ISSN 2644-3260 (online)

Artificial Intelligence in *Medical Imaging*

Artif Intell Med Imaging 2022 June 28; 3(3): 55-86

Published by Baishideng Publishing Group Inc

AIMI

Artificial Intelligence in Medical Imaging

Contents

Bimonthly Volume 3 Number 3 June 28, 2022

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Medical Imaging*, Bilge Gursel, MD, Professor, Ondokuz Mayis University School of Medicine Radiation Oncology Department, Atakum, Samsun 55200, Turkey. bgursel@omu.edu.tr

AIMS AND SCOPE

The primary aim of *Artificial Intelligence in Medical Imaging* (*AIMI*, *Artif Intell Med Imaging*) is to provide scholars and readers from various fields of artificial intelligence in medical imaging with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

AIMI mainly publishes articles reporting research results obtained in the field of artificial intelligence in medical imaging and covering a wide range of topics, including artificial intelligence in radiology, pathology image analysis, endoscopy, molecular imaging, and ultrasonography.

INDEXING/ABSTRACTING

The *AIMI* is now abstracted and indexed in Reference Citation Analysis, China Science and Technology Journal Database.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: *Ying-Yi Yuan*; **Production Department Director:** *Xu Guo*; **Editorial Office Director:** *Yun-Xiaojiao Wu*.

NAME OF JOURNAL
Artificial Intelligence in Medical Imaging

ISSN
ISSN 2644-3260 (online)

LAUNCH DATE
June 28, 2020

FREQUENCY
Bimonthly

EDITORS-IN-CHIEF
Xue-Li Chen, Caroline Chung, Jun Shen

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2644-3260/editorialboard.htm

PUBLICATION DATE
June 28, 2022

INSTRUCTIONS TO AUTHORS
https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS
https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH
https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS
https://www.wjgnet.com/bpg/GerInfo/288

PUBLICATION MISCONDUCT
https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE
https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS
https://www.wjgnet.com/bpg/GerInfo/239

ONLINE SUBMISSION
https://www.f6publishing.com

AIMI Artificial Intelligence in Medical Imaging

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Med Imaging 2022 June 28; 3(3): 55-69

DOI: 10.35711/aimi.v3.i3.55

ISSN 2644-3260 (online)

MINIREVIEWS

Enhancing medical-imaging artificial intelligence through holistic use of time-tested key imaging and clinical parameters: Future insights

Prakash Nadkarni, Suleman Adam Merchant

Specialty type: Radiology, nuclear medicine and medical imaging

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): D
Grade E (Poor): 0

P-Reviewer: Jamshidi M, Czech Republic; Sato H, Japan; Singh M, United States

A-Editor: Yao QG, China

Received: January 19, 2022
Peer-review started: January 19, 2022
First decision: March 12, 2022
Revised: April 12, 2022
Accepted: June 16, 2022
Article in press: June 16, 2022
Published online: June 28, 2022

Prakash Nadkarni, College of Nursing, University of Iowa, Iowa City, IA 52242, United States

Suleman Adam Merchant, Department of Radiology, LTM Medical College & LTM General Hospital, Mumbai 400022, Maharashtra, India

Corresponding author: Suleman Adam Merchant, MD, Former Dean, Professor & Head, Department of Radiology, LTM Medical College & LTM General Hospital, Sion Hospital, Mumbai 400022, Maharashtra, India. suleman.a.merchant@gmail.com

Abstract

Much of the published literature in Radiology-related Artificial Intelligence (AI) focuses on single tasks, such as identifying the presence or absence or severity of specific lesions. Progress comparable to that achieved for general-purpose computer vision has been hampered by the unavailability of large and diverse radiology datasets containing different types of lesions with possibly multiple kinds of abnormalities in the same image. Also, since a diagnosis is rarely achieved through an image alone, radiology AI must be able to employ diverse strategies that consider all available evidence, not just imaging information. Using key imaging and clinical signs will help improve their accuracy and utility tremendously. Employing strategies that consider all available evidence will be a formidable task; we believe that the combination of human and computer intelligence will be superior to either one alone. Further, unless an AI application is explainable, radiologists will not trust it to be either reliable or bias-free; we discuss some approaches aimed at providing better explanations, as well as regulatory concerns regarding explainability ("transparency"). Finally, we look at federated learning, which allows pooling data from multiple locales while maintaining data privacy to create more generalizable and reliable models, and quantum computing, still prototypical but potentially revolutionary in its computing impact.

Key Words: Medical imaging; Artificial intelligence; Federated learning; holistic approach; Quantum computing; Future insights

Core Tip: It is necessary to understand the principles of how different artificial intelligence (AI) approaches work to appreciate their respective strengths and limitations. While advances in deep neural net research in Radiology are impressive, their focus must shift from applications that perform only single recognition task, to those that perform realistic multi-recognition tasks that radiologists perform daily. Humans use multiple problem-solving strategies, applying each as needed. Similarly, realistic AI solutions must combine multiple approaches. Good radiologists are also good clinicians. AI must similarly be able to use all available evidence, not imaging information alone, and not just one/Limited aspects of imaging. Both humans and computer algorithms (including AI) can be biased. A way to reduce bias, as well as prevent failure, is better explainability – the ability to clearly describe the workings of a particular application to a subject-matter expert unfamiliar with AI technology. Federated learning allows more generalizable and accurate machine-learning models to be created by preserving data privacy, concerns about which form a major barrier to large-scale collaboration. While the physical hurdles to implementing Quantum computing at a commercial level are formidable, this technology has the potential to revolutionize all of computing.

Citation: Nadkarni P, Merchant SA. Enhancing medical-imaging artificial intelligence through holistic use of time-tested key imaging and clinical parameters: Future insights. *Artif Intell Med Imaging* 2022; 3(3): 55-69
URL: https://www.wjgnet.com/2644-3260/full/v3/i3/55.htm
DOI: https://dx.doi.org/10.35711/aimi.v3.i3.55

INTRODUCTION

As medical knowledge's volume and complexity advances, electronic clinical decision support will become increasingly important in healthcare delivery, and increasingly likely to use Artificial Intelligence (AI). Historically, AI approaches have been diverse. However, even senior radiologists, *e.g.* [1], have inaccurately considered AI, machine learning, and deep learning as synonymous. We therefore summarize these approaches, considering their strengths and weaknesses.

Symbolic approaches

These, the focus of "classical" AI (1950s-1990s), embody the use of high-level abstractions ("symbols") that represent the concepts that humans (often experts) use in solving non-numerical problems. They are most closely related to traditional computer science/software development. In fact, they are mainstream enough that specific terms (instead of "AI") are preferred to describe a given approach. Among the successes:

Business-rule systems (BRS or "Expert Systems")[2]: These allow human experts, working either with software developers or with graphical user interfaces, to embody their knowledge of a particular area to offer domain-specific advice/diagnosis. Robust open-source tools such as Drools[3] are available for building BRS.

Constraint programming systems[4]: Constraint satisfaction involves finding a solution to a multivariate problem given a set of constraints on those variables. When the constraints are numeric, techniques such as linear programming[5] (which preceded symbolic AI and is applied in numerous business-operations problems) work better. Some software, such as Frontline Solver(TM)[6] (of which Microsoft Excel's "Solver" add-in is a lightweight version) handles both numerical and symbolic constraints.

Data-driven approaches

(Also called "machine learning" or ML): These are used to make predictions, or decisions based on those predictions, by manipulating numbers, or entities transformed into numbers, rather than symbols. They are most useful in domains where human experts have not formulated problem-solving strategies, but data is available that, if analyzed to discover patterns, can guide such formulation.

Understandably, ML approaches have received a major boost in today's "big data" era. Approaches that employ probabilities, such as Bayesian inferencing[7], have become viable: prior probabilities that could only be guessed at previously (using highly subjective "expert judgment") can now be computed directly from data (*e.g.*, EHRs/public-health registries), with the caveat that these reflect local conditions – *e.g.*, incidence of specific infectious diseases – and will vary with the data source.

All data-driven approaches use iterative mathematical optimization techniques (originally pioneered by Isaac Newton and his contemporaries) to converge onto solutions. In ML parlance, the optimization process is called "training".

ML APPROACHES ARE SUBDIVIDED INTO

Statistical learning

The use of statistical methods to discover patterns or fit predictive models to data. These techniques originated in the late 19^{th} century (linear regression/correlation), though they have advanced to tackling vast numbers of input variables (also called "features" in ML) and vastly more diverse problems. Human expertise is involved in identifying the features (numeric or categorical) relevant to the problem, and in transforming them to a form suitable for analysis. (For example, a variable comprising of N categories – *e.g.*, gender/race – can be transformed into (N-1) one-or-zero variables using a simple technique called "one-hot encoding"[8]). Almost all statistical learning (SL) methods have been developed by researchers with an applied math/statistics background. Individual methods might make specific assumptions about the nature of the variables (*e.g.*, that they have a Gaussian distribution, or that their effects are additive).

Artificial neural networks

(The term "artificial" is typically implied and therefore usually dropped in both the full phrase and the abbreviation.) This family of approaches, which began in the 1950s, also results in the creation of predictive models. It is now prominent enough to deserve its own subsection, below.

Neural networks: Deep learning: Neural Networks (NNs) are inspired by the microstructural anatomy and functioning of animals' central nervous systems: software that simulates two or more layers of "neuron"-like computational units ("cells"). Each layer's cells send their output to cells in the next – and in approaches called "recurrent NNs", provide "feedback" to earlier layers as well. However, NNs employ mathematical techniques under the hood, notably mathematical "activation functions" for individual cells. The activation function for a neuron typically transforms inputs of large positive or negative numbers into outputs with a smaller range (*e.g.*, zero to one, or ± 1). An activation function may also incorporate a threshold, *i.e.*, the output is zero unless the input exceeds a particular value.

"Deep" NNs, their modern incarnation, have many more layers than older ("shallow") NNs. ("Deep learning" is ML performed by DNNs). NNs differ from Statistical learning in two ways.

NNs make few or no assumptions about variables' characteristics: their statistical distributions don't matter, and their inter-relationships may be non-linear (typically, unknown). Consequently, NNs may sometimes yield accurate predictive models where traditional SL fails.

While NNs can use human-expert-supplied features, they don't have to. For image input, DNNs can *discover* features directly from the raw pixels/voxels. The initial layer discovers basic feature such as regional lines, subsequent layers assemble these into shapes, and so on: LeCun *et al*'s classic Nature paper describes this process[9], which parallels the cat visual cortex's operation, as discovered by Nobelists David Hubel and Torsten Wiesel[10]. After training, the initial layers can be reused for other image-recognition problems, a phenomenon called *Transfer Learning* (TL)[11]: Starting training with layers that recognize basic features is faster than starting from scratch.

TL is also widely used in DNN-based natural language processing (NLP) for medical text: BERT[12], a giant DNN trained by a Google team on the entire contents of Wikipedia and Google Books, was used to bootstrap the training of BioBERT, trained on the full text of PubMed and PubMed Central[13]. Choudhary *et al*[14] review medical-imaging applications of *Domain adaptation*, a special case of TL, where a DNN trained on a set of labeled images (*e.g.*, relating to a particular medical condition) are reused for images for a different, but related, condition, either as-is or after an accelerated training process.

This gain in power isn't free. The number of computations involved goes up non-linearly with the number of layers[15], and so much more compute power is required: Notably, abundant random-access-memory (RAM) and the use of general-purpose Graphics Processing Units (GPUs)[16], which perform mathematical operations on sequences of numbers in parallel. (In fact, the theoretical advances embodied in diverse modern DNN architectures would be infeasible without powerful hardware).

DNNs require vastly more data than SL to discover reliable features which human experts may find obvious. Data volume isn't enough: One must also try to eliminate bias by using diverse data. (We address bias in section 3).

Certain arithmetic-based issues manifest when the number of layers becomes large - production DNNs can have hundreds of layers - and inputs from each layer pass to the next. Underneath the hood, numbers are being multiplied. When a large sequence of numbers that are all either larger or less than 1 get multiplied repeatedly, the product tends to infinity or to zero: For example, 2 multiplied by itself 64 times is approximately 1.88×10^{19}.

In DNNs, the consequences of repeated multiplication, called the "Exploding Gradient" or "Vanishing Gradient" problems, can thwart the training process. These are both prevented by batch normalization (BN), which re-adjusts the numerical values of all the outputs of each hidden layer during each iteration of the optimization training, so that the average of the outputs is zero and their standard deviation is one. Apart from speeding learning, BN allows more layers to be added to the DNN, and hence one can tackle harder problems.

Because of their performance characteristics - DNNs have achieved better accuracy than previous methods, on numerous benchmarks, in a variety of domains - most current AI research focuses on DNNs.

Table 1 summarizes the differences between the symbolic, statistical and DNN approaches.

Training in machine learning: ML models can be trained in one of two ways: Supervised Learning: The objective here is to predict a category (presence/absence or severity of a lesion/disease) or a numeric (interval) value. Category prediction is also called “classification”. The training data contains the answers: Either in the output variable/s for tabular data, or for images, human annotation/Labeling that identifies specific object categories (including their region of interest, if multiple categories coexist within an image).

Unsupervised Learning: Here, the objective is to discover patterns in the data, thereby achieving dimension reduction (*i.e.*, a more compact, parsimonious representation of the data).

Semi-supervised learning: The drawback of supervised learning is that for unstructured data (narrative text, images) annotation/Labeling is human-intensive, as well as costly if it involves human expertise that must be paid for. Semi-supervised learning uses a combination of (some) labeled and (mostly) unlabeled data, under the assumption that unlabeled data points close to (or in the same cluster as) labeled data points are likely to share the same category/class.

Statistical learning techniques can be either supervised or unsupervised. Examples of supervised techniques are: Multivariate linear regression/general linear models, which predict interval values; logistic regression and support vector machines, which predict categories; K-nearest neighbor and Classification and Regression Trees (CART), which predict either. Unsupervised SL methods include clustering algorithms, principal components/factor analysis and Latent Dirichlet Allocation.

DNNs, which need very large amounts of data, have motivated the development of semi-supervised methods. They are intrinsically suited for classification. For interval-value prediction with image data, they typically perform or assist in segmentation (which can work with/without supervision), after which numeric volumes can be computed from the demarcated voxels.

Preprocessing: Before training, the data is typically pre-processed with one or more steps. Pre-processing makes the training (and hence predictions) more reliable. The strategies used depend on the kind of data (numeric *vs* image). Some strategies are general, while others are problem specific (we occasionally refer to the latter). Among these steps are: Detecting suspected erroneous values including unrealistic outliers (*e.g.*, non-physiological clinical-parameter values). The adage “Garbage In, Garbage Out” applies to all facets of computing.

Replacing missing/erroneous values (“imputing”): An entire subfield of applied statistics is devoted to this problem. Strategies include picking the average value across all data points, average value for the individual patient, interpolated values (for time-series data), *etc.* In general, SL algorithms, many of which mandate either imputing all missing values or dropping the data point/s in question, are more vulnerable to missing values than DL.

Standardizing: Adjusting numeric values so that disparate variables are represented on the same scale. For variables with a Gaussian (“Normal”) distribution, each value is subtracted from the variable’s mean and the result divided by the variable’s standard deviation, with the sign preserved. For non-Gaussian variables, the value is subtracted from the median and divided by the inter-quartile range. (Batch normalization, discussed earlier, was inspired by standardizing).

For images, editing out artefacts extraneous to the content to be analyzed - *e.g.*, superimposed text labels or rulers to indicate object size. We come back to this issue later.

Sources of error: Overfitting and hidden stratification: A strength of DNNs, stated earlier, is their ability to discover features from raw data. Sometimes, this can also be a weakness: *Overfitting* occurs when any ML model is led astray by incidental but irrelevant features in the input. Apart from working unreliably with a new dataset, an overfitted model often making mistakes that humans never would. A DNN for diagnosing skin malignancies used a ruler/scale’s presence to infer cancerous lesions, whose dimensions are usually recorded diligently[17]. Similarly, textual labels on plain musculoskeletal radiographs were confused with internal-fixation implants, lowering accuracy[18].

Several strategies minimize the risk of overfitting, in addition to making reporting of results more honest: Cross-validation: The training data is partitioned into a certain number, N (*e.g.*, 10), of approximately equal slices. The training is conducted N times, each time sequentially withholding 1 slice (*i.e.*, only the remaining N-1 slices are used), and the results are averaged.

Withholding of test data from training: A portion of the data is completely withheld from the training process. After the ML model is fully trained with the training data, it is evaluated with the test data, and results are (or should be) reported against the test data only.

Regularization: This is a general term for computational techniques that reduce the likelihood of overfitting during the operation of the training algorithm’s optimization phase. The most well-known and general approach is to *penalize model complexity*: the fewer the number of variables that remain in the final trained model, the less the complexity. Originally applied to linear and logistic regression[19], where Lasso and Ridge Regression respectively include penalties that are linear and quadratic in the final number of variables, it is also used for DL.

Table 1 Comparison of symbolic artificial intelligence, statistical learning and deep learning (Nadkarni P & Merchant SA)

	Symbolic AI	Statistical learning (SL)	Deep learning (DL)
Entities manipulated	Both symbols and numbers	Numbers (most representing interval data, but some representing categories)	Same as SL, can be applied to the same problems
Algorithm design	Requires computer-science knowledge & traditional software skills, including user-interface design	Less customization needed, but problem-specific pre-processing of data (*e.g.*, statistical standard-ization is necessary)	Same as SL
Domain expert role	Work closely and extensively with software developer, Evaluate output of algorithm for a set of test cases against desired output	To identify variables/features of interest, annotating training data, and evaluating results and individual features' relative importance. Must evaluate results for novelty	Same as SL, but features can be discovered from raw data, so may not need designation. Annotation is more burdensome because much more data is typically needed
Data inputs	Expert and software work closely to design software and create test cases	Rows of data, annotated text, or images. For supervised learning, the output variable's value for each instance is also supplied	Same as SL, in some forms of DL, notably for image processing, features are discovered from raw data
Partitioning of input data	(Not applicable)	Divided into training data and test data	Same as SL
Generalizability	Limited to modest: Typically required tailored solutions, especially for the user interface	More generalizable than symbolic AI, but success depends on careful feature selection, choice of method and whether the data matches the method's assumptions (*e.g.*, Gaussian distribution, additive effects)	DL methods are "non-parametric" and rely on few or no assumptions about the variables/features in the data

AI: Artificial intelligence; SL: Statistical learning; DL: Deep learning.

A regularization approach specific to DLs is Dropout: disabling a certain fraction of neurons in hidden layers of a multilayer network during each cycle of training. Li *et al*[20] provide theoretical reasons why dropout can interfere with batch normalization, discussed above, resulting in performance degradation. They recommend that dropout be employed only after the last hidden layer where BN is used, and that the proportion of disabled neurons not exceed 50% (and should usually be much smaller).

