표시해주는 모델(JFD-01A [JLK inspection, Seoul, Korea]: 대장 내시경 이미지 자료를 기반으로 용기 또는 함몰 패턴을 자동으로 감지하여 표시, JFD-02A [JLK inspection]: 위 내시경 이미지 자료를 통하여 용기 또는 함몰 부위, 붉은색 또는 흰색으로 변색된 부위 등을 자동으로 감지하여 표시, 2등급 의료기기)이 2019년 10월에 식품의약품안전처 허가를 획득하였다. 많은 소프트웨어 관련 코드들이 오픈소스로 개방이

되어 있고 이를 활용하여 자율적인 연구 및 적용이 가능하다. 인공신경망의 layer의 숫자나 learning epoch (가중치를 업데이트하는 학습을 순방향, 역방향으로 1회 시행), batch size (한 번 학습에 처리하는 데이터의 수), iteration (몇 개의 batch를 학습 시에 사용하는지에 대한 수), optimizer 등의 변동이 가능한 hyperparameter를 변화시키거나 기존에 수립된 pre-trained model을 이용하여 classifier를 추가하여 fine-tuning (transfer learning)하는 등의 방식을 통하여 다양한 적용, 응용이 가능한 시대이다. 최근에는 모델을 최적화하기 위하여 난해한 알고리즘을 조정하는 것에서 벗어나 자동으로 hyperparameter를 최적화하는 자동화 머신러닝 도구(i.e., AutoGluon)가 발표되어 인공지능의 진입 장벽을 낮추고 대중화하려는 시도 또한 이루어지고 있다.

결론

인공지능 알고리즘, 특히 딥러닝 알고리즘은 입력값과 출력값을 가지고 이를 학습하여 새로운 입력값에도 출력값을 잘 예측하도록 훈련하는 black-box와 같은 도구이다. 이는 입력되는 데이터의 양과 질에 따라 그 성능이 좌우된다. 임상 연구도 마찬가지로 얼마나 질 높은 데이터를 대량을 구축하는지가 증거 수준에 영향을 미치는 것을 부정할 수 없다. 임상의사로서 질 높은 임상 데이터를 구축하는 것이 중요하고 이를 기반으로 한 모델 개발과 검증 역시 중요하다. 현재는 누구나 인공지능을 사용하는 시대이다. 의료 현장에 직접 사용되기 위하여 해석 가능성과 같은 제한사항들이 아직 존재하지만 흥미롭고 신기하기까지 한 인공지능의 작동 방식은 의료 분야의 적용에 매력적이다. 인공지능 알고리즘에 대한 합리적인 논의를

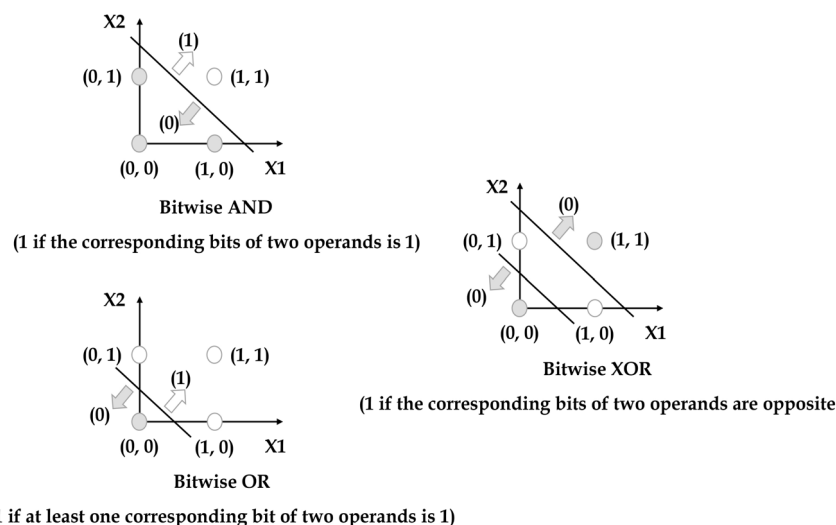


Fig. 5. Limitation of a single perceptron.

통하여 활발한 개발이 이루어지기를 기대한다.