A related problem, *Hidden Stratification*[21] occurs when a category contains sub-categories ("strata") unrecognized during problem analysis: here, performance on some strata may be poor. Thus, Rueckel *et al*[22] cite an example of severe pneumothorax being recognized accurately only in those images where a chest tube (inserted to provide an outlet for trapped air) is present[23]. While mild pneumothorax is treated conservatively without a tube, misdiagnosing a yet-to-be-treated, severe pneumothorax has serious consequences.

Nakkiran *et al*[24] had earlier observed the phenomenon of "*double descent.*" For some problems, when a DNN classifier is trained on increasingly larger datasets, performance intially gets worse. Later, when the training dataset has become much larger, performance gets better. This could be explained by hidden stratification. The somewhat-larger dataset is heterogenous in unconsidered ways, but the instances of minority sub-categories are too few to learn from, so they only serve to degrade performance. With much larger datasets, these instances become numerous enough to yield a signal that the DNN can use to discriminate more accurately.

The need for a holistic, system based approach

Most recent research in radiology AI has focused on DNNs: The following is just a brief list of DL applications. (This list is not intended to be comprehensive). Binary (Yes/no) classification: Elbow fractures[25], rib fractures[26], orthopedic implants[27], pneumothorax[28], pulmonary embolism[29], lung cancer[30], pulmonary tuberculosis (where several commercial applications exist)[31]. Multi-category classification (grading/staging): Anterior cruciate ligament injuries[32], hip fracture[33]. Segmentation with quantitation: Pulmonary edema[34], epicardial fat[35,36]; gliomas[37,38]; liver metastases[39,40]; spleen[41], and brain infarcts[42]. While impressive, much more is needed to apply AI to realistic problems, especially when intended for deployment in teleradiology scenarios where onsite skill/experience is often lacking. We summarize the issues here before discussing each issue in detail. The focus on DNN applications that perform only a single task, while proliferating the number of publications in the literature, does little to advance the likelihood of practical deployment. Depending on the problem, humans use multiple problem-solving strategies. Similarly, realistic solutions must combine multiple AI approaches, in addition to old-fashioned software engineering (such as intuitive and robust user interfaces). Good radiologists are also good clinicians. AI must be able to use all available evidence, including collective wisdom gained over decades of experience. Both humans and AI can be biased; this susceptibility must be recognized. Among the numerous ways to reduce bias, one must consider explainability – the ability to clearly describe the workings of a particular application to a subject-matter expert unfamiliar with AI technology.

The Limitations of Uni-tasking: As Krupinski notes[1], most DNNs in radiology uni-task. Thus, a DNN specialized for rib-fracture recognition will, even if outperforming radiologists, ignore concurrent tuberculosis, pneumothorax, or Flail Chest, unless trained for the same. For that matter, DNN tuberculosis (TB) diagnosis considering only consolidation/cavitation/mediastinal lymph nodes may miss TB in children. In one series of pediatric patients with pleural effusions, 22% had TB; in 41% of these, effusion was the only radiologic TB sign[43]. We have noticed that these effusions may be lamellar and track upwards, akin to pleural thickening, without being overtly visible, unlike the usual pleural effusions. In fact, in our experience, a lamellar effusion in a child is a good pointer towards the presence of a Primary Complex of TB.

No clinical radiologist uni-tasks: "Savant Syndrome" describes humans with exceptional skill in one area who are mentally challenged otherwise. Overspecialized DNNs suffer, in effect, from perceptual blindness. This phenomenon can be induced experimentally in normal humans by overwhelming their cognitive abilities: in a famous experiment, where subjects had to watch a basketball-game video and count the number of passes one team made, half the subjects failed to notice an intermingling gorilla-suited actor in the center of several scenes[44].

Based on general-purpose vision (GPV) studies, features learned in one specialized uni-tasking recognition problem (*e.g.*, cats) transfer poorly to a related problem (*e.g.*, recognizing horses). GPV has advanced because of the public availability of datasets, most notably ImageNet[45], which contain a vast number of object categories, often with multiple categories per image. The images are annotated by crowdsourcing: each object is indicated with a bounding box. Any DL approach expecting to perform well in a challenge to identify these objects cannot be over-specialized. (Unfortunately, DNNs trained on ImageNet perform very poorly with radiology images: Transfer learning is not guaranteed to work).

We believe that focusing short-term on research publications addressing relatively simple problems (with much research being PhD-thesis-driven) retards overall progress. Historically, symbolic AI's notorious addiction to this approach, accompanied by hype that greatly outpaced actual achievement, led to several "AI Winters"[46,47], steep funding drops following disillusionment. McDermott (a symbolic AI researcher) raised such concerns in a famous 1976 paper, "Artificial Intelligence Meets Natural Stupidity"[48].

Moving toward multi-tasking: There is no reason (besides the costs of compensating radiologists for their time) why radiographic modality-specific ImageNet equivalents cannot be created. Collections of images for trauma patients where multiple lesions are likely to be present may be a good starting point. One could also reuse the vast amount of existing annotated images for uni-tasking-DL research: Federated DL (see section 5.1) may help to test new, broader, lesion-recognition algorithms.

While DNNs excel at the important subtask of pattern recognition, they alone would not suffice to move radiology AI into the clinic, as now discussed.

The right strategy for the right subtask: Decades of research in cognitive psychology, especially observations of human expertise, have shown that humans use different strategies to different problems. In his classic, "Conceptual Blockbusting", Adams *et al*[49] identifies strategies as varied as: General-purpose critical thinking; knowledge of science and mathematics (including calculus); visualization; and applying ethical constraints.

The psychologists Daniel Kahneman and Amos Tversky, founders of "behavioral economics" (Kahneman got a Nobel– Tversky was deceased by then) postulate two modes of thinking. These are "System 1" – "lower level", rapid, intuitive, and reflex ("short-cut")– and "System 2" – "higher level", slow, deliberate, considering multiple sources of information, and requiring concentration. (We return to this work later.) As noted by Lawton[50], DNNs embody System 1 thinking, while statistical and symbolic approaches embody System 2. Both must be used together.

What applies to humans also applies to electronic systems. Symbolic, statistical and NN approaches have been combined in several ways: In new domains where little practical human experience has accumulated, statistical learning has led to discovery of patterns that can then be encoded as rules or in decision trees, which originated symbolic AI.

While symbolic AI can identify differential diagnosis for a given clinical presentation, statistical AI, using data from local sources or from the literature, can compute probabilities to rank these diagnoses, as well as sensitivity/positive predictive value of individual findings (including test results) to suggest the way forward.

Symbolic approaches are easier for human experts to understand (because they parallel deliberative human problem-solving approaches), and so are often used to "explain" patterns discovered by DNNs. (We discuss explainability in Section 4).

In radiology AI, Rudie *et al*[51] combine DNN with symbolic/statistical AI (Bayesian networks) for differential diagnosis of brain lesions. Doing this on a large scale across multiple radiology domains has the potential to improve clinical decision making.

Using all available evidence: In sufficiently diverse patient populations, attribution of diagnoses to detected radiographic lesions requires evidence from history, physical exam, non-radiology investigations, plus knowledge of prevalence. Our recommendation to combine all such information to make better decisions is not unique: Kwon *et al*[52] also suggest a Radiology AI that approach that combines multiple evidence sources (imaging plus clinical variables) for COVID-19 prognostication, while Jamshidi *et al*[53] also recommend a combined approach for COVID-19 diagnosis and treatment.

We provide examples below. An upper-lobe cavity on a chest X-ray could suggest neoplastic processes, mycobacterial infection, intracellular fungal infection (histoplasma, coccidiosis), *etc*. Serological confirmation plus newer technologies (*e.g.*, GenXPert for tuberculosis[54]) assist diagnosis.

The failure to elicit a proper history can be expensive and traumatizing. One of us (S.A.M.) encountered a young girl who had been repeatedly evaluated under general anesthesia for possible ectopic ureter localization, because of failure to make one simple observation on the plain radiograph. A subsequent Multidetector CT exam concluded erroneously that the incontinence was due to a vesicovaginal fistula, which is extremely rare in children, more so if acquired. This erroneous diagnosis could have been avoided by a simple observation (a slight gap in the pubic symphysis) and one simple question: When did symptoms start? (From birth). This suggested the correct diagnosis: female epispadias, which a pediatric surgeon confirmed.

Recognizing midline shift (MLS), plus trans-tentorial and other herniations, allows better triaging for intracranial bleeds or head trauma[55,56]). Xiao *et al*[57] describe an algorithm to MLS of the brain on CT, with a sensitivity of 94% and specificity of 100%, comparable to radiologists.

In head injury, ear-nose-throat bleeds/pneumocephalus suggest basilar skull fractures[58], which are non-displaced and difficult to detect unless looked for diligently.

Pneumothorax diagnosis by DNNs[59], while useful, could increase accuracy for Tension Pneumothorax by additionally looking for simple radiological signs like - inversion of the diaphragm, tracheal shift/shift of mediastinal structures to the opposite side (Figure 1).

AI for rib-fracture recognition[60] can be complemented by the clinical finding of "Flail Chest", which seriously impairs respiratory physiology[61] and may occur when three or more ribs are broken in at least two places.

Combining AI with other technologies: A major thrust of medical AI is in making other technologies, both existing and novel, much "smarter", reducing error by assisting manual tasks and decision-making performed by the radiologist or operator.

Applications in Interventional Radiology: The Royal Free Hospital in London employs an AI-backed keyhole procedure for stenting, coupled with Optical coherence tomography (OCT). While OCT allows viewing the inside of a blood vessel, the AI software automatically measures vessel diameter to enhance decision-making by the interventionist[62]. Similar roles are possible in interventions such as robotic intussusception–where visualization of the ileocecal junction and reflux into terminal ileum could be taken as end points of the procedure.

AI-assisted 3-D Printing of biological tissue such as heart valves, blood vessel grafts and possibly complete organs is discussed in[63].

BIASES IN RADIOLOGY

Artificial Intelligence needs real Intelligence to guide it. Truly intelligent humans are distinguished from the merely smart by intellectual humility and flexibility: as noted in Robson's "The Intellect Trap"[64], they constantly consider the possibility of being wrong, and abandon long-held beliefs when these are invalidated by new evidence. Tetlock's work on human expertise also emphasizes flexibility's importance; both in adapting to reality, as well as in problem-solving strategies. As discussed in section 2.2, AI approaches must be flexible too.

Tversky and Kahneman emphasize that, because of its reflex nature, System 1 thinking is prone to bias. Also, because System 2 requires sustained mental effort (which can cause fatigue), System 1 often contaminates System 2 thought, leading to errors or bias. Busby *et al*[65] cite this work in their excellent article on bias in radiology. An early paper by Egglin and Feinstein considers context bias in radiology [66], where certain aspects of patients' initial presentation to their clinicians led radiologists to give less weight to alternative diagnoses.

Electronic applications can be biased just as humans are. The sources of bias are several. Symbolic approaches may reflect the biases of their human creators. Machine-learning approaches that rely on humans to specify relevant features/input variables may be biased if the features chosen are inappropriate, or if relevant features are omitted.

If features are discovered entirely by DL, the data itself may be biased or non-representative. An early version of Facebook's artificial-vision system misidentified bare-chested black males as "primates"[67] because of too few samples in the training data.

DOI: 10.35711/aimi.v3.i3.55

Figure 1 Tension Pneumothorax computed tomography topogram. A large left Tension Pneumothorax herniating across the midline with a marked shift of the mediastinal structures to the opposite side. Arrowheads denote a displaced trachea. Image courtesy Dr. Anagha Joshi, Prof & Head (Radiology), LTMMC & LTMGH, Mumbai.

EXPLAINABILITY OF AI

Explainability is the ability to describe the internal workings of a particular AI model (which may apply one or more techniques to a practical problem) to a human expert who intimately knows the problem's-domain but not AI technology. Molnar's book on Interpretable ML[68] is an excellent reference. From this perspective, ML techniques are classified into *"white-box"* (explainable in terms resembling ordinary language), and *"black-box"* models, which cannot be readily explained, because they rely on complex mathematical functions/concepts.

What determines "Black-Box" vs "White-Box"?

Explainability is determined by the following factors: The choice of technique. In general, Symbolic AI (and techniques that display output as symbols, such as decision trees) are most understandable/explainable.

Statistical techniques are less explainable. Tversky and Kahneman found in their studies of cognitive errors that people find statistical concepts – such as the phenomenon of regression to the mean due to random processes– more difficult to understand than symbols. In the real-life example of the "Monty Hall problem"[69], at least 1000 PhDs, including the great mathematician Paul Erdos, had difficulty believing the correct answer, which is an application of Bayesian reasoning that causes a revision of posterior probabilities when new evidence arrives. Therefore, the explainer must often educate the human expert in statistics before addressing the specifics of the application.

In DNNs, the "explanation" is actually a large set of numbers, corresponding to the weights of the inputs of each "neuron" to the neurons to which it connects, along with descriptions of the mathematical transformation/s involved. This is so far removed from everyday experience as to be practically incomprehensible (though there is active research in converting this information into explanatory visuals).

The classification of a particular technique as "black-box" or "white-box" is somewhat arbitrary, depending on the beholder, and on the domain expert's background knowledge. For example, Loyola-Gonzales[70] classifies Support Vector Machines (SVMs) as "black-box". However, SVMs, developed by applied statistician Vladimir Vapnik's group at Bell Labs[71] , are mathematically very closely related to regression[72], but try to optimize a different mathematical function (maximized separation between instances of different classes *vs* minimized sum-of-least-squares deviations between observed and predicted values). Multivariate regression (linear, logistic, *etc.*) is taught in enough practically oriented college-level statistics courses for non-statisticians (*e.g.*, business majors, life scientists, medical researchers) to be widely understood.

The complexity of individual problems: Any model with hundreds of input variables (such as the regression models used by macro-economists) will be intrinsically hard to comprehend.

Business-Rule systems are naturally expressed in ordinary language, and so are in principle, highly explainable. However, R1, devised by McDermott[73] to configure Digital Equipment Equipment's VAX minicomputers based on a customer's needs, eventually used 2500 rules. Proving that a BRS is internally consistent - that is, no rule contradicts any other rule in the system- is known to be combinatorically hard. "Understanding" the principles of a large BRS does not make it any easier to debug if its output is incorrect.

Whether human-understandable input needs to be modified into an unfamiliar form to make it amenable to computation. This is the case with SVMs when employed for optical character recognition: the image of each letter is converted to a set of numeric features. In the extreme case, radiographic images are transformed by DNNs from individual pixels into hundreds of features that are "discovered" from the raw data, with each subsequent layer in the DNN representing composite features of increasing complexity.

The consequences of non-explainability

The concerns about explainability are closely tied to two risks: Bias: If you cannot explain the application (to a human expert, or to a jury if the application's use is challenged legally), how can you show that it is not biased? "Because the computer says so" is unpersuasive.

Failure: DNNs that process images often make unexplained, bizarre mistakes – misidentifications or failure to identify, as noted by Heaven D[74]. Explanations for such mistakes' origins are not obvious in "post-mortems" even to DNN experts. One approach to forestalling such errors is to deliberately attempt to fool image-classification DNNs by generating "fakes" using another "adversary" DNN to make tweaks (minor or not-so-minor) to authentic images, which are then supplied as training input to the classification-DNN[75]. However, while adversarial networks have reduced misidentifications, they do not offer cast-iron guarantees that a mistake will never be made. As in the cliché, absence of evidence (of defects) is not evidence of absence.

Failure can have consequences ranging from the merely frustrating to the near-apocalyptic. A famous example of the latter was the Soviets' satellite-based Early-Missile-Warning System, which, in 1983, flagged 5 missiles from US sites heading toward the USSR[76]. A retaliatory nuclear strike, which would have started World War 3, was averted by Lt. Col. Stanislav Petrov, who reasoned that this was a false alarm – an intentional US attack would need many more missiles – and disobeyed standing orders (to relay the warning up the command-chain) by deciding to wait for confirming evidence, which never arrived.

Approaches toward making "Black-Box" AI more explainable

In general, such approaches are specific to the problem being addressed, as Molnar makes clear. One can show the impact of the values of individual input variables/features on the output variable (*e.g.*, categorization, risk score) using a technique called Deep Taylor Decomposition (DTD)[77], based on the Taylor series taught in intermediate-level Calculus. Lauritsen *et al*[78] use DTD as part of an explanation module for predicting four categories of acute critical illness in inpatients based on EHR data. DTD works when the number of input variables is modest (this paper used 33 clinical parameters), and the variables correspond to concepts in the domain. It would not be useful for very numerous, transformed, or automatically discovered variables.

Sometimes, a detailed technical explanation may not be necessary: one can simply test with enough test cases where the system's output matched that of human experts. For images, delineating areas of interest with highlight boxes can draw the user's attention. (This is a standard technique employed by object-recognition systems on benchmark datasets such as ImageNet). This technique has the drawback that in case of erroneous diagnosis, merely drawing the user's attention to regions of interest may not suffice.

Also, "absence of evidence is not evidence of absence". For a "black-box" system with a critical bug that manifests under uncommon circumstances, you will discover the problem only when it happens. In a complex-system (non-AI) context, Jon Bentley, in his classic work "Programming Pearls"[79] cites a colleague who implemented what he thought was a performance optimization in a FORTRAN compiler. Two years later, the compiler crashed during use. The colleague traced the crash to his "optimization", which had never been invoked in the interim and crashed the very first time it was activated in production.

Loyola-Gonzales[70] suggests combining a white-box and black-box approach (the order depending on the problem) in a pipeline, so that the output of the first is processed into a more human-understandable approach by the second.

Regulatory concerns

Certain software applications for tasks previously requiring specialized human skills have already received FDA approval and are in wide use. For example, smartphone-deployable electrocardiogram (EKG)-interpretation programs report standard EKG parameters as well as a few abnormal signals such as Ventricular Premature Beats. Given the increasing deployment of Software as a Medical Device (SaMD), and the possibility of catastrophic medical error when operated (semi-) autonomously, national regulatory bodies are naturally concerned about standardizing the processes of development and testing of SaMD to prevent such errors.

The FDA has specified an action plan, including guidelines for best ML practices, version control when the algorithm is changed, and protection of patient data[80]. The European Commission's proposal for regulation is much wider, encompassing uses of AI across all of society[81]: Human Rights Watch has criticized this proposal[82] on the grounds that it currently does not offer sufficient protection for the social safety net when such software functions autonomously to make decisions concerning, for example, eligibility of individuals for benefits.

FUTURE DIRECTIONS

Federated machine learning

ML in general, and DL specifically, need lots of data to achieve desired accuracy. Volume alone does not suffice: the data must also be sufficiently diverse (*i.e.*, coming from multiple locales) to minimize bias. The obvious solution, physical pooling of data. faces the following barriers: Data privacy - which is less of an issue with digital radiography, where DICOM metadata containing identifiable information can be removed. Mistrust – a formidable hurdle when academic or commercial consortia bring rivals together. The technique of *Federated Learning* (FL), originally pioneered by Google as an application of their well-known MapReduce algorithm[83] allows iteratively training an ML model across geographically separated hardware: The ML algorithm is distributed, while data remains local, thereby ensuring data privacy. It can be employed for both statistical and deep learning.

Typically, a central server coordinates computations across multiple distributed clients. At start-up, the server sends the clients initialization information. The clients commence computation. When each client is done, it sends its results back to the server, which collates all clients' results. For the next iteration, the server sends updates to each client, which then computes again. The process continues until the ML training completes convergence.

FL's drawbacks are Internet-based communication overhead, which limits training speed, and greater difficulty of analysis of any detected residual bias. Ng *et al*[84] provide a detailed technology overview. Sheller *et al*[85] use FL to replicate prior analysis of a 10-institution brain-tumor-image-dataset derived from The Cancer Genome Atlas (TCGA). Sarma *et al*[86] describe 3-institution FL-based training on whole-prostate segmentation from MRIs, while Navia-Vazquez *et al*[87] describe an approach for Federated Logistic Regression.

In balance, FL's finessing of data privacy issues enables addressing of problems at scales not previously possible, with the greater data volume and diversity ensuring better accuracy and generalizability.

Quantum computing

See our previous work, Merchant *et al*[88], for an exploration of this rapidly progressing and revolutionary field. Here, we only provide a basic introduction and address some issues not covered in that paper.

Quantum mechanics describes the rules governing the properties and behavior of matter at the molecular and subatomic levels. Established technologies such as digital photography and nuclear radiography (based on the photoelectric effect), the integrated circuit (based on semi-conduction of electricity by certain materials), and the laser (based on coherent emission of photons) are all applications of quantum mechanics.

Quantum computing (QC) uses the phenomenon of *quantum superposition*, in which matter at the atomic/subatomic level can exist (briefly) in two different states simultaneously, as the basis for computing hardware design. Unlike the bit in an ordinary computer, which can be either 1 or 0, the quantum bit ("qubit") can be both 1 and 0 simultaneously, so that an array of N qubits could represent 2^N states simultaneously.

QC can, in theory, help solve certain computational problems (called NP-hard problems, where NP = "non-deterministic polynomial"[89]). The time taken to solve an NP-hard problem by brute force (*i.e.*, trying out every possible solution, which is the only way to solve such a problem exactly) increases exponentially as the problem size grows linearly. For example, cracking the widely used Advanced Encryption Standard-256 (with 256 bits) would take all the world's (non-quantum) computers working together, longer than the age of the Universe. In 1994, Peter Shor's theoretical work[90] showed that a "quantum computer" with enough qubits could solve a particular NP-hard problem (factoring the product of 2 large prime numbers, used in AES-256) in polynomial time, making cryptographic attacks feasible.

The physical challenge is to maintain the qubits stable for a sufficiently long time to accomplish some computation (thus far, such stability has been achieved at temperatures close to absolute zero). In addition, for a computer based on qubits, prototypical work suggests that replacing the conducting elements (the interconnecting wires in an integrated circuit) with light-conducting elements (so-called optical computing[91]) may be the way forward[92].

There are also theoretical considerations as to the kinds of problems for which QC will offer benefits. Thus, Aaronson[93] points out that we don't yet know if the class of problems involved in the optimization (training) phase of DNNs will benefit: while we can hope that they do, the simulations must still be performed to show that this will be the case. Similar concerns are echoed by Sarma[94], who expresses uncertainty about the timeline for QC to become commercially feasible.

Despite the risks of hype and disillusion, it may be worth remembering Arthur C. Clarke's dictum about the future: "If an elderly but distinguished scientist says that something is possible, he is almost certainly right; but if he says that it is impossible, he is very probably wrong"[95]. If quantum computing becomes commercially viable, almost every aspect of computing (and therefore, every technology that depends on computing) will benefit vastly. The Quantum Internet, Intelligent Edge devices, Edge Computing, Quantum Artificial Intelligence, Quantum Artificial Intelligence Algorithms and their applications in Augmented Reality/Virtual Reality and a more immersive Metaverse experience (for teaching/simulations, actual interactions *etc.*); are some of the exciting future developments/enhancements based on Quantum Computing that we have discussed in our previous paper.

CONCLUSION

Combining the wisdom (of both knowledge and meta-knowledge – *i.e.*, problem-solving strategies) gained over the years, with the tremendous versatility of AI algorithms will maximize the utility of AI applications in medical imaging for everyday clinical care. However, scaling up the use of multiple algorithmic strategies and sources of evidence is challenging. Because of its sheer diversity and volume, radiologists' experiential knowledge is very hard to encode in a form that allows instant retrieval. This difficulty applies even to its subset, "artificial general intelligence" (AGI), also known as "common sense". Common sense, apart from being not so common across humans, turns out to be surprisingly hard to implement, because of the sheer breadth of information that must be encoded into computable form.

We see two ways forward: The first long-term and less feasible, the second possible today. Allocating massive effort and resources to create medical/radiology AGI. Using software technology (including AI) to extend the human mind, much as access to Web search engines has vastly democratized access to considerable specialized knowledge.

In the latter approach, AI technology can be ubiquitous, integrated, and often functioning behind the scenes for tedious, monotonous and time-consuming tasks (as suggested by Krupinski[1], but still leaving humans in control of critical decisions.

FOOTNOTES

Author contributions: Both authors have contributed equally.

Conflict-of-interest statement: All the authors declare that they have no conflict of interest.

Country/Territory of origin: India

ORCID number: Prakash Nadkarni 0000-0002-9628-4700; Suleman Adam Merchant 0000-0001-6513-450X.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Krupinski EA**. Artificial Intelligence and Teleradiology: Like It or Leave It? 2019. Southwest Telehealth Resource Center Blog. Last Accessed: Dec 9, 2021. Available from: https://southwesttrc.org/blog/2019/artificial-intelligence-teleradiology-it-or-leave-it
2 **Ross R**. Principles of the Business Rule Approach. Pearson Education Inc: Boston; 2003. 372 p. ISBN: 020-178-8934
3 **Browne P**. JBoss Drools Business Rules. Packt Publishing: Birmingham, UK; 2009 ISBN: 9781847196064
4 **Apt KR**. Principles of Constraint Programming. Cambridge University Press: Cambridge, UK; 2003. ISBN: 052-182-5830
5 **Stevenson WJ**. Operations Management. 14th Ed. ed. McGraw-Hill: New York, NY; 2020. ISBN: 978-1260238891
6 **Frontline Solvers**. Solver Technology: Optimization. 2021. Last Accessed: Dec 10, 2021. Available from:

https://www.solver.com/solver-technology
7 **Pearl J**. Causality: Models, Reasoning, and Inference. Cambridge University Press: Cambridge, UK; 2000. ISBN: 978-0-521-77362-1
8 **Brownlee J**. Ordinal and One-Hot Encodings for Categorical Data. 2020. Last Accessed: Dec 1, 2021. Available from: https://machinelearningmastery.com/one-hot-encoding-for-categorical-data/
9 **LeCun Y**, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436-444 [PMID: 26017442 DOI: 10.1038/nature14539]
10 **Hubel DH**, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 1962; **160**: 106-154 [PMID: 14449617 DOI: 10.1113/jphysiol.1962.sp006837]
11 **Brownlee J**. A Gentle Introduction to Transfer Learning for Deep Learning. 2019. Last Accessed: Dec 1, 2021. Available from: https://machinelearningmastery.com/transfer-learning-for-deep-learning/
12 **Jacob D**, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2019. Available from: https://arxiv.org/abs/1810.04805v2
13 **Lee J**, Yoon W, Kim S, Kim D, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020; **36**: 1234-1240 [PMID: 31501885 DOI: 10.1093/bioinformatics/btz682]
14 **Choudhary A**, Tong L, Zhu Y, Wang MD. Advancing Medical Imaging Informatics by Deep Learning-Based Domain Adaptation. *Yearb Med Inform* 2020; **29**: 129-138 [PMID: 32823306 DOI: 10.1055/s-0040-1702009]
15 **Sze V**, Chen Y-H, Yang T-J, Emer J. Efficient Processing of Deep Neural Networks. Morgan & Claypool Publishers: New York, NY; 2020. 342 p. ISBN: 978-1681738352
16 **Wikipedia**. Graphical Processing Unit. 2021. Available from: https://en.wikipedia.org/wiki/Graphics_processing_unit
17 **Winkler JK**, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, Thomas L, Lallas A, Blum A, Stolz W, Haenssle HA. Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition. *JAMA Dermatol* 2019; **155**: 1135-1141 [PMID: 31411641 DOI: 10.1001/jamadermatol.2019.1735]
18 **Yi PH**, Malone PS, Lin CT, Filice RW. Deep Learning Algorithms for Interpretation of Upper Extremity Radiographs: Laterality and Technologist Initial Labels as Confounding Factors. *AJR Am J Roentgenol* 2022; **218**: 714-715 [PMID: 34755522 DOI: 10.2214/AJR.21.26882]
19 **Harrell FE**. Regression Modeling Strategies with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis 2nd Ed. ed. Springer; 2015. ISBN: 978-3319194240
20 **Li X**, Chen S, Hu X, Yang J. Understanding the Disharmony Between Dropout and Batch Normalization by Variance Shift. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2019; 2682-2690. Available from: https://openaccess.thecvf.com/content_CVPR_2019/html/Li_Understanding_the_Disharmony_Between_Dropout_and_Batch_Normalization_by_Variance_CVPR_2019_paper.html
21 **Oakden-Rayner L**, Dunnmon J, Carneiro G, Ré C. Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging. *Proc ACM Conf Health Inference Learn (2020)* 2020; **2020**: 151-159 [PMID: 33196064 DOI: 10.1145/3368555.3384468]
22 **Rueckel J**, Trappmann L, Schachtner B, Wesp P, Hoppe BF, Fink N, Ricke J, Dinkel J, Ingrisch M, Sabel BO. Impact of Confounding Thoracic Tubes and Pleural Dehiscence Extent on Artificial Intelligence Pneumothorax Detection in Chest Radiographs. *Invest Radiol* 2020; **55**: 792-798 [PMID: 32694453 DOI: 10.1097/RLI.0000000000000707]
23 **Oakden-Rayner L**, Dunnmon J, Carneiro G, Ré C. Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging. *Proc ACM Conf Health Inference Learn (2020)* 2020; **2020**: 151-159 [PMID: 33196064 DOI: 10.1145/3368555.3384468]
24 **Nakkiran P**, Kaplun G, Bansal Y, Yang T, Barak B, Sutskever I. Deep Double Descent: Where Bigger Models and More Data Hurt. 2019. [DOI: 10.48550/arXiv.1912.02292]
25 **Rayan JC**, Reddy N, Kan JH, Zhang W, Annapragada A. Binomial Classification of Pediatric Elbow Fractures Using a Deep Learning Multiview Approach Emulating Radiologist Decision Making. *Radiol Artif Intell* 2019; **1**: e180015 [PMID: 33937781 DOI: 10.1148/ryai.2019180015]
26 **Wu M**, Chai Z, Qian G, Lin H, Wang Q, Wang L, Chen H. Development and Evaluation of a Deep Learning Algorithm for Rib Segmentation and Fracture Detection from Multicenter Chest CT Images. *Radiol Artif Intell* 2021; **3**: e200248 [PMID: 34617026 DOI: 10.1148/ryai.2021200248]
27 **Patel R**, Thong EHE, Batta V, Bharath AA, Francis D, Howard J. Automated Identification of Orthopedic Implants on Radiographs Using Deep Learning. *Radiol Artif Intell* 2021; **3**: e200183 [PMID: 34350407 DOI: 10.1148/ryai.2021200183]
28 **Thian YL**, Ng D, Hallinan JTPD, Jagmohan P, Sia SY, Tan CH, Ting YH, Kei PL, Pulickal GG, Tiong VTY, Quek ST, Feng M. Deep Learning Systems for Pneumothorax Detection on Chest Radiographs: A Multicenter External Validation Study. *Radiol Artif Intell* 2021; **3**: e200190 [PMID: 34350409 DOI: 10.1148/ryai.2021200190]
29 **Pan I**. Deep Learning for Pulmonary Embolism Detection: Tackling the RSNA 2020 AI Challenge. *Radiol Artif Intell* 2021; **3**: e210068 [PMID: 34617031 DOI: 10.1148/ryai.2021210068]
30 **Jacobs C**, Setio AAA, Scholten ET, Gerke PK, Bhattacharya H, M Hoesein FA, Brink M, Ranschaert E, de Jong PA, Silva M, Geurts B, Chung K, Schalekamp S, Meersschaert J, Devaraj A, Pinsky PF, Lam SC, van Ginneken B, Farahani K. Deep Learning for Lung Cancer Detection on Screening CT Scans: Results of a Large-Scale Public Competition and an Observer Study with 11 Radiologists. *Radiol Artif Intell* 2021; **3**: e210027 [PMID: 34870218 DOI: 10.1148/ryai.2021210027]
31 **Tavaziva G**, Harris M, Abidi SK, Geric C, Breuninger M, Dheda K, Esmail A, Muyoyeta M, Reither K, Majidulla A, Khan AJ, Campbell JR, David PM, Denkinger C, Miller C, Nathavitharana R, Pai M, Benedetti A, Ahmad Khan F. Chest X-ray Analysis With Deep Learning-Based Software as a Triage Test for Pulmonary Tuberculosis: An Individual Patient Data Meta-Analysis of Diagnostic Accuracy. *Clin Infect Dis* 2022; **74**: 1390-1400 [PMID: 34286831 DOI: 10.1093/cid/ciab639]
32 **Namiri NK**, Flament I, Astuto B, Shah R, Tibrewala R, Caliva F, Link TM, Pedoia V, Majumdar S. Deep Learning for Hierarchical Severity Staging of Anterior Cruciate Ligament Injuries from MRI. *Radiol Artif Intell* 2020; **2**: e190207 [PMID: 32793889 DOI: 10.1148/ryai.2020190207]
33 **Krogue JD**, Cheng KV, Hwang KM, Toogood P, Meinberg EG, Geiger EJ, Zaid M, McGill KC, Patel R, Sohn JH, Wright

A, Darger BF, Padrez KA, Ozhinsky E, Majumdar S, Pedoia V. Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning. *Radiol Artif Intell* 2020; **2**: e190023 [PMID: 33937815 DOI: 10.1148/ryai.2020190023]

34 **Horng S**, Liao R, Wang X, Dalal S, Golland P, Berkowitz SJ. Deep Learning to Quantify Pulmonary Edema in Chest Radiographs. *Radiol Artif Intell* 2021; **3**: e190228 [PMID: 33937857 DOI: 10.1148/ryai.2021190228]

35 **Commandeur F**, Goeller M, Razipour A, Cadet S, Hell MM, Kwiecinski J, Chen X, Chang HJ, Marwan M, Achenbach S, Berman DS, Slomka PJ, Tamarappoo BK, Dey D. Fully Automated CT Quantification of Epicardial Adipose Tissue by Deep Learning: A Multicenter Study. *Radiol Artif Intell* 2019; **1**: e190045 [PMID: 32090206 DOI: 10.1148/ryai.2019190045]

36 **Schoepf UJ**, Abadia AF. Greasing the Skids: Deep Learning for Fully Automated Quantification of Epicardial Fat. *Radiol Artif Intell* 2019; **1**: e190140 [PMID: 33937806 DOI: 10.1148/ryai.2019190140]

37 **Eijgelaar RS**, Visser M, Müller DMJ, Barkhof F, Vrenken H, van Herk M, Bello L, Conti Nibali M, Rossi M, Sciortino T, Berger MS, Hervey-Jumper S, Kiesel B, Widhalm G, Furtner J, Robe PAJT, Mandonnet E, De Witt Hamer PC, de Munck JC, Witte MG. Robust Deep Learning-based Segmentation of Glioblastoma on Routine Clinical MRI Scans Using Sparsified Training. *Radiol Artif Intell* 2020; **2**: e190103 [PMID: 33937837 DOI: 10.1148/ryai.2020190103]

38 **Wu S**, Li H, Quang D, Guan Y. Three-Plane-assembled Deep Learning Segmentation of Gliomas. *Radiol Artif Intell* 2020; **2**: e190011 [PMID: 32280947 DOI: 10.1148/ryai.2020190011]

39 **Nakamura Y**, Higaki T, Tatsugami F, Zhou J, Yu Z, Akino N, Ito Y, Iida M, Awai K. Deep Learning-based CT Image Reconstruction: Initial Evaluation Targeting Hypovascular Hepatic Metastases. *Radiol Artif Intell* 2019; **1**: e180011 [PMID: 33937803 DOI: 10.1148/ryai.2019180011]

40 **Vorontsov E**, Cerny M, Régnier P, Di Jorio L, Pal CJ, Lapointe R, Vandenbroucke-Menu F, Turcotte S, Kadoury S, Tang A. Deep Learning for Automated Segmentation of Liver Lesions at CT in Patients with Colorectal Cancer Liver Metastases. *Radiol Artif Intell* 2019; **1**: 180014 [PMID: 33937787 DOI: 10.1148/ryai.2019180014]

41 **Humpire-Mamani GE**, Bukala J, Scholten ET, Prokop M, van Ginneken B, Jacobs C. Fully Automatic Volume Measurement of the Spleen at CT Using Deep Learning. *Radiol Artif Intell* 2020; **2**: e190102 [PMID: 33937830 DOI: 10.1148/ryai.2020190102]

42 **Christensen S**, Mlynash M, MacLaren J, Federau C, Albers GW, Lansberg MG. Optimizing Deep Learning Algorithms for Segmentation of Acute Infarcts on Non-Contrast Material-enhanced CT Scans of the Brain Using Simulated Lesions. *Radiol Artif Intell* 2021; **3**: e200127 [PMID: 34350404 DOI: 10.1148/ryai.2021200127]

43 **Merino JM**, Carpintero I, Alvarez T, Rodrigo J, Sánchez J, Coello JM. Tuberculous pleural effusion in children. *Chest* 1999; **115**: 26-30 [PMID: 9925059 DOI: 10.1378/chest.115.1.26]

44 **Most SB**, Simons DJ, Scholl BJ, Jimenez R, Clifford E, Chabris CF. How not to be seen: the contribution of similarity and selective ignoring to sustained inattentional blindness. *Psychol Sci* 2001; **12**: 9-17 [PMID: 11294235 DOI: 10.1111/1467-9280.00303]

45 **ImageNet**. org. ImageNet: About. 2021. Last Accessed: Dec 10, 2021. Available from: https://image-net.org/about.php

46 **Shead S**. Researchers: Are we on the cusp of an ‘AI winter’? BBC News. 2020. Available from: https://www.bbc.com/news/technology-51064369

47 **Wikipedia**. AI Winter. 2020. Available from: https://en.wikipedia.org/wiki/AI_winter

48 **McDermott D**. Artificial Intelligence meets Natural Stupidity. ACM SIGART Bulletin. 1976; 57: 4-9 [DOI: 10.1145/1045339.1045340]

49 **Adams JL**. Conceptual Blockbusting (4th Ed). Basic Bookks: San Francisco, CA; 2001. ISBN: 978-0738205373