REFERENCES

1. Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. *World J Gastroenterol* 2019;25:1666-1683.
2. Ebigbo A, Palm C, Probst A, et al. A technical review of artificial intelligence as applied to gastrointestinal endoscopy: clarifying the terminology. *Endosc Int Open* 2019;7:E1616-E1623.
3. Imler TD, Morea J, Kahi C, et al. Multi-center colonoscopy quality measurement utilizing natural language processing. *Am J Gastroenterol* 2015;110:543-552.
4. de Groof AJ, Struyvenberg MR, van der Putten J, et al. Deep-learning system detects neoplasia in patients with Barrett's esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. *Gastroenterology* 2019 Nov 21. [Epub ahead of print]
5. de Groof J, van der Sommen F, van der Putten J, et al. The Argos project: the development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterol J* 2019;7:538-547.
6. Ebigbo A, Mendel R, Probst A, et al. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut* 2019 Sep 20. [Epub ahead of print]
7. Sehgal V, Rosenfeld A, Graham DG, et al. Machine learning creates a simple endoscopic classification system that improves dysplasia detection in Barrett's esophagus amongst non-expert endoscopists. *Gastroenterol Res Pract* 2018;2018:1872437.
8. van der Sommen F, Zinger S, Curvers WL, et al. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy* 2016;48:617-624.
9. Swager AF, van der Sommen F, Klomp SR, et al. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017;86:839-846.
10. Struyvenberg MR, van der Sommen F, Swager AF, et al. Improved Barrett's neoplasia detection using computer-assisted multi-frame analysis of volumetric laser endomicroscopy. *Dis Esophagus* 2019 Jul 31. [Epub ahead of print]
11. Hong J, Park BY, Park H. Convolutional neural network classifier for distinguishing Barrett's esophagus and neoplasia endomicroscopy images. *Conf Proc IEEE Eng Med Biol Soc* 2017;2017:2892-2895.
12. Guo L, Xiao X, Wu C, et al. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastrointest Endosc* 2020;91:41-51.
13. Cai SL, Li B, Tan WM, et al. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). *Gastrointest Endosc* 2019;90:745-753.e2.
14. Nakagawa K, Ishihara R, Aoyama K, et al. Classification for invasion depth of esophageal squamous cell carcinoma using a deep neural network compared with experienced endoscopists. *Gastrointest Endosc* 2019;90:407-414.
15. Tokai Y, Yoshio T, Aoyama K, et al. Application of artificial intelligence using convolutional neural networks in determining the invasion depth of esophageal squamous cell carcinoma. *Esophagus* 2020 Jan 24. [Epub ahead of print]
16. Ghatwary N, Zolgharni M, Ye X. Early esophageal adenocarcinoma detection using deep learning methods. *Int J Comput Assist Radiol Surg* 2019;14:611-621.
17. 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.
18. Zheng W, Zhang X, Kim JJ, et al. High accuracy of convolutional neural network for evaluation of *Helicobacter pylori* infection based on endoscopic images: preliminary experience. *Clin Transl Gastroenterol* 2019;10:e00109.
19. Shichijo S, Endo Y, Aoyama K, et al. Application of convolutional neural networks for evaluating *Helicobacter pylori* infection status on the basis of endoscopic images. *Scand J Gastroenterol* 2019;54:158-163.
20. Nakashima H, Kawahira H, Kawachi H, Sakaki N. Artificial intelligence diagnosis of *Helicobacter pylori* infection using blue laser imaging-bright and linked color imaging: a single-center prospective study. *Ann Gastroenterol* 2018;31:462-468.
21. Itoh T, Kawahira H, Nakashima H, Yata N. Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images. *Endosc Int Open* 2018;6:E139-E144.
22. Shichijo S, Nomura S, Aoyama K, et al. Application of convolutional neural networks in the diagnosis of *Helicobacter pylori* infection based on endoscopic images. *EBioMedicine* 2017;25:106-111.
23. Huang CR, Sheu BS, Chung PC, Yang HB. Computerized diagnosis of *Helicobacter pylori* infection and associated gastric inflammation from endoscopic images by refined feature selection using a neural network. *Endoscopy* 2004;36:601-608.
24. Cho BJ, Bang CS, Park SW, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. *Endoscopy* 2019;51:1121-1129.
25. 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:E1310.
26. Zhu Y, Wang QC, Xu MD, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. *Gastrointest Endosc* 2019;89:806-815.e1.
27. Kubota K, Kuroda J, Yoshida M, Ohta K, Kitajima M. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. *Surg Endosc* 2012;26:1485-1489.
28. 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.
29. Kanesaka T, Lee TC, Uedo N, et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. *Gastrointest Endosc* 2018;87:1339-1344.
30. Lee JH, Kim YJ, Kim YW, et al. Spotting malignancies from gastric endoscopic images using deep learning. *Surg Endosc* 2019;33:3790-3797.
31. 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.
32. Chen D, Wu L, Li Y, et al. Comparing blind spots of unsedated ultrafine, sedated, and unsedated conventional gastroscopy with and without artificial intelligence: a prospective, single-blind, 3-parallel-group, randomized, single-center trial. *Gastrointest Endosc* 2020;91:332-339.e3.
 33. Shung DL, Au B, Taylor RA, et al. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. *Gastroenterology* 2020;158:160-167.
 34. Rotondano G, Cipolletta L, Grossi E, et al. Artificial neural networks accurately predict mortality in patients with nonvariceal upper GI bleeding. *Gastrointest Endosc* 2011;73:218-226.e2.
 35. Das A, Ben-Menachem T, Farooq FT, et al. Artificial neural network as a predictive instrument in patients with acute non-variceal upper gastrointestinal hemorrhage. *Gastroenterology* 2008;134:65-74.
 36. Grossi E, Marmo R, Intraligi M, Buscema M. Artificial neural networks for early prediction of mortality in patients with non variceal upper GI bleeding (UGIB). *Biomed Inform Insights* 2008; 1:7-19.
 37. Sallis BF, Erkert L, Moñino-Romero S, et al. An algorithm for the classification of mRNA patterns in eosinophilic esophagitis: Integration of machine learning. *J Allergy Clin Immunol* 2018; 141:1354-1364.e9.
 38. Luo H, Xu G, Li C, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multi-centre, case-control, diagnostic study. *Lancet Oncol* 2019;20:1645-1654.
 39. Cho BJ, Bang CS. Artificial intelligence for the determination of a management strategy for diminutive colorectal polyps: hype, hope, or help. *Am J Gastroenterol* 2020;115:70-72.
 40. Price WN 2nd, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA* 2019;322:1765-1766.
 41. Poursabzi-Sangdeh F, Goldstein DG, Hofman JM, Vaughan JW, Wallach H. Manipulating and measuring model interpretability. *arXiv preprint* 2018;arXiv:1802.07810.
 42. Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal* 2019;54:30-44.
 43. Kurzweil R. The singularity is near : when humans transcend biology. Seoul: Gimmyoungsa, 2007.
 44. Gidon A, Zolnik TA, Fidzinski P, et al. Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science* 2020;367:83-87.