50 **Lawton G**. Neuro-symbolic AI emerges as powerful new approach. 2020. Last Accessed: Dec 10, 2021. Available from: https://searchenterpriseai.techtarget.com/feature/Neuro-symbolic-AI-seen-as-evolution-of-artificial-intelligence

51 **Rudie JD**, Rauschecker AM, Xie L, Wang J, Duong MT, Botzolakis EJ, Kovalovich A, Egan JM, Cook T, Bryan RN, Nasrallah IM, Mohan S, Gee JC. Subspecialty-Level Deep Gray Matter Differential Diagnoses with Deep Learning and Bayesian Networks on Clinical Brain MRI: A Pilot Study. *Radiol Artif Intell* 2020; **2**: e190146 [PMID: 33937838 DOI: 10.1148/ryai.2020190146]

52 **Kwon YJF**, Toussie D, Finkelstein M, Cedillo MA, Maron SZ, Manna S, Voutsinas N, Eber C, Jacobi A, Bernheim A, Gupta YS, Chung MS, Fayad ZA, Glicksberg BS, Oermann EK, Costa AB. Combining Initial Radiographs and Clinical Variables Improves Deep Learning Prognostication in Patients with COVID-19 from the Emergency Department. *Radiol Artif Intell* 2021; **3**: e200098 [PMID: 33928257 DOI: 10.1148/ryai.2020200098]

53 **Jamshidi MB**, Lalbakhsh A, Talla J, Peroutka Z, Hadjilooei F, Lalbakhsh P, Jamshidi M, Spada L, Mirmozafari M, Dehghani M, Sabet A, Roshani S, Bayat-Makou N, Mohamadzade B, Malek Z, Jamshidi A, Kiani S, Hashemi-Dezaki H, Mohyuddin W. Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment. *IEEE Access* 2020; **8**: 109581-109595 [PMID: 34192103 DOI: 10.1109/ACCESS.2020.3001973]

54 **Helb D**, Jones M, Story E, Boehme C, Wallace E, Ho K, Kop J, Owens MR, Rodgers R, Banada P, Safi H, Blakemore R, Lan NT, Jones-López EC, Levi M, Burday M, Ayakaka I, Mugerwa RD, McMillan B, Winn-Deen E, Christel L, Dailey P, Perkins MD, Persing DH, Alland D. Rapid detection of Mycobacterium tuberculosis and rifampin resistance by use of on-demand, near-patient technology. *J Clin Microbiol* 2010; **48**: 229-237 [PMID: 19864480 DOI: 10.1128/JCM.01463-09]

55 **Bartels RH**, Meijer FJ, van der Hoeven H, Edwards M, Prokop M. Midline shift in relation to thickness of traumatic acute subdural hematoma predicts mortality. *BMC Neurol* 2015; **15**: 220 [PMID: 26496765 DOI: 10.1186/s12883-015-0479-x]

56 **Chiewvit P**, Tritakarn SO, Nanta-aree S, Suthipongchai S. Degree of midline shift from CT scan predicted outcome in patients with head injuries. *J Med Assoc Thai* 2010; **93**: 99-107 [PMID: 20196418]

57 **Xiao F**, Liao CC, Huang KC, Chiang IJ, Wong JM. Automated assessment of midline shift in head injury patients. *Clin Neurol Neurosurg* 2010; **112**: 785-790 [PMID: 20663606 DOI: 10.1016/j.clineuro.2010.06.020]

58 **Simon LV**, Newton EJ. Basilar Skull Fractures. StatPearls. Treasure Island (FL) 2022. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29261908

59 **Rueckel J**, Huemmer C, Fieselmann A, Ghesu FC, Mansoor A, Schachtner B, Wesp P, Trappmann L, Munawwar B, Ricke J, Ingrisch M, Sabel BO. Pneumothorax detection in chest radiographs: optimizing artificial intelligence system for

accuracy and confounding bias reduction using in-image annotations in algorithm training. *Eur Radiol* 2021; **31**: 7888-7900 [PMID: 33774722 DOI: 10.1007/s00330-021-07833-w]

60 **Meng XH**, Wu DJ, Wang Z, Ma XL, Dong XM, Liu AE, Chen L. A fully automated rib fracture detection system on chest CT images and its impact on radiologist performance. *Skeletal Radiol* 2021; **50**: 1821-1828 [PMID: 33599801 DOI: 10.1007/s00256-021-03709-8]

61 **Perera TB**, King KC. Flail Chest. StatPearls. Treasure Island (FL) 2022. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30475563

62 **Mageit S**. Artificial intelligence delivers boost for heart attack patients. NHS Royal Free London NHS Foundation Trust 2021. Last Accessed: May 18, 2021. Available from: https://www.royalfree.nhs.uk/news-media/news/artificial-intelligence-delivers-boost-for-heart-attack-patients/

63 **Ewumi O**. AI in 3-D Bioprinting. 2021. Last Accessed: Jan 18, 2022. Available from: https://dataconomy.com/2021/04/ai-in-3d-bioprinting/

64 Robson D. The Intellect Trap. W. W. Norton & Company; 2019. ISBN: 978-0393651423

65 **Busby LP**, Courtier JL, Glastonbury CM. Bias in Radiology: The How and Why of Misses and Misinterpretations. *Radiographics* 2018; **38**: 236-247 [PMID: 29194009 DOI: 10.1148/rg.2018170107]

66 **Egglin TK**, Feinstein AR. Context bias. A problem in diagnostic radiology. *JAMA* 1996; **276**: 1752-1755 [PMID: 8940325 DOI: 10.1001/jama.276.21.1752]

67 **Mac R**. Facebook Apologizes After A.I. Puts 'Primates' Label on Video of Black Men. New York Times. 2021. Sept 3, 2021. Available from: https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html

68 **Molnar C**. Interpretable machine learning: A Guide for Making Black Box Models Explainable; 2019

69 **Wikipedia**. Monty Hall Problem. 2018. Available from: https://en.wikipedia.org/wiki/Monty_Hall_problem

70 **Loyola-Gonzalez O**. Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View. *IEEE access* 2019; **7**: 154096-154113 [DOI: 10.1109/ACCESS.2019.2949286]

71 **Cortes C**, Vapnik V. Support-Vector Networks. Machine Learning. 1995; 20(3):273–97. [DOI: 10.1007/BF00994018]

72 **Hastie T**, Tibshirani R, Friedman J. Elements of Statistical Learning: Data Mining, Inference and Prediction (2nd Ed). Springer: Stanford, CA; 2008

73 **McDermott J**. R1: An Expert in the Computer Systems Domain. Proceedings of the First AAAI Conference on Artificial Intelligence; Stanford, California 1980. Available from: https://aaai.org/Papers/AAAI/1980/AAAI80-076.pdf

74 **Heaven D**. Why deep-learning AIs are so easy to fool. *Nature* 2019; **574**: 163-166 [PMID: 31597977 DOI: 10.1038/d41586-019-03013-5]

75 **Brownlee J**. A Gentle Introduction to Generative Adversarial Networks (GANs). 2019. Last Accessed: Dec 1, 2021. Available from: https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/

76 **Wikipedia**. 1983 Soviet nuclear false alarm incident. 2021. Available from: https://en.wikipedia.org/wiki/1983_Soviet_nuclear_false_alarm_incident

77 **Montavon G**, Bach S, Binder A, Samek W, Muller K-R. Explaining NonLinear Classification Decisions with Deep Taylor Decomposition. Pattern Recognition. 2017; 65: 211-222. Available from: https://arxiv.org/pdf/1512.02479.pdf

78 **Lauritsen SM**, Kristensen M, Olsen MV, Larsen MS, Lauritsen KM, Jørgensen MJ, Lange J, Thiesson B. Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nat Commun* 2020; **11**: 3852 [PMID: 32737308 DOI: 10.1038/s41467-020-17431-x]

79 **Bentley JL**. Programming Pearls (2nd edition). Addison-Wesley: Reading, MA; 1999. ISBN: 0-201-11889-0

80 **US Food and Drug Administration**. Artificial Intelligence and Machine Learning in Software as a Medical Device: Action Plan. 2019 (updated 2021). Last Accessed: 4/1/2022. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

81 **European Commission**. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. April 21 2021. Last Accessed: 4/1/2022. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

82 **Human Rights Watch**. How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net: Questions and Answers. 2022. Last Accessed: 4/1/2022. Available from: https://www.hrw.org/news/2021/11/10/how-eus-flawed-artificial-intelligence-regulation-endangers-social-safety-net

83 **Dean J**, Ghemawat S. editors. MapReduce: Simplified Data Processing on Large Clusters Sixth Symposium on Operating System Design and Implementation; 2004. Available from: http://static.googleusercontent.com/media/research.google.com/en/us/archive/mapreduce-osdi04.pdf

84 **Ng D**, Lan X, Yao MM, Chan WP, Feng M. Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets. *Quant Imaging Med Surg* 2021; **11**: 852-857 [PMID: 33532283 DOI: 10.21037/qims-20-595]

85 **Sheller MJ**, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, Milchenko M, Xu W, Marcus D, Colen RR, Bakas S. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 2020; **10**: 12598 [PMID: 32724046 DOI: 10.1038/s41598-020-69250-1]

86 **Sarma KV**, Harmon S, Sanford T, Roth HR, Xu Z, Tetreault J, Xu D, Flores MG, Raman AG, Kulkarni R, Wood BJ, Choyke PL, Priester AM, Marks LS, Raman SS, Enzmann D, Turkbey B, Speier W, Arnold CW. Federated learning improves site performance in multicenter deep learning without data sharing. *J Am Med Inform Assoc* 2021; **28**: 1259-1264 [PMID: 33537772 DOI: 10.1093/jamia/ocaa341]

87 **Navia-Vazquez A**, Vazquez-Lopez M, Cid-Sueiro J. Double Confidential Federated Machine Learning Logistic Regression for Industrial Data Platforms. Proceedings of the 37th International Conference on Machine Learning; Vienna, Austria 2020. Available from: http://www.tsc.uc3m.es/~navia/FL-ICML2020/DCFML-FL_ICML2020-A_Navia_Vazquez_et_al.pdf

88 **Merchant SA**, Shaikh MJ, Nadkarni PM. Tuberculosis Conundrum - Current and Future Scenarios: A proposed

comprehensive approach combining Laboratory, Imaging and Computing Advances. *World J Radiol* 2022
89 **Wikipedia**. NP-hardness. Wikimedia foundation; 2022. Last Accessed: 8/3/2015. Available from: https://en.wikipedia.org/wiki/NP-hardness
90 **Shor PW**. Algorithms for quantum computation: discrete logarithms and factoring. Proceedings 35th Annual Symposium on Foundations of Computer Science: IEEE Comput. *Soc* 1994; **p**: 124-134 [DOI: 10.1109/sfcs.1994.365700]
91 **Wikipedia**. Optical Computing. 2022. Last Accessed: 4/1/2022. Available from: https://en.wikipedia.org/wiki/Optical_computing
92 **Pires F**. Research Opens the Door to Fully Light-Based Quantum Computing 2020. Last Accessed: 4/1/2022. Available from: https://www.tomshardware.com/news/research-opens-the-door-to-fully-light-based-quantum-computing
93 **Aaronson S**. The Limits of Quantum Computers. Scientific American. 62-69
94 **Sarma SD**. Quantum computing has a hype problem. MIT Technology Review. March 28, 2022. Available from: https://www.technologyreview.com/2022/03/28/1048355/quantum-computing-has-a-hype-problem/
95 **Clarke AC**. Arthur C. Clarke Quotes. 2022. Last Accessed: 4/1/2022. Available from: https://www.brainyquote.com/quotes/arthur_c_clarke_100793

Artificial Intelligence in Medical Imaging

Submit a Manuscript: https://www.f6publishing.com

Artif Intell Med Imaging 2022 June 28; 3(3): 70-86

DOI: 10.35711/aimi.v3.i3.70

ISSN 2644-3260 (online)

MINIREVIEWS

Advances and horizons for artificial intelligence of endoscopic screening and surveillance of gastric and esophageal disease

Byung Soo Yoo, Kevin V Houston, Steve M D'Souza, Alsiddig Elmahdi, Isaac Davis, Ana Vilela, Parth J Parekh, David A Johnson

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0

Grade B (Very good): B

Grade C (Good): 0

Grade D (Fair): D

Grade E (Poor): E

P-Reviewer: Fakhradiyev I, Kazakhstan; Liu Y, China

Received: January 28, 2022

Peer-review started: January 28, 2022

First decision: March 28, 2022

Revised: May 18, 2022

Accepted: June 20, 2022

Article in press: June 20, 2022

Published online: June 28, 2022

Byung Soo Yoo, Steve M D'Souza, Alsiddig Elmahdi, Isaac Davis, Ana Vilela, Department of Internal Medicine, Eastern Virginia Medical School, Norfolk, VA 23507, United States

Kevin V Houston, Department of Internal Medicine, Virginia Commonwealth University, Richmond, VA 23298, United States

Parth J Parekh, David A Johnson, Division of Gastroenterology, Department of Internal Medicine, Eastern Virginia Medical School, Norfolk, VA 23507, United States

Corresponding author: David A Johnson, MD, MACG, FASGE, MACP, Division of Gastroenterology, Department of Internal Medicine, Eastern Virginia Medical School, 886 Kempsville Road Suite 114, Norfolk, VA 23507, United States. dajevms@aol.com

Abstract

The development of artificial intelligence in endoscopic assessment of the gastrointestinal tract has shown progressive enhancement in diagnostic acuity. This review discusses the expanding applications for gastric and esophageal diseases. The gastric section covers the utility of AI in detecting and characterizing gastric polyps and further explores prevention, detection, and classification of gastric cancer. The esophageal discussion highlights applications for use in screening and surveillance in Barrett's esophagus and in high-risk conditions for esophageal squamous cell carcinoma. Additionally, these discussions highlight applications for use in assessing eosinophilic esophagitis and future potential in assessing esophageal microbiome changes.

Key Words: Artificial intelligence; Endoscopy; Gastric cancer; Gastric polyps; Barrett's esophagus; Esophageal cancer

Core Tip: The application of artificial intelligence (AI) in gastroenterology has demonstrated broad utility in esophageal and gastric disease diagnosis and management. The current data shows that AI can be used for gastric polyp and cancer detection and characterization as well as screening and surveillance for esophageal cancer and its high-risk conditions such as Barrett's esophagus. The AI systems can also apply in conditions such as achalasia, post-caustic esophageal injuries, and eosinophilic esophagitis.

Citation: Yoo BS, Houston KV, D'Souza SM, Elmahdi A, Davis I, Vilela A, Parekh PJ, Johnson DA. Advances and horizons for artificial intelligence of endoscopic screening and surveillance of gastric and esophageal disease. *Artif Intell Med Imaging* 2022; 3(3): 70-86

URL: https://www.wjgnet.com/2644-3260/full/v3/i3/70.htm

DOI: https://dx.doi.org/10.35711/aimi.v3.i3.70

INTRODUCTION

Artificial intelligence (AI) has emerged as a new tool with a wide applicability and has transformed every aspect of society including medicine. This technology is an assimilation of human intelligence through computer algorithms to perform specific tasks[1-3]. Machine learning (ML) and deep learning (DL) are techniques of AI. A ML system refers to automatically built mathematical algorithms from data sets that form decisions with or without human supervision[1-3]. A DL system is a subdomain of ML in which AI self-creates algorithms that connects multi-layers of artificial neural networks[1-3].

The recent expansion of research involving AI has shed light on the potential applications in gastrointestinal diseases. Researchers have developed computer aided diagnosis (CAD) systems based on DL to enhance detection and characterization of lesions. CAD systems are now being investigated in numerous studies involving Barrett's esophagus, esophageal cancers, inflammatory bowel disease, and detection and characterization of colonic polyps[4].

In this review, we aim to evaluate the evidence on the role of AI in endoscopic screening and surveillance of gastric and esophageal diseases. In addition, we also provide the current limitations and future directions associated with eosinophilic esophagitis and esophageal microbiome (Figure 1).

MATERIALS AND METHODS

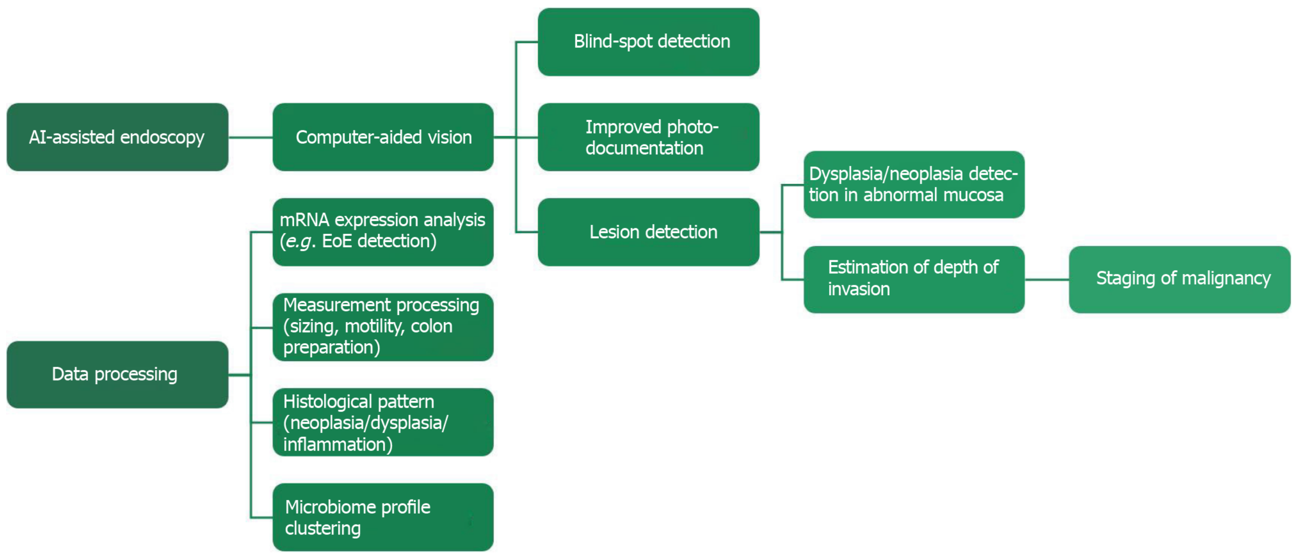
A literature search to identify all relevant articles on the use of AI in endoscopic screening and surveillance of gastric and esophageal diseases was conducted. The search was conducted utilizing PubMed, Medline, and Reference Citation Analysis (RCA) electronic database. We performed a systematic search from January 1998 to January 2022 with search words and key terms including "artificial intelligence", "deep learning", "neural network", "endoscopy", "endoscopic screening", "gastric disease", esophageal disease", "gastric cancer", "gastric polyps", "Barrett's esophagus", "eosinophilic esophagitis", "microbiome".

AI AND GASTRIC POLYPS

Gastric polyps represent abnormal tissue growth, the majority of which do not cause symptoms and, as such, are often found incidentally in patients undergoing upper gastrointestinal endoscopy for an unrelated condition[5]. The incidence of gastric polyps ranges from 1% to 6%, depending on geographical location and predisposing factors, such as *Helicobacter pylori* (*H. pylori*) infection and PPI use[6]. While most polyps are not neoplastic, certain subtypes carry malignant potential with a rater of cancerization as high as 20%[7]. Therefore, the primary utility of polyp detection is cancer prevention. The necessity for detection and recognition of precancerous gastric polyps and the fact that most are incidental findings are a crossroad that has helped propel research and advancement in the field of AI computer-assisted systems for upper-endoscopy.

Detection of gastric polyps

One way to increase accurate detection of gastric polyps is by ensuring complete mapping of the stomach during esophagogastroduodenoscopy (EGD). WISENSE is a real-time quality improvement system that uses deep convolutional neural network (DCNN) and deep reinforcement learning to monitor blind spots, track procedural time and, generate photo documentation during EGD. One of the datasets used to train the network of learning and classifying gastric sites utilized 34513 qualified EGD images. Images were labeled into 26 different sites based on the guidelines of the ESGE and Japanese



DOI: 10.35711/aimi.v3.i3.70

Figure 1 Artificial intelligence -assisted endoscopy and data processing are the currently demonstrated uses for Artificial intelligence. AI: Artificial intelligence; EoE: Eosinophilic esophagitis; mRNA: Messenger ribonucleic acid.

systematic screening protocol. The system was tested using a single-center randomized-control trial. A total of 324 patients were randomized, with 153 of them undergoing EGD with WISENSE assistance. The rate of blind spots (number of unobserved sites in each patient/26) was significantly lower for WISENSE group compared to the control group, 5.86% *vs* 22.46%. Additionally, the system led to increased inspection time and completeness of photodocumentation[8].

A year after the previously mentioned study, the developers renamed WISENSE to ENDOANGEL and further explored the systems capability of identifying blind spots in three different types of EGD; sedated conventional EGD (C-EGD), non-sedated ultrathin transoral endoscopy (U-toe), and non-sedated C-EGD[9]. ENDOANGEL was tested using a prospective single-center, single-blind, randomized, 3-parallel group study. The study results indicated that with the assistance of ENDOANGEL the blind spot rate was significantly reduced for all three EGD modalities. The greatest reduction was seen in the sedated C-EGD group and demonstrated 84.77% reduction. Non-sedated U-TOE and C-EGD blind spot rate decreased by 24.24% and 26.45%, respectively[9]. The major benefit of ENDOANGEL is that it provided real-time prompting when blind spots were identified, thereby allowing the endoscopist to re-examine the missing parts and improve overall visualization. Furthermore, through reduction in total blind spots the authors extrapolate that ENDOANGEL has the potential to mitigate the skill variation between endoscopists[9].

While neither of the above-mentioned systems are specifically designed for the detection of polyps, these encourage and assist endoscopists in completing complete and thorough visualization of stomach during upper endoscopy, a task that has become more daunting over the years as the workload of endoscopists continues to increase. Multiple research groups have created various automated computer-aided vision methods to help detect gastric polyps in real time. Billah *et al*[10] proposed a system that uses multiresolution analysis of color textural features. These color wavelet (CW) features are used in conjunction with CNN features of real time videoframes to train a linear support vector machine (SVM). The fusion of all three features then allows the SVM to differentiate between polyp and non-polyp. The program was trained using more than 100 videos from various sources, resulting in greater than 14000 images being used. This proposed model was then tested on a standard public database and achieved a detection rate of 98.65 %, sensitivity of 98.79%, and specificity 98.52%.

One of the commonly encountered problems with regard to developing computer-aided polyp detection systems is identification of small polyps. To address this problem, Zhang *et al*[11] constructed a CNN using enhanced single shot multibox detector (SSD) architecture that they termed SSD for gastric-polyps (SSD-GPNet). This system was designed to circumvent the problem of lost information that occurs during the process of max-pooling utilized by the SSD feature pyramid during object detection. By reusing this lost information, their new algorithm maximized the quantity of information that could be utilized and therefore increased detection accuracy. The system was tested on 404 images containing gastric polyps, the majority of which were categorized as small. According to the authors, the system was able to achieve real-time gastric polyp detection with a mean average precision of 90.4% utilizing a speed of 50 frames per second[11].

Recently, Cao *et al*[7] developed a system that further improves upon the traditional feature pyramid to identify small polyps as well as those that are more difficult to distinguish from surrounding mucosa due to similarity in features. Their proposed system contains a 'feature fusion and extraction module' which allows the program to combine features from multiple levels of view without diluting the information obtained from adjacent levels. In doing so, program continues to create new feature pyramids which deepens the network, retains more high-level semantic and low-level detailed texture information. The retention and fusion of such information allows the system to distinguish gastric polyps from gastric folds. The system was trained using 1941 images with polyps. To overcome the small data set, the authors utilized random data augmentation which consists of changing image hue and saturation, rotation of the image, *etc.* The system demonstrated a mean sensitivity of 91.6% and recall of 86.2% (proportion correctly identified true positives), after 10-fold validation testing[7]. Unfortunately, the authors do not provide detection results regarding those polyps they deemed difficult to discern from gastric folds. Nonetheless, the development of an augmented data set and a high level of sensitivity show promise with regards to overall polyp detection rates.

Characterization of gastric polyps

There are numerous types of gastric polyps and most of them do not carry any malignant potential. The two classes of polyps with the highest potential for malignancy are hyperplastic polyps and gastric adenomas. Gastric adenomas, or raised intraepithelial neoplasia, represent direct precursor lesions to adenocarcinoma and rarely appear in the presence of normal gastric mucosa. Instead, they are often found on a background of chronic mucosal injury, such as chronic gastritis and gastric atrophy[6]. Therefore, many of the AI systems that have been developed to assist endoscopists in the prevention of gastric cancer focus on the characterization and identification of known gastric cancer precursor lesions such as gastric atrophy and intestinal metaplasia, rather than characterizing all the various types of polyps. Characterization of gastric polyps relies heavily on image-enhanced endoscopy (IEE). Especially modalities such as narrow-band imaging (NBI) and blue laser imaging with or without magnification.

Xu *et al*[12] utilized various IEE images to train their DCNN system, named ENDOANGEL, to detect and diagnose gastric precancerous conditions, specifically gastric atrophy and intestinal metaplasia, in real time. The performance of their AI model tested using a prospective video set achieved an accuracy of 87.8%, sensitivity of 96.7% and specificity of 73.0% with regards to identification of gastric atrophy. In the prospective video set test for intestinal metaplasia the system achieved an accuracy, sensitivity, and specificity of 89.8%, 94.6%, and 83.7%, respectively[12]. Additionally, the system performance was tested against that of endoscopist with varying degrees of expertise (for a subset 24 patients). Overall, the program performed similarly to 4 expert endoscopists (those with 5 or more years of training including 3 or more in IEE). Compared to 5 nonexpert endoscopists (those with 2 years of endoscopic experience and 1 year of experience in IEE) who had a mean accuracy of 75.0%, sensitivity of 82.8% and specificity of 59.4% for GA and an accuracy of 73.6%, sensitivity of 73.8%, and specificity of 73.3% for IM, ENDOANGEL performed significantly better[12].

Limitations of AI in gastric polyps

To the best of our knowledge, there have been no randomized control trials to evaluate the clinical efficacy of AI automated gastric polyp detection systems. However, the accuracy, sensitivity, and specificity of those mentioned here, as well as others not mentioned, indicate great potential in assisting endoscopist to detect gastric polyps. With the further development of AI systems to not only detect but, to characterize these gastric lesions, the potential clinical utility is further increased. AI systems with fully developed CADe and CADx can be developed to aid rapid and effective decision making for identifying lesions that should be targeted for biopsy. Such systems may also improve other patient outcomes by mitigating the difference in endoscopist experience.

AI AND GASTRIC CANCER

Gastric cancer (GC) is the fifth most common cancer in the world and the fourth most fatal cancer[11]. The 5-year survival rate is greater than 90% when diagnosed at early stages, making early detection particularly important[7]. Alarmingly, in 2019, more than 80% of GCs in China were diagnosed at advanced stages, signifying inadequate early detection[12]. Risk factors for GC include *H. pylori* infection, alcohol use, smoking, diet, race and gender[13]. Due to the non-specific nature of symptoms, most GC is usually diagnosed at later stages which makes prognosis poor[14].

Although endoscopic imaging is the most effective method of detection, visualization can be difficult. The reasons for this include the subtle changes in mucosa (elevations, depressions, redness or atrophy) that can be mistaken for gastritis or intestinal metaplasia, especially when found in a region with background gastritis[15]. Further, the subjective nature of identification makes detection endoscopist dependent with reported miss rates as high as 14% and 26%[15,16]. In addition to the limitations in detecting mucosal changes, endoscopy is historically poor at predicting depth of invasion with studies reporting only 69% to 79% accuracy[17]. This is important because accurately predicting depth of invasion can aid in guiding management and surgical planning.

Over the past several decades, AI has expanded towards new horizons in medicine and image recognition. Recently, DL has become more widely applied in the prevention and detection of GC. Medical image recognition in locating tumors is called "image segmentation". Importantly, image segmentation determines diagnostic accuracy for evaluation and surgical planning in GC. DL has been shown to improve image segmentation *via* three networks; supervised network, semi-supervised network, and unsupervised network[18]. Supervised learning networks comprise the majority. These networks use large data sets that are preemptively labeled. Convolutional Neural Networks (CNN) are supervised learning networks which have demonstrated high performance in image recognition tasks [18].

Prevention, detection and classification of gastric cancer

For prevention of GC, it is important to optimize the diagnosis and eradication of *H. pylori*. In 2018, Itoh *et al*[19] developed a CNN-based system which was trained on 149 images to diagnose *H. pylori*. The results showed 86.7% sensitivity and 86.7% specificity which significantly outcompetes traditional endoscopy and the researchers concluded that CNN-aided endoscopy may improve diagnostic yield in H. pylori endoscopy.

A 2020 systematic review and meta-analysis. reviewed 8 studies with 1719 patients and found a pooled sensitivity and specificity of 0.87 (95%CI 0.72-0.94) and 0.86 (95%CI 0.77-0.92), respectively in predicting *H. pylori* infection. In addition, the study showed an 82% accuracy of AI for differentiating between post eradication images and non-infected images[20]. The authors were also able to identify 2 studies where discrimination using AI, between *H. pylori* infected and post-eradicated images was analyzed, revealing an accuracy of 77%. While the authors state external validity as a limitation of this study, the results cannot be ignored in the context of prior studies. Accordingly, AI may have a role in diagnosis as well as confirmation of treatment.

Along with eradication of *H. pylori*, prevention also comes in the form of detecting precancerous lesions. These lesions include erosion, polyps and ulcers which may develop into gastric cancer if they are not detected early. In 2017, Zhang *et al*[21] developed a CNN known as the Gastric Precancerous Disease Network (GPDNET) to categorize precancerous gastric disease. This AI demonstrated an accuracy of 88.90% in classifying lesions as either polyps, erosions or ulcers.

As previously mentioned, GC is often discovered in late stages, which thereby makes improvements in early detection, particularly important. Deep learning algorithms have shown promise with this regard. A study by Li *et al*[22] demonstrated significantly higher diagnostic accuracy in CNN trained (90.91%) endoscopy compared to non-experts (69.79 and 73.61%) (P < 0.001 with kappa scores of 0.466 and 0.331). The researchers looked at CNN-based analysis of gastric lesions observed by magnifying endoscopy with narrow band imaging (M-NBI) and found a 91.8% sensitivity, 90.64 specificity and 90.91 accuracy in diagnosing early gastric cancer (EGC). While specificity was like that of experts, sensitivity of EGC detection was superior to both experts (78.24 and 81.18) and non-experts (77.65 and 74.12). The researchers attributed this to a lack of subjectivity which is inherent to human endoscopy. Ikenoyama *et al*[23] constructed their CNN using 13584 images from 2639 early GC lesions and compared its diagnostic ability to 67 endoscopists. Results showed faster processing as well as a 26.5% higher diagnostic sensitivity in CNN compared to endoscopists. This further demonstrates the potential for AI to improve efficiency in diagnosing GC.

The role of AI is not limited to early detection. Hirasawa *et al*[24] constructed a CNN trained with 13584 images to detect both early (T1) and advanced GC (T2-4). They demonstrated an overall sensitivity of 92.2% in diagnosing gastric cancer. The diagnostic yield was further accentuated at diameters of 6mm or greater with a sensitivity of 98.6%. All invasive lesions were correctly identified as cancer during this study. Despite these promising results, there were false positives that lead to a positive predictive value (PPV) of only 30.6%.

In addition to CNN, fully convolutional neural networks (FCN) use pixel level classification to allow for more robust image segmentation[25]. When it comes to distinguishing cancer from precancerous disease, FCN has shown promise. In 2019, Lee *et al*[26] used data from 200 normal, 220 ulcer and 367 cancer cases to build the Inception-ResNet-v2 FCN which was able to distinguish between cancer and normal as well as cancer and ulcer at accuracies above 90%. In a 2019 study by Nguyen *et al* Inception-ResNet-v2 was used to further classify neoplasms based on severity. Five categories were assessed: EGC, advanced GC, high grade dysplasia, low grade dysplasia and non-neoplasm. The result was a weighted average accuracy of 84.6% in classifying neoplasm[27].

Depth of invasion of gastric cancer

Depth of invasion is an important characteristic when it comes to accordant direction for best management of GC[17]. The current evidence suggests that early stages of EGCs with depth limited to the mucosal (M) or superficial submucosal layers (SM1) can be managed with endoscopic submucosal dissection or endoscopic mucosal resection[17]. Invasion into the deeper submucosal layer will require surgery. In 2018, Zhu *et al*[17] built a CNN computer-aided detection (CNN-CAD) system to determine depth of invasion of GC. The results showed accuracy of 89.16% which was significantly higher than that of endoscopists (69% to 79%). PPV and NPV were 89.66% and 88.97%, respectively. Endoscopists had values of 55.86% and 91.01%. This enhanced ability to predict invasion supports the assertion that

CNN has shown utility in helping endoscopists detect, classify, and predict prognosis of GC.

Limitations of AI in gastric cancer

Supervised learning networks show promise in the prevention of cancer through detection of *H. pylori* and precancerous lesions as well as promise in detection and classification of neoplasm. AI has not only demonstrated superiority to traditional endoscopists when it comes to identifying GC stage but also at determining depth of invasion which can dramatically improve prognosis in a disease with inadequacy of early detection. There is utility when it comes to helping less experienced endoscopists. Despite their superior diagnostic efficacy, supervised learning networks are not immune to false positives and false negatives. Because they rely heavily on the quality and quantity of learning samples, they may interpret poor images of intestinal metaplasia or atrophy as GC and are data dependent[25]. Semi-supervised and unsupervised learning networks are potential alternatives as they are not entirely data dependent[18].

AI AND BARRETT'S ESOPHAGUS

The American Cancer Society's estimates about 19260 new cases of esophageal cancer (EC) diagnosed (15310 in men and 3950 in women) and about 15530 deaths from EC (12410 in men and 3120 in women) in the United States in 2021[28]. It is the seventh most common cancer and the sixth leading cause of cancer related mortality worldwide[29]. The two major histological types of EC are adenocarcinoma (AC) and squamous cell carcinoma (SCC)[30]. For SCC alone, the primary causal risk factors vary geographically. Over the past 40 years, the incidence of AC, which typically arises in the lower third of the esophagus, has risen faster than any other cancer in the Western world, and rates continue to rise even among new birth cohorts. Conversely, the incidence of SCC has declined in these same populations. As such, AC is now the predominant subtype of esophageal cancer in North America, Australia and Europe. Like AC, the incidence of Barrett esophagus has increased in many Western populations[31].

Barret's esophagus (BE) is a change of the normal squamous epithelium of the distal esophagus to a columnar-lined intestinal metaplasia, and the main risk factors associated with its the development are long-standing gastroesophageal reflux disease (GERD), male gender, central obesity, and age over 50 years[32]. It is thought to follow a linear progression from nondysplastic BE to low-grade dysplasia to high-grade dysplasia and finally to cancer. The presence of regions of dysplasia in BE increases the risk of progression and guides treatment considerations. Early detection of dysplastic lesions and cancer confined to the mucosa allows for minimally invasive curative endoscopic treatment, which provides a less invasive method of treatment than surgical resection and/or neo adjuvant therapy for advanced lesions. However, the evaluation and assessment of BE is challenging for both expert and nonexpert endoscopists. The appearance of dysplasia may be subtle, and segmental biopsy samples may not detect patchy dysplasia[33,34].

Current challenges in Barrett's esophagus

Results from a multicentric cohort study support that missed esophageal cancer is relatively frequent at routine upper gastrointestinal endoscopies in tertiary referral centers, with an overall MEC rate as high as 6.4% among newly diagnosed esophageal cancer patients[35]. Additionally, a recent meta-analysis showed a high miss rate of 25% for high grade dysplasia and cancer within 1 year of a negative index examination, the reasons for this are likely multifactorial, including the lack of recognition of subtle lesions, lack of detailed inspection of the esophageal mucosa, non-optimum cleaning techniques, and less experienced endoscopists[34].

Optical identification and diagnosis of dysplasia would guide treatment decisions during endoscopy for BE. The limitations of current screening and surveillance strategies impulse to improve diagnostic accuracy and risk stratification of patients with BE. In recent years, many new endoscopic techniques have been developed, such as magnification endoscopy, chromoendoscopy, confocal laser endomicroscopy, and volumetric laser endomicroscopy, most of which are expensive and take a long time for endoscopists to learn. Differences in endoscopists' interpretations of the images can also lead to differences in diagnosis[36].

AI and convolutional neural network

A proposed use of AI during upper endoscopy will be with live video images that will be sent to the AI application and analyzed in real time. The application will be able to detect areas suspicious for neoplasia and measure the size and morphology of lesions. It will alert the endoscopist to suspicious areas either with a screen alert or location box. The endoscopist can then decide if the area needs to be sampled based on the characterization provided by the machine or managed endoscopically[34]. Therefore, AI can assist in by using methods of DL to identify and process in real-time endoscopic data that may not consciously appreciated by humans such as subtle changes in color and texture to aid in taking targeted biopsies rather than random biopsies.

AI uses several machine learning methods, one that is frequently used is CNN, a form of DL which receives input (*e.g.* endoscopic images), learns specific features (*e.g.* pit pattern), and processes this information through multilayered neural networks to produce an output (*e.g.* presence or absence of neoplasia). Several layers of neurons can exist to make a single decision to call a grouping of pixels on an image either normal tissue or dysplasia. The advantages that AI appears to confer per-endoscopy is a removal of the inter-observer or intra-observer variability in identification of non-normal lesions, combined with rapid, objective analysis of all visual inputs in such a way that is consistent and not subject to fatigue. This advanced technology of CAD can allow endoscopists to take targeted, high-yield biopsies in real-time. Compared to taking random biopsies per the Seattle protocol or using enhanced imaging, CAD may increase efficiency and accuracy for making a diagnosis by limiting the chance of missing neoplastic mucosa. Moreover, CAD may decrease risk by decreasing sedation time secondary to decreased procedure length[37].

AI use with white light imaging

Van der Sommen *et al*[38] in 2016 collected 100 images from 44 BE patients and created a machine learning algorithm which used texture and color filters to detect early neoplasia in BE. The sensitivity and specificity of the system were 83% for the per-image analysis and 86% and 87% for the per-patient analysis, respectively. Therefore, the automated computer algorithm developed was able to identify early neoplastic lesions with reasonable accuracy, suggesting that automated detection of early neoplasia in Barrett's esophagus is feasible.

In a study by de Groof *et al*[39], six experts identified likely neoplastic tissue in the same image and used these expert-delineated images to train the computer algorithm to identify neoplastic BE and non-dysplastic BE in test cases. The resulting sensitivity and specificity of the computer algorithm was 0.95 and 0.85 respectively. de Groof *et al*[40] developed a deep learning system using high-definition white light endoscopy images of over 10000 images of normal GI tract followed by 690 images of early neoplastic lesions and 557 non dysplastic Barrett's epithelium to detect, delineate the lesion, and pinpoint high yielding biopsy sites withing the lesion. This group was able to externally validate their CAD system demonstrating a better accuracy of 88% in detecting early neoplastic lesions compared with an accuracy of 73% with endoscopists. Ebigbo *et al*[41] were also able to validate a CNN system to detect EAC in real time with the endoscopic examination of 14 patients using 62 images and showed a sensitivity of 83.7% and specificity of 100%.

Hashimoto *et al*[42] collected 916 images from 70 patients with early neoplastic BE and 916 control images from 30 normal BE patients and then trained a CNN algorithm on ImageNet. The researchers analyzed 458 images using the CNN algorithm. The accuracy, sensitivity, and specificity of the system for detecting early neoplastic BE were 95.4%, 96.4%, and 94.2%, respectively.

AI use with volumetric laser endomicroscopy and confocal laser endomicroscopy

The volumetric laser endomicroscopy system has the capacity to provide three-dimensional circumferential data of the entire distal esophagus up to 3-mm tissue depth. This large volume of data in real-time remains difficult for most experts to analyze. AI has the potential to better interpret such complex data [43].

Interpretation of volumetric laser endomicroscopy (VLE) images from BE patients can be quite difficult and requires a steep learning curve. An AI software called intelligent real-time image segmentation has been developed to identify VLE features by different color schemes. A pink color scheme indicates a hyper-reflective surface which implies increased cellular crowding, increased maturation, and a greater nuclear to cytoplasmic ratio. A blue color scheme indicates a hypo-reflective surface which implies abnormal BE epithelial gland morphology. An orange color scheme indicates lack of layered architecture which differentiates squamous epithelium from BE[44].

Swager *et al*[45], created an algorithm to retrospectively identify early BE neoplasia on *ex vivo* VLE images showing a sensitivity of 90% and specificity of 93% in detection with better performance than the clinical VLE prediction score. A CAD system reported by Struyvenberg *et al*[46] analyzed multiple neighboring VLE frames and showed improved neoplasia detection in BE with an area under the curve of 0.91.

Future of AI and applications in Barrett's esophagus

Ali *et al*[47] at the University of Oxford reported on one a deep learning tool to automatically estimate the Prague classification and total area affected by columnar metaplasia in patients with Barrett's esophagus. They propose a novel methodology for measuring the risk score automatically, enabling the quantification of the area of Barrett's epithelium and islands, as well as a 3-dimensional (3D) reconstruction of the esophageal surface, enabling interactive 3D visualization. This pilot study used a depth estimator network is used to predict endoscope camera distance from the gastric folds. By segmenting the area of Barrett's epithelium and gastroesophageal junction and projecting them to the estimated mm distances, they were able to measure C&M scores including the area of Barrett's epithelium. The derived endoscopy artificial intelligence system was tested on a purpose-built 3D printed esophagus phantom with varying areas of Barrett's epithelium and on 194 high-definition videos from 131 patients with C&M values scored by expert endoscopists. The endoscopic phantom video data demonstrated a 97.2% for C&M and island measurements, while the accuracy for the area of Barrett's epithelium it was 98.4% compared with ground-truth[47].

This is the first study to demonstrate that Barrett's circumferential and maximal lengths and total affected area can be automatically quantified. While further optimization and extensive validation are required, this tool may be an important component of deep learning-based computer-aided detection systems to improve the effectiveness of surveillance programs for Barrett's esophagus patients[48].

The studies show promising results and as AI systems develop, it will be important that they are tested and validated in real-world settings, in diverse patient populations, with physicians of varying expertise, with different endoscope types and in different practice settings. Commercially developed AI will need to demonstrate cost-effective care that will provide meaningful value and impact on patient care and outcomes. The field continues to expand and promises to impact the field of BE detection, diagnosis, and endoscopic treatment[33,49].

ACHALASIA AND AI

Achalasia is an esophageal motility disorder characterized by impaired peristalsis and relaxation of the lower esophageal sphincter. While the pathophysiology is incompletely understood, it is thought to be related to loss of inhibitory neurons in the myenteric plexus. Symptoms include dysphagia to both solids and liquids as well as heartburn, chest pain and other nonspecific symptoms. In fact, 27%-42% of patients are initially misdiagnosed as GERD[50].

High-resolution manometry (HRM) is the gold standard[51]. A limitation of manometry is that it cannot differentiate between achalasia and pseudo achalasia, a disorder which is often malignancy presenting as achalasia[52]. As such, the utility of endoscopy comes in ruling out malignancy and endoscopic biopsy is an important part of the diagnostic algorithm. Endoscopy can also be used to rule out other obstructive lesions or GERD[53]. However, HRM is vital in classification of achalasia subtypes which guides treatment and prognosis.

The Chicago Classification system is based on manometric differences between three subtypes. All three have impaired EGJ relaxation[54]. Subtype 1 has aperistalsis with the absence of pan esophageal pressurization. Subtype 2 has aperistalsis with pressurization greater than 30 mmHg and subtype three is characterized by abnormal spastic contractions with or without periods of pan esophageal pressurization. While types 1 and 2 can be corrected with Heller myotomy, type 3 patients are more likely to benefit from more extensive myotomy[55].

Functional lumen imaging probe and AI

The functional lumen imaging probe (FLIP) device that uses high resolution impedance planimetry to measure cross sectional area and pressure to provide a 3D model of achalasia. It has been shown to be just as good as manometry in diagnosing achalasia and has also shown application in cases where clinical suspicion is high, but manometry is equivocal[56]. Because FLIP is performed during endoscopy, it can help identify patients who do not respond to manometry.

Despite its ability to diagnose achalasia, FLIP has limited data available in its ability to differentiate between achalasia subtypes. If it were able to do this, it could essentially combine the steps of endoscopic evaluation, diagnosis, and classification of achalasia. Machine learning may have a role here.

In 2020, Carlson *et al*[57] were able to demonstrate the application of supervised machine learning in using FLIP to characterize achalasia subtypes in a study of 180 patients. The AI was able to differentiate type 3 achalasia from non-spastic subtypes with an accuracy of 90% while the control group did so with an accuracy of 78%. The machine was also able to further classify achalasia into subtype 1, 2 and 3 with an accuracy of 71% compared to the 55% accuracy of the control group. This is an important application given the differences in prognosis and management based on subtype.

Achalasia and cancer

Esophageal cancer is a rare consequence of achalasia with reported risks ranging from 0.4%-9.2%[58]. One meta-analysis found a risk of SCC of 308.1 per 1000000 per year[59]. One study found that 8.4% of 331 patients with achalasia developed Barrett's esophagus after undergoing pneumatic dilation[60]. While there are no established guidelines for cancer screening in patients with achalasia, some studies have suggested 3-year interval screening for patients with achalasia for 10 or more years[58].

Given the association between achalasia and esophageal cancer, enhanced imaging in high-risk patients should have value and applications of AI in this population are warranted.

POST CAUSTIC INGESTION AND AI

In the United States, there were over 17000 cases of caustic injury which accounted for about 9% of poisoning cases[61]. Endoscopy has been determined to be an important part of diagnosis and prognosis for these cases of post-caustic ingestion[62,63]. Typically, the Zargar classification is used to help guide evaluation with patients graded 0 through IV. Those with grade III or above typically had complications or death[64]. Artificial intelligence in endoscopy and the role for post-caustic ingestion has not been evaluated. It is reasonable to postulate that with advances in other areas of upper endoscopy in evaluation of the GI lumen for precancerous lesions, achalasia, esophageal carcinoma that there is a role for evaluation of the GI lumen for grading of caustic injury. Further studies are necessary to evaluate whether there is a role for AI assistance in evaluation and if there would be a significant difference in patient outcomes after implementation.

AI AND ESOPHAGEAL SQUAMOUS CELL CARCINOMA

Esophageal cancer has been a large area of investigation due the aggressive disease course and high morbidity and mortality outcomes. It has been reported to be as high as the eighth most common cancer and sixth leading cause of cancer-related death world-wide[65]. As of 2020, there are higher risk geographic areas of concern regarding esophageal cancer in South-Central Asia being the third overall leading cause of cancer-related mortality in males and in the region of Eastern and Southern Africa esophageal cancer ranks second and third in male cancer-mortality respectively. Eastern Africa is also the third leading cause of female related cancer incidence and mortality[66].

Of the two major subtypes of esophageal cancer esophageal squamous cell carcinoma (ESCC) is the predominant histological type world-wide[67]. Classically, ESCC has been associated with risk factors including gender, race, tobacco and alcohol consumption, diet and nutrient intake[67]. Recently, poor oral health and microbiome changes have been associated with the development or predisposition of ESCC[68,69]. By the time of diagnosis of ESCC, disease course is typically found at an advanced stage and often requires highly invasive treatment contributing to poor prognosis, morbidity, and mortality rates. Investigation into early screening is critical, but as with implementation of any mass screening, the method must be evaluated for the benefit of screening tests to reduce cancer *vs* the risk of over-diagnosing and putting patients through high-risk procedures. It should be noted that there may be specific benefits in implementation of screening in high-risk populations and geographic areas in areas of Africa and Asia. Being an area with high rates of esophageal and gastric cancer, a research study across seven cities in the Henan Province of China enrolled 36154 people for screening using endoscopy and biopsy[70]. They found 46% of patients had precancerous lesions, 2.42% had confirmed cancer. Of those with this confirmed cancer diagnosis, 84% of them had an early stage that underwent prompt treatment with a success rate of 81%. Their study concluded that early detection was crucial in reducing their rate of esophageal and gastric carcinoma in that region[70].

Early-stage detection of ESCC

Early detection is important for improving outcomes for ESCC. Historically, conventional white light endoscopy with biopsy was the gold standard for diagnosis of esophageal cancer[71]. The limitation of this for ESCC is that clinical suspicion needs to be high to perform the procedure and the cancer must be of significant size to be identified on endoscopy. The emergence of chromoendoscopy, using chemicals such as iodine, allowed a staining technique to better detect ESCC. But this procedure can often cause irritation in patients due to mucosal irritation to the GI tract and it increases procedural time per patient.

Alternatively, the emergence of narrow band imaging offers an image-enhancing technique using wavelength filters to observe mucosal differences and vascular patterns on the GI tract that correlates with esophageal cancer (among other uses stated throughout this article). The downside of NBI is that detection rate is dependent on endoscopist experience and subject-ability in processing the information given[71]. Despite these methods, a large multi-center retrospective cohort study by Rodríguez de Santiago *et al*[35] analyzed over 123000 patients undergoing EGD and found a miss rate of esophageal cancer of 6.4% with a follow-up diagnosis made within 36 mo by repeat endoscopy. This miss rate was present regardless of histologic subtype of esophageal adenocarcinoma or ESCC. Their analysis found that less experienced endoscopists and smaller lesions were associated with the missed detection. Their study acknowledges that there was a low use of chromoendoscopy due to small proportion of early neoplasms across the study and a lack of digital chromoendoscopy at their institutions at the time of the study which may limit applicability[35]. But this still suggests conventional techniques have higher miss rates and newer technology or innovative technique development are essential in assisting and creating a better standard for ESCC detection and to provide a basis for better screening in this aggressive disease.

AI systems – early detection, screening, surveillance

The use of endoscopic AI has recently showed potential to change the diagnostic evaluation for many different gastrointestinal tract diseases. Due to the novelty, ESCC guidelines for use of AI in clinical practice is still being determined.

The use of AI specifically in high-risk populations, may provide great utility to reduce rates of ESCC. Early detection through AI has shown promise through early studies. Ohmori *et al*[72] used a CNN and showed an accuracy of the AI system for diagnosing ESCC was comparable to that of experienced endoscopists. The system achieved a 76% PPV for detection using non-magnified images and in the differentiation of ESCC using magnified images. Horie *et al*[73], one of the pioneer investigators of AI in GI endoscopy used a CNN-based AI system to detect ESCC. Their study results showed that their CNN took only 27 s to analyze 1118 images and correctly detected esophageal cancer cases with 98% sensitivity[73]. Thus, it is reasonable that beyond the use of AI systems for evaluation for high-risk patients, at a population-based level, AI systems could be utilized to analyze endoscopic images of patients of medium to low risk that are undergoing EGD for other reasons.

A study by Cai *et al*[74] specifically developed and validated a computer-aided detection using a DNN to be used for screening for early ESCC. Out of 1332 abnormal and 1096 normal images from 746 patients, they compared their system to 16 endoscopists of various experience levels. Their results showed that the DNN-CAD had an accuracy of 91% compared to their senior endoscopist of 88% and junior endoscopists of 77%. More importantly, after taking the results separately, they allowed the endoscopists to refer to the data and this improved the average diagnostic ability of the endoscopists from an overall average accuracy from 81 to 91%, sensitivity from 74 to 89%, and NPV from 79 to 90% [74].

Depth of invasion

Beyond identifying ESCC at a superficial level for diagnosis, the ability to accurately assess the depth of invasion is important, because it best guides intradisciplinary treatment options[75]. Criteria for diagnosis can be divided into two broad categories: non-magnified endoscopy and magnified endoscopy[75]. In non-magnified endoscopy, macroscopic identifiers are observed such as protrusions and depressions. Magnified endoscopy observes the blood vessel patterns using narrow-based imaging or blue laser imaging; criteria of invasion up to 200 μm (SM1) are candidates for resection because of their lower risk of metastasis[75]. Alternatively, SM2-3 are considered higher risk of metastasis and require consideration for esophagectomy[75]. This diagnostic identification is shown to have endoscopist variability.

The AI systems using CNN have recently emerged to assist the endoscopist and create a higher standard for depth of invasion detection to match or have higher rates than those of expert endoscopists. Evidence was shown by Tokai *et al*[76], where they used a CNN to differentiate between SM1 and SM2. This was a retrospective study, and 1791 test images were prepared and reviewed by the CNN compared with review by 13 expert endoscopists and found that the AI system demonstrated higher diagnostic accuracy for invasion depth than those of endoscopists.

To determine clinical application from still-images to video, a more recent study by Shimamoto *et al* [77] utilized real-time assessment of video images for ESCC and compared their AI model with those of expert endoscopists and found that accuracy, sensitivity, and specificity with non-magnified endoscopy were 87%, 50%, and 99% for the AI system and 85%, 45%, 97% for the experts. Accuracy, sensitivity, and specificity with magnified endoscopy was 89%, 71%, and 95% for the AI system and 84%, 42%, 97% for the experts. This suggests that with more inexperienced endoscopists, AI can offer a similar or even higher standard and allow for better patient outcomes with higher depth of invasion diagnosis.

Newer advances in the field of endoscopic AI may offer the potential for diagnosis without biopsy. The Japan esophageal society introduced a classification system for endoscopic diagnosis of ESCC by analyzing intrapapillary capillary loops which help estimate depth of invasion and make a visual diagnosis for ESCC. Although this classification can be endoscopist-dependent, in combination with AI systems, study by Zhao *et al*[78] used a computer assisted model to allow objective image evaluation and assist in classification of EPCLs and found that their model was 89% accurate in diagnosing the lesion. This was in comparison to accuracy of 92% by senior endoscopists (greater than 15 years), 82% by mid-level endoscopists (10-15 years), and 73% by junior endoscopists (5-10 years). While it is likely not to replace histopathological confirmation, being able to diagnose at a high rate could help more efficiently allocate resources and provide faster diagnosis to help guide clinical intervention in this highly aggressive disease.

In summary, implementation of any cancer-screening for primary prevention is going to require careful analysis of risk-benefits through large-scale medical studies. It is clear that ESCC has a significant presence world-wide and of particular healthcare burden in geographic areas of Africa and Asia. ESCC studies have suggested that implementation of screening can benefit high-risk populations in these areas. AI in endoscopy has emerged with promise in showing consistent results in both early detection, quicker diagnosis, and non-inferior rates of success for the studied patients. Implementation of AI with endoscopic screening of high-risk populations for ESCC should be considered in the coming years as the technology becomes more widely available.

FUTURE PERSPECTIVES FOR AI AND ESOPHAGEAL DISEASES AND MICROBIOME

Eosinophilic esophagitis (EoE)

Eosinophilic esophagitis is a food allergen-mediated inflammatory disease affecting the esophagus. It is traditionally associated with atopic conditions such as asthma and atopic dermatitis[79]. Treatment includes food-elimination diets, proton-pump inhibitors, and topical steroids[79].

Initial diagnosis of eosinophilic esophagitis (EoE) involves mucosal biopsy demonstrating > 15 eosinophils per high-powered field (400× magnification)[79]. In addition to this peripheral eosinophil count (PEC), other histological features may be present in EoE, and can be used to characterize the disease state and to assess for response to therapy, including epithelial thickness, eosinophilic abscess, surface layering, and epithelial alteration[80]. These features have been used to develop a histologic scoring system for diagnosis, the EoEHSS[80]. Both PEC and EoEHSS are evaluated by a pathologist, and are time-consuming processes. EoEHSS additionally requires training and there appears to be inter-observer variability. The need for a more precise and automated process has let to machine learning approaches. Several groups have developed platforms for automated analysis of biopsy images that utilized a deep-convolutional neural network approach to distinguish downscaled biopsy images for features of EoE[81,82]. One platform was able to distinguish between normal tissue, candidiasis, and EoE with 87% sensitivity and 94% specificity. Another platform was able to achieve 82.5% sensitivity and 87% specificity in distinguishing between EoE and controls, despite the potential limitations of image downscaling[82].

In addition to improving efficiency and precision of current diagnostic methods for EoE, AI is a promising tool for the development of new diagnostic methods to subclassify disease and guide treatment. One approach is through evaluation of tissue mRNA expression for unique factors that can classify or subclassify EoE. One group used mRNA transcript patterns to develop a probability score for EoE, in comparison to GERD and controls[83]. This diagnostic model was found to have a 91% diagnostic sensitivity and 93% specificity[83]. Additionally, this EoE predictive score was able to demonstrate response to steroid treatment[83]. Further work may develop new diagnostic criteria, methods for subclassification of disease, and to assess for various therapeutic options.

Esophageal microbiome

Current understanding of the commensal microbiome has developed through various techniques, including 16s rRNA sequencing to describe genus-level composition or shotgun sequencing to describe strain-level composition of a sample microbial community[84]. Various ML models, specifically DL, have been utilized to develop descriptive techniques, disease prediction models based on composition and for exploration of novel therapeutic targets[85].

Initial work on the esophageal microbiome described two compositional types: Type I, associated with the healthy population, mainly consisting of gram-positive flora, including *Streptococcus spp*., and a Type II, associated with GERD and BE, with higher prevalence of gram-negative anaerobes[86]. Later work stratified esophageal microbiome communities into three types, a *Streptococcus spp*. predominant (Cluster 2), *Prevotella spp*. predominant (Cluster 3), and an intermediate abundance type (Cluster 1)[87]. Further work has identified specific flora or groups of flora associated with various disease states as well as a gradient of composition from proximal to distal esophagus[69].

The ML models can be used to expand on this work using both supervised and unsupervised methods. Random Forest classifiers and Least Absolute Shrinkage and Selection Operator feature selection have been used to analyze shotgun genomics data and classify disease state and stage several GI disorders, including colorectal cancer and Crohn's disease[87-90]. In addition to descriptive methods, machine learning has been used to develop models to predict disease progression in primary sclerosing cholangitis[91]. Finally, correlation-based network analysis methods have been used to assess response to intervention, such as symptomatic response to probiotics and association with microbial changes[92]. Within esophageal disease, a neural network framework has been used to develop a microbiome profile for classification of phenotypes, including datasets from patients with BE and EAC[93]. Future work has the potential to further develop microbiome-based models for detection, assessment of progression, and development of new therapeutics for several esophageal disease states.

DISCUSSION

The emerging use of AI in medicine has the potential for practice changing effects. During the diagnostic process, better visualization techniques, including CAD can assist endoscopists in detection of lesions[94]. When malignancy is detected, AI can be used to predict extent of disease[94]. Following diagnosis, CNN can be used to predict response to treatment as well as risk of recurrence[94].

Of the multiple AI techniques with demonstrated use, some are more likely to be more adaptable to everyday use by clinicians. AI-assisted endoscopy is already being utilized in the area of colorectal disease, with products available on the market to assist with adenoma detection rate and early detection [95]. Given the compatibility of AI solutions with current endoscopic devices, it is likely that broader applications of these systems to other areas of the GI tract are approaching[96].

Some limitations exist in the use of AI-based techniques. First, the quality and number of learning samples significantly affects the accuracy of predictive algorithms. This primarily affects supervised learning networks, where the use of labeled sample data affects the quality of training, and can affect overall accuracy. This concept is sometimes referred to as "garbage in, garbage out." For example, in the detection of gastric cancer, supervised learning algorithms that rely heavily on the quality and quantity of samples may interpret poor images of intestinal metaplasia or atrophy as GC and are heavily data dependent[24]. Semi-supervised and unsupervised learning networks are potential alternatives as they are not entirely data dependent[19]. Another possible limitation is the role of confounding factors- lack of population diversity in training models may lead to lack of generalizability of AI systems to alternate populations.

Finally, privacy will be important to maintain when translated to clinical practice, in both the improvement of training models as well as in patient care. Further legislative discussion is needed to ensure adequate privacy when patient medical data is used and potentially shared for use in ongoing training of AI models[97]. Additionally, this further digitization and storage of patient data will require appropriate security within adapting healthcare system infrastructures[97,98].

CONCLUSION

Clearly, the rapidly developing application of artificial intelligence has shown its wide applicability in gastroenterology and continues to be investigated for the accuracy in endoscopic diagnosis of esophageal and gastric diseases. The esophagogastric diseases including gastric polyps, gastric cancer, BE, achalasia, post-caustic ingestion, ESCC, eosinophilic esophagitis have distinct features that AI can be utilized. The current systems propose a sound base for an AI system that envelops all the esophago-gastric diseases. Although this area of active research is very encouraging, further work is needed to better define the specific needs in assessing disease states as well as the cost effectiveness before incorporating AI as a standard tool for daily practice.

FOOTNOTES

Author contributions: Johnson DA, Parekh PJ, and Yoo BS all contributed to the construction of the project; all authors wrote and edited the manuscript.

Conflict-of-interest statement: All the authors declare that they have no conflict of interest.

Country/Territory of origin: United States

ORCID number: Byung Soo Yoo 0000-0002-8501-7922; Kevin V Houston 0000-0002-8441-0132; Steve M D'Souza 0000-0003-3772-2616; Alsiddig Elmahdi 0000-0003-4808-7232; Isaac Davis 0000-0003-1962-6719; Ana Vilela 0000-0002-0597-0247; Parth J Parekh 0000-0003-4750-775X; David A Johnson 0000-0002-8737-0711.

Corresponding Author's Membership in Professional Societies: American College of Gastroenterology; American Society for Gastrointestinal Endoscopy.

S-Editor: Liu JH
L-Editor: A
P-Editor: Liu JH

REFERENCES

1 **Kröner PT**, Engels MM, Glicksberg BS, Johnson KW, Mzaik O, van Hooft JE, Wallace MB, El-Serag HB, Krittanawong C. Artificial intelligence in gastroenterology: A state-of-the-art review. *World J Gastroenterol* 2021; **27**: 6794-6824 [PMID: 34790008 DOI: 10.3748/wjg.v27.i40.6794]

2 **Min JK**, Kwak MS, Cha JM. Overview of Deep Learning in Gastrointestinal Endoscopy. *Gut Liver* 2019; **13**: 388-393 [PMID: 30630221 DOI: 10.5009/gnl18384]

3 **Yang YJ**, Bang CS. Application of artificial intelligence in gastroenterology. *World J Gastroenterol* 2019; **25**: 1666-1683

[PMID: 31011253 DOI: 10.3748/wjg.v25.i14.1666]
4 **Correia FP**, Lourenço LC. Artificial intelligence application in diagnostic gastrointestinal endoscopy - Deus ex machina? *World J Gastroenterol* 2021; **27**: 5351-5361 [PMID: 34539137 DOI: 10.3748/wjg.v27.i32.5351]
5 **Islam RS**, Patel NC, Lam-Himlin D, Nguyen CC. Gastric polyps: a review of clinical, endoscopic, and histopathologic features and management decisions. *Gastroenterol Hepatol (N Y)* 2013; **9**: 640-651 [PMID: 24764778]
6 **Kővári B**, Kim BH, Lauwers GY. The pathology of gastric and duodenal polyps: current concepts. *Histopathology* 2021; **78**: 106-124 [PMID: 33382489 DOI: 10.1111/his.14275]
7 **Cao C**, Wang R, Yu Y, Zhang H, Sun C. Gastric polyp detection in gastroscopic images using deep neural network. *PLoS One* 2021; **16**: e0250632 [PMID: 33909671 DOI: 10.1371/journal.pone.0250632]
8 **Wu L**, Zhang J, Zhou W, An P, Shen L, Liu J, Jiang X, Huang X, Mu G, Wan X, Lv X, Gao J, Cui N, Hu S, Chen Y, Hu X, Li J, Chen D, Gong D, He X, Ding Q, Zhu X, Li S, Wei X, Li X, Wang X, Zhou J, Zhang M, Yu HG. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut* 2019; **68**: 2161-2169 [PMID: 30858305 DOI: 10.1136/gutjnl-2018-317366]
9 **Chen D**, Wu L, Li Y, Zhang J, Liu J, Huang L, Jiang X, Huang X, Mu G, Hu S, Hu X, Gong D, He X, Yu H. 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 [PMID: 31541626 DOI: 10.1016/j.gie.2019.09.016]
10 **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**: 9545920 [PMID: 28894460 DOI: 10.1155/2017/9545920]
11 **Zhang X**, Chen F, Yu T, An J, Huang Z, Liu J, Hu W, Wang L, Duan H, Si J. Real-time gastric polyp detection using convolutional neural networks. *PLoS One* 2019; **14**: e0214133 [PMID: 30908513 DOI: 10.1371/journal.pone.0214133]
12 **Xu M**, Zhou W, Wu L, Zhang J, Wang J, Mu G, Huang X, Li Y, Yuan J, Zeng Z, Wang Y, Huang L, Liu J, Yu H. Artificial intelligence in the diagnosis of gastric precancerous conditions by image-enhanced endoscopy: a multicenter, diagnostic study (with video). *Gastrointest Endosc* 2021; **94**: 540-548.e4 [PMID: 33722576 DOI: 10.1016/j.gie.2021.03.013]
13 **Machlowska J**, Baj J, Sitarz M, Maciejewski R, Sitarz R. Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. *Int J Mol Sci* 2020; **21** [PMID: 32512697 DOI: 10.3390/IJMS21114012]
14 **Jiang K**, Jiang X, Pan J, Wen Y, Huang Y, Weng S, Lan S, Nie K, Zheng Z, Ji S, Liu P, Li P, Liu F. Current Evidence and Future Perspective of Accuracy of Artificial Intelligence Application for Early Gastric Cancer Diagnosis With Endoscopy: A Systematic and Meta-Analysis. *Front Med (Lausanne)* 2021; **8**: 629080 [PMID: 33791323 DOI: 10.3389/fmed.2021.629080]
15 **Hosokawa O**, Tsuda S, Kidani E, Watanabe K, Tanigawa Y, Shirasaki S, Hayashi H, Hinoshita T. Diagnosis of gastric cancer up to three years after negative upper gastrointestinal endoscopy. *Endoscopy* 1998; **30**: 669-674 [PMID: 9865554 DOI: 10.1055/S-2007-1001386]
16 **Menon S**, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? *Endosc Int Open* 2014; **2**: E46-E50 [PMID: 26135259 DOI: 10.1055/s-0034-1365524]
17 **Zhu Y**, Wang QC, Xu MD, Zhang Z, Cheng J, Zhong YS, Zhang YQ, Chen WF, Yao LQ, Zhou PH, Li QL. 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 [PMID: 30452913 DOI: 10.1016/j.gie.2018.11.011]
18 **Li Y**, da Zhou TTL, Shen XZ. Application of deep learning in image recognition and diagnosis of gastric cancer. *Artif Intell Gastrointest Endosc* 2021; **2**: 12-24 [DOI: 10.37126/aige.v2.i2.12]
19 **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 [PMID: 29399610 DOI: 10.1055/s-0043-120830]
20 **Bang CS**, Lee JJ, Baik GH. Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy. *J Med Internet Res* 2020; **22**: e21983 [PMID: 32936088 DOI: 10.2196/21983]
21 **Zhang X**, Hu W, Chen F, Liu J, Yang Y, Wang L, Duan H, Si J. Gastric precancerous diseases classification using CNN with a concise model. *PLoS One* 2017; **12**: e0185508 [PMID: 28950010 DOI: 10.1371/journal.pone.0185508]
22 **Li L**, Chen Y, Shen Z, Zhang X, Sang J, Ding Y, Yang X, Li J, Chen M, Jin C, Chen C, Yu C. Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer* 2020; **23**: 126-132 [PMID: 31332619 DOI: 10.1007/S10120-019-00992-2]
23 **Ikenoyama Y**, Hirasawa T, Ishioka M, Namikawa K, Yoshimizu S, Horiuchi Y, Ishiyama A, Yoshio T, Tsuchida T, Takeuchi Y, Shichijo S, Katayama N, Fujisaki J, Tada T. Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists. *Dig Endosc* 2021; **33**: 141-150 [PMID: 32282110 DOI: 10.1111/den.13688]
24 **Hirasawa T**, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, Ohnishi T, Fujishiro M, Matsuo K, Fujisaki J, Tada T. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 2018; **21**: 653-660 [PMID: 29335825 DOI: 10.1007/s10120-018-0793-2]
25 **Shelhamer E**, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017; **39**: 640-651 [PMID: 27244717 DOI: 10.1109/TPAMI.2016.2572683]
26 **Lee JH**, Kim YJ, Kim YW, Park S, Choi YI, Park DK, Kim KG, Chung JW. Spotting malignancies from gastric endoscopic images using deep learning. *Surg Endosc* 2019; **33**: 3790-3797 [PMID: 30719560 DOI: 10.1007/s00464-019-06677-2]
27 **Nguyen DT**, Lee MB, Pham TD, Batchuluun G, Arsalan M, Park KR. Enhanced Image-Based Endoscopic Pathological Site Classification Using an Ensemble of Deep Learning Models. *Sensors (Basel)* 2020; **20** [PMID: 33105736 DOI: 10.3390/S20215982]
28 **Siegel RL**, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* 2022; **72**: 7-33 [PMID: 35020204 DOI: 10.3322/caac.21708]

29 **Bray F**, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68**: 394-424 [PMID: 30207593 DOI: 10.3322/caac.21492]

30 **Grille VJ**, Campbell S, Gibbs JF, Bauer TL. Esophageal cancer: the rise of adenocarcinoma over squamous cell carcinoma in the Asian belt. *J Gastrointest Oncol* 2021; **12**: S339-S349 [PMID: 34422398 DOI: 10.21037/jgo-2019-gi-08]

31 **Thrift AP**. Global burden and epidemiology of Barrett oesophagus and oesophageal cancer. *Nat Rev Gastroenterol Hepatol* 2021; **18**: 432-443 [PMID: 33603224 DOI: 10.1038/s41575-021-00419-3]

32 **Shaheen NJ**, Falk GW, Iyer PG, Gerson LB; American College of Gastroenterology. ACG Clinical Guideline: Diagnosis and Management of Barrett's Esophagus. *Am J Gastroenterol* 2016; **111**: 30-50; quiz 51 [PMID: 26526079 DOI: 10.1038/ajg.2015.322]

33 **Hamade N**, Sharma P. 'Artificial intelligence in Barrett's Esophagus'. *Ther Adv Gastrointest Endosc* 2021; **14**: 26317745211049964 [PMID: 34671724 DOI: 10.1177/26317745211049964]

34 **Berzin TM**, Parasa S, Wallace MB, Gross SA, Repici A, Sharma P. Position statement on priorities for artificial intelligence in GI endoscopy: a report by the ASGE Task Force. *Gastrointest Endosc* 2020; **92**: 951-959 [PMID: 32565188 DOI: 10.1016/j.gie.2020.06.035]

35 **Rodríguez de Santiago E**, Hernanz N, Marcos-Prieto HM, De-Jorge-Turrión MÁ, Barreiro-Alonso E, Rodríguez-Escaja C, Jiménez-Jurado A, Sierra-Morales M, Pérez-Valle I, Machado-Volpato N, García-Prada M, Núñez-Gómez L, Castaño-García A, García García de Paredes A, Peñas B, Vázquez-Sequeiros E, Albillos A. Rate of missed oesophageal cancer at routine endoscopy and survival outcomes: A multicentric cohort study. *United European Gastroenterol J* 2019; **7**: 189-198 [PMID: 31080603 DOI: 10.1177/2050640618811477]

36 **Li N**, Jin SZ. Artificial intelligence and early esophageal cancer. *Artif Intell Gastrointest Endosc* 2021; **2**: 198-210 [DOI: 10.37126/aige.v2.i5.198]

37 **Chang K**, Jackson CS, Vega KJ. Artificial intelligence in Barrett's esophagus: A renaissance but not a reformation. *Artif Intell* 2020; **1**: 28-32 [DOI: 10.37126/aige.v1.i2.28]

38 **van der Sommen F**, Zinger S, Curvers WL, Bisschops R, Pech O, Weusten BL, Bergman JJ, de With PH, Schoon EJ. Computer-aided detection of early neoplastic lesions in Barrett's esophagus. *Endoscopy* 2016; **48**: 617-624 [PMID: 27100718 DOI: 10.1055/s-0042-105284]

39 **de Groof J**, van der Sommen F, van der Putten J, Struyvenberg MR, Zinger S, Curvers WL, Pech O, Meining A, Neuhaus H, Bisschops R, Schoon EJ, de With PH, Bergman JJ. 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 [PMID: 31065371 DOI: 10.1177/2050640619837443]

40 **de Groof AJ**, Struyvenberg MR, van der Putten J, van der Sommen F, Fockens KN, Curvers WL, Zinger S, Pouw RE, Coron E, Baldaque-Silva F, Pech O, Weusten B, Meining A, Neuhaus H, Bisschops R, Dent J, Schoon EJ, de With PH, Bergman JJ. 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* 2020; **158**: 915-929.e4 [PMID: 31759929 DOI: 10.1053/j.gastro.2019.11.030]

41 **Ebigbo A**, Mendel R, Probst A, Manzeneder J, Prinz F, de Souza LA Jr, Papa J, Palm C, Messmann H. Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus. *Gut* 2020; **69**: 615-616 [PMID: 31541004 DOI: 10.1136/gutjnl-2019-319460]

42 **Hashimoto R**, Requa J, Dao T, Ninh A, Tran E, Mai D, Lugo M, El-Hage Chehade N, Chang KJ, Karnes WE, Samarasena JB. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointest Endosc* 2020; **91**: 1264-1271.e1 [PMID: 31930967 DOI: 10.1016/j.gie.2019.12.049]

43 **Yu H**, Singh R, Shin SH, Ho KY. Artificial intelligence in upper GI endoscopy - current status, challenges and future promise. *J Gastroenterol Hepatol* 2021; **36**: 20-24 [PMID: 33448515 DOI: 10.1111/jgh.15354]

44 **Trindade AJ**, McKinley MJ, Fan C, Leggett CL, Kahn A, Pleskow DK. Endoscopic Surveillance of Barrett's Esophagus Using Volumetric Laser Endomicroscopy With Artificial Intelligence Image Enhancement. *Gastroenterology* 2019; **157**: 303-305 [PMID: 31078625 DOI: 10.1053/j.gastro.2019.04.048]

45 **Swager AF**, van der Sommen F, Klomp SR, Zinger S, Meijer SL, Schoon EJ, Bergman JJGHM, de With PH, Curvers WL. Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017; **86**: 839-846 [PMID: 28322771 DOI: 10.1016/j.gie.2017.03.011]

46 **Struyvenberg MR**, van der Sommen F, Swager AF, de Groof AJ, Rikos A, Schoon EJ, Bergman JJ, de With PHN, Curvers WL. Improved Barrett's neoplasia detection using computer-assisted multiframe analysis of volumetric laser endomicroscopy. *Dis Esophagus* 2020; **33** [PMID: 31364700 DOI: 10.1093/DOTE/DOZ065]

47 **Ali S**, Bailey A, Ash S, Haghighat M; TGU Investigators, Leedham SJ, Lu X, East JE, Rittscher J, Braden B. A Pilot Study on Automatic Three-Dimensional Quantification of Barrett's Esophagus for Risk Stratification and Therapy Monitoring. *Gastroenterology* 2021; **161**: 865-878.e8 [PMID: 34116029 DOI: 10.1053/j.gastro.2021.05.059]

48 **Byrne MF**, Critchley-Thorne RJ. Move Over, Colon. It's Time for the Esophagus to Take Center Stage for Artificial Intelligence and Computer-Aided Detection of Barrett's! *Gastroenterology* 2021; **161**: 802-804 [PMID: 34197829 DOI: 10.1053/j.gastro.2021.06.071]

49 **Ebigbo A**, Palm C, Messmann H. Barrett esophagus: What to expect from Artificial Intelligence? *Best Pract Res Clin Gastroenterol* 2021; **52-53**: 101726 [PMID: 34172253 DOI: 10.1016/j.bpg.2021.101726]

50 **Spechler SJ**, Souza RF, Rosenberg SJ, Ruben RA, Goyal RK. Heartburn in patients with achalasia. *Gut* 1995; **37**: 305-308 [PMID: 7590421 DOI: 10.1136/GUT.37.3.305]

51 **Kahrilas PJ**, Bredenoord AJ, Fox M, Gyawali CP, Roman S, Smout AJ, Pandolfino JE; International High Resolution Manometry Working Group. The Chicago Classification of esophageal motility disorders, v3.0. *Neurogastroenterol Motil* 2015; **27**: 160-174 [PMID: 25469569 DOI: 10.1111/nmo.12477]

52 **Kahrilas PJ**, Kishk SM, Helm JF, Dodds WJ, Harig JM, Hogan WJ. Comparison of pseudoachalasia and achalasia. *Am J Med* 1987; **82**: 439-446 [PMID: 3548347 DOI: 10.1016/0002-9343(87)90443-8]

53 **Tuason J**, Inoue H. Current status of achalasia management: a review on diagnosis and treatment. *J Gastroenterol* 2017; **52**: 401-406 [PMID: 28188367 DOI: 10.1007/s00535-017-1314-5]

54 **Yadlapati R**, Kahrilas PJ, Fox MR, Bredenoord AJ, Prakash Gyawali C, Roman S, Babaei A, Mittal RK, Rommel N, Savarino E, Sifrim D, Smout A, Vaezi MF, Zerbib F, Akiyama J, Bhatia S, Bor S, Carlson DA, Chen JW, Cisternas D, Cock C, Coss-Adame E, de Bortoli N, Defilippi C, Fass R, Ghoshal UC, Gonlachanvit S, Hani A, Hebbard GS, Wook Jung K, Katz P, Katzka DA, Khan A, Kohn GP, Lazarescu A, Lengliner J, Mittal SK, Omari T, Park MI, Penagini R, Pohl D, Richter JE, Serra J, Sweis R, Tack J, Tatum RP, Tutuian R, Vela MF, Wong RK, Wu JC, Xiao Y, Pandolfino JE. Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0©. *Neurogastroenterol Motil* 2021; **33**: e14058 [PMID: 33373111 DOI: 10.1111/nmo.14058]

55 **Pandolfino JE**, Kwiatek MA, Nealis T, Bulsiewicz W, Post J, Kahrilas PJ. Achalasia: a new clinically relevant classification by high-resolution manometry. *Gastroenterology* 2008; **135**: 1526-1533 [PMID: 18722376 DOI: 10.1053/j.gastro.2008.07.022]

56 **Carlson DA**, Lin Z, Kahrilas PJ, Sternbach J, Hungness ES, Soper NJ, Balla M, Listernick Z, Tye M, Ritter K, Craft J, Ciolino JD, Pandolfino JE. High-Resolution Impedance Manometry Metrics of the Esophagogastric Junction for the Assessment of Treatment Response in Achalasia. *Am J Gastroenterol* 2016; **111**: 1702-1710 [PMID: 27698386 DOI: 10.1038/ajg.2016.442]

57 **Carlson DA**, Kou W, Rooney KP, Baumann AJ, Donnan E, Triggs JR, Teitelbaum EN, Holmstrom A, Hungness E, Sethi S, Kahrilas PJ, Pandolfino JE. Achalasia subtypes can be identified with functional luminal imaging probe (FLIP) panometry using a supervised machine learning process. *Neurogastroenterol Motil* 2021; **33**: e13932 [PMID: 32608147 DOI: 10.1111/nmo.13932]

58 **Torres-Aguilera M**, Remes Troche JM. Achalasia and esophageal cancer: risks and links. *Clin Exp Gastroenterol* 2018; **11**: 309-316 [PMID: 30233226 DOI: 10.2147/CEG.S141642]

59 **Tustumi F**, Bernardo WM, da Rocha JRM, Szachnowicz S, Seguro FC, Bianchi ET, Sallum RAA, Cecconello I. Esophageal achalasia: a risk factor for carcinoma. A systematic review and meta-analysis. *Dis Esophagus* 2017; **30**: 1-8 [PMID: 28859394 DOI: 10.1093/dote/dox072]

60 **Leeuwenburgh I**, Scholten P, Caljé TJ, Vaessen RJ, Tilanus HW, Hansen BE, Kuipers EJ. Barrett's esophagus and esophageal adenocarcinoma are common after treatment for achalasia. *Dig Dis Sci* 2013; **58**: 244-252 [PMID: 23179142 DOI: 10.1007/s10620-012-2157-9]

61 **Gummin DD**, Mowry JB, Spyker DA, Brooks DE, Fraser MO, Banner W. 2016 Annual Report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 34th Annual Report. *Clin Toxicol (Phila)* 2017; **55**: 1072-1252 [PMID: 29185815 DOI: 10.1080/15563650.2017.1388087]

62 **Boskovic A**, Stankovic I. Predictability of gastroesophageal caustic injury from clinical findings: is endoscopy mandatory in children? *Eur J Gastroenterol Hepatol* 2014; **26**: 499-503 [PMID: 24642691 DOI: 10.1097/MEG.0000000000000060]

63 **Poley JW**, Steyerberg EW, Kuipers EJ, Dees J, Hartmans R, Tilanus HW, Siersema PD. Ingestion of acid and alkaline agents: outcome and prognostic value of early upper endoscopy. *Gastrointest Endosc* 2004; **60**: 372-377 [PMID: 15332026 DOI: 10.1016/s0016-5107(04)01722-5]

64 **Zargar SA**, Kochhar R, Mehta S, Mehta SK. The role of fiberoptic endoscopy in the management of corrosive ingestion and modified endoscopic classification of burns. *Gastrointest Endosc* 1991; **37**: 165-169 [PMID: 2032601 DOI: 10.1016/s0016-5107(91)70678-0]

65 **Domper Arnal MJ**, Ferrández Arenas Á, Lanas Arbeloa Á. Esophageal cancer: Risk factors, screening and endoscopic treatment in Western and Eastern countries. *World J Gastroenterol* 2015; **21**: 7933-7943 [PMID: 26185366 DOI: 10.3748/wjg.v21.i26.7933]

66 **Ferlay J**, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, Bray F. Cancer statistics for the year 2020: An overview. *Int J Cancer* 2021 [PMID: 33818764 DOI: 10.1002/ijc.33588]

67 **Abnet CC**, Arnold M, Wei WQ. Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 2018; **154**: 360-373 [PMID: 28823862 DOI: 10.1053/j.gastro.2017.08.023]

68 **Sepehr A**, Kamangar F, Fahimi S, Saidi F, Abnet CC, Dawsey SM. Poor oral health as a risk factor for esophageal squamous dysplasia in northeastern Iran. *Anticancer Res* 2005; **25**: 543-546 [PMID: 15816626]

69 **D'Souza SM**, Houston K, Keenan L, Yoo BS, Parekh PJ, Johnson DA. Role of microbial dysbiosis in the pathogenesis of esophageal mucosal disease: A paradigm shift from acid to bacteria? *World J Gastroenterol* 2021; **27**: 2054-2072 [PMID: 34025064 DOI: 10.3748/wjg.v27.i18.2054]

70 **Lu YF**, Liu ZC, Li ZH, Ma WH, Wang FR, Zhang YB, Lu JB. Esophageal/gastric cancer screening in high-risk populations in Henan Province, China. *Asian Pac J Cancer Prev* 2014; **15**: 1419-1422 [PMID: 24606476 DOI: 10.7314/apjcp.2014.15.3.1419]

71 **Yip HC**, Chiu PW. Endoscopic diagnosis and management of early squamous cell carcinoma of esophagus. *J Thorac Dis* 2017; **9**: S689-S696 [PMID: 28815064 DOI: 10.21037/jtd.2017.06.57]

72 **Ohmori M**, Ishihara R, Aoyama K, Nakagawa K, Iwagami H, Matsuura N, Shichijo S, Yamamoto K, Nagaike K, Nakahara M, Inoue T, Aoi K, Okada H, Tada T. Endoscopic detection and differentiation of esophageal lesions using a deep neural network. *Gastrointest Endosc* 2020; **91**: 301-309.e1 [PMID: 31585124 DOI: 10.1016/j.gie.2019.09.034]

73 **Horie Y**, Yoshio T, Aoyama K, Yoshimizu S, Horiuchi Y, Ishiyama A, Hirasawa T, Tsuchida T, Ozawa T, Ishihara S, Kumagai Y, Fujishiro M, Maetani I, Fujisaki J, Tada T. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 2019; **89**: 25-32 [PMID: 30120958 DOI: 10.1016/j.gie.2018.07.037]

74 **Cai SL**, Li B, Tan WM, Niu XJ, Yu HH, Yao LQ, Zhou PH, Yan B, Zhong YS. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). *Gastrointest Endosc* 2019; **90**: 745-753.e2 [PMID: 31302091 DOI: 10.1016/j.gie.2019.06.044]

75 **Chiam KH**, Shin SH, Choi KC, Leiria F, Militz M, Singh R. Current Status of Mucosal Imaging with Narrow-Band Imaging in the Esophagus. *Gut Liver* 2021; **15**: 492-499 [PMID: 32307976 DOI: 10.5009/gnl20031]

76 **Tokai Y**, Yoshio T, Aoyama K, Horie Y, Yoshimizu S, Horiuchi Y, Ishiyama A, Tsuchida T, Hirasawa T, Sakakibara Y,

Yamada T, Yamaguchi S, Fujisaki J, Tada T. Application of artificial intelligence using convolutional neural networks in determining the invasion depth of esophageal squamous cell carcinoma. *Esophagus* 2020; **17**: 250-256 [PMID: 31980977 DOI: 10.1007/s10388-020-00716-x]

77 **Shimamoto Y**, Ishihara R, Kato Y, Shoji A, Inoue T, Matsueda K, Miyake M, Waki K, Kono M, Fukuda H, Matsuura N, Nagaike K, Aoi K, Yamamoto K, Nakahara M, Nishihara A, Tada T. Real-time assessment of video images for esophageal squamous cell carcinoma invasion depth using artificial intelligence. *J Gastroenterol* 2020; **55**: 1037-1045 [PMID: 32778959 DOI: 10.1007/s00535-020-01716-5]

78 **Zhao YY**, Xue DX, Wang YL, Zhang R, Sun B, Cai YP, Feng H, Cai Y, Xu JM. Computer-assisted diagnosis of early esophageal squamous cell carcinoma using narrow-band imaging magnifying endoscopy. *Endoscopy* 2019; **51**: 333-341 [PMID: 30469155 DOI: 10.1055/a-0756-8754]

79 **Dellon ES**, Liacouras CA, Molina-Infante J, Furuta GT, Spergel JM, Zevit N, Spechler SJ, Attwood SE, Straumann A, Aceves SS, Alexander JA, Atkins D, Arva NC, Blanchard C, Bonis PA, Book WM, Capocelli KE, Chehade M, Cheng E, Collins MH, Davis CM, Dias JA, Di Lorenzo C, Dohil R, Dupont C, Falk GW, Ferreira CT, Fox A, Gonsalves NP, Gupta SK, Katzka DA, Kinoshita Y, Menard-Katcher C, Kodroff E, Metz DC, Miehlke S, Muir AB, Mukkada VA, Murch S, Nurko S, Ohtsuka Y, Orel R, Papadopoulou A, Peterson KA, Philpott H, Putnam PE, Richter JE, Rosen R, Rothenberg ME, Schoepfer A, Scott MM, Shah N, Sheikh J, Souza RF, Strobel MJ, Talley NJ, Vaezi MF, Vandenplas Y, Vieira MC, Walker MM, Wechsler JB, Wershil BK, Wen T, Yang GY, Hirano I, Bredenoord AJ. Updated International Consensus Diagnostic Criteria for Eosinophilic Esophagitis: Proceedings of the AGREE Conference. *Gastroenterology* 2018; **155**: 1022-1033.e10 [PMID: 30009819 DOI: 10.1053/j.gastro.2018.07.009]

80 **Collins MH**, Martin LJ, Alexander ES, Boyd JT, Sheridan R, He H, Pentiuk S, Putnam PE, Abonia JP, Mukkada VA, Franciosi JP, Rothenberg ME. Newly developed and validated eosinophilic esophagitis histology scoring system and evidence that it outperforms peak eosinophil count for disease diagnosis and monitoring. *Dis Esophagus* 2017; **30**: 1-8 [PMID: 26857345 DOI: 10.1111/dote.12470]

81 **Guimarães P**, Keller A, Fehlmann T, Lammert F, Casper M. Deep learning-based detection of eosinophilic esophagitis. *Endoscopy* 2022; **54**: 299-304 [PMID: 34058769 DOI: 10.1055/a-1520-8116]

82 **Rosier PFWM**. Referring to: Santis-Moya F, Calvo CI, Rojas T, Dell'Oro A, Baquedano P, Saavedra A. Urodynamic and clinical features in women with overactive bladder: When to suspect concomitant voiding dysfunction? *Neurourol Urodyn* 2021; **40**: 2050-2052 [PMID: 34369016 DOI: 10.1002/nau.24766]

83 **Czyzewski T**, Daniel N, Rochman M, Caldwell JM, Osswald GA, Collins MH, Rothenberg ME, Savir Y. Machine Learning Approach for Biopsy-Based Identification of Eosinophilic Esophagitis Reveals Importance of Global features. *IEEE Open J Eng Med Biol* 2021; **2**: 218-223 [PMID: 34505063 DOI: 10.1109/ojemb.2021.3089552]

84 **Sallis BF**, Erkert L, Moñino-Romero S, Acar U, Wu R, Konnikova L, Lexmond WS, Hamilton MJ, Dunn WA, Szepfalusi Z, Vanderhoof JA, Snapper SB, Turner JR, Goldsmith JD, Spencer LA, Nurko S, Fiebiger E. An algorithm for the classification of mRNA patterns in eosinophilic esophagitis: Integration of machine learning. *J Allergy Clin Immunol* 2018; **141**: 1354-1364.e9 [PMID: 29273402 DOI: 10.1016/j.jaci.2017.11.027]

85 **Durazzi F**, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep* 2021; **11**: 3030 [PMID: 33542369 DOI: 10.1038/s41598-021-82726-y]

86 **Marcos-Zambrano LJ**, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovik V, Aasmets O, Berland M, Gruca A, Hasic J, Hron K, Klammsteiner T, Kolev M, Lahti L, Lopes MB, Moreno V, Naskinova I, Org E, Paciência I, Papoutsoglou G, Shigdel R, Stres B, Vilne B, Yousef M, Zdravevski E, Tsamardinos I, Carrillo de Santa Pau E, Claesson MJ, Moreno-Indias I, Truu J. Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment. *Front Microbiol* 2021; **12**: 634511 [PMID: 33737920 DOI: 10.3389/fmicb.2021.634511]

87 **Yang L**, Lu X, Nossa CW, Francois F, Peek RM, Pei Z. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. *Gastroenterology* 2009; **137**: 588-597 [PMID: 19394334 DOI: 10.1053/j.gastro.2009.04.046]

88 **Deshpande NP**, Riordan SM, Castaño-Rodríguez N, Wilkins MR, Kaakoush NO. Signatures within the esophageal microbiome are associated with host genetics, age, and disease. *Microbiome* 2018; **6**: 227 [PMID: 30558669 DOI: 10.1186/s40168-018-0611-4]

89 **Yachida S**, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, Watanabe H, Masuda K, Nishimoto Y, Kubo M, Hosoda F, Rokutan H, Matsumoto M, Takamaru H, Yamada M, Matsuda T, Iwasaki M, Yamaji T, Yachida T, Soga T, Kurokawa K, Toyoda A, Ogura Y, Hayashi T, Hatakeyama M, Nakagama H, Saito Y, Fukuda S, Shibata T, Yamada T. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* 2019; **25**: 968-976 [PMID: 31171880 DOI: 10.1038/s41591-019-0458-7]

90 **Wirbel J**, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck JS, Voigt AY, Palleja A, Ponnudurai R, Sunagawa S, Coelho LP, Schrotz-King P, Vogtmann E, Habermann N, Niméus E, Thomas AM, Manghi P, Gandini S, Serrano D, Mizutani S, Shiroma H, Shiba S, Shibata T, Yachida S, Yamada T, Waldron L, Naccarati A, Segata N, Sinha R, Ulrich CM, Brenner H, Arumugam M, Bork P, Zeller G. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* 2019; **25**: 679-689 [PMID: 30936547 DOI: 10.1038/s41591-019-0406-6]

91 **Douglas GM**, Hansen R, Jones CMA, Dunn KA, Comeau AM, Bielawski JP, Tayler R, El-Omar EM, Russell RK, Hold GL, Langille MGI, Van Limbergen J. Multi-omics differentially classify disease state and treatment outcome in pediatric Crohn's disease. *Microbiome* 2018; **6**: 13 [PMID: 29335008 DOI: 10.1186/s40168-018-0398-3]

92 **Pereira P**, Aho V, Arola J, Boyd S, Jokelainen K, Paulin L, Auvinen P, Färkkilä M. Bile microbiota in primary sclerosing cholangitis: Impact on disease progression and development of biliary dysplasia. *PLoS One* 2017; **12**: e0182924 [PMID: 28796833 DOI: 10.1371/journal.pone.0182924]

93 **Seo M**, Heo J, Yoon J, Kim SY, Kang YM, Yu J, Cho S, Kim H. Methanobrevibacter attenuation *via* probiotic intervention reduces flatulence in adult human: A non-randomised paired-design clinical trial of efficacy. *PLoS One* 2017; **12**: e0184547 [PMID: 28937980 DOI: 10.1371/journal.pone.0184547]

94 **Lo C**, Marculescu R. MetaNN: accurate classification of host phenotypes from metagenomic data using neural networks. *BMC Bioinformatics* 2019; **20**: 314 [PMID: 31216991 DOI: 10.1186/s12859-019-2833-2]

95 **Pecere S**, Milluzzo SM, Esposito G, Dilaghi E, Telese A, Eusebi LH. Applications of Artificial Intelligence for the Diagnosis of Gastrointestinal Diseases. *Diagnostics (Basel)* 2021; **11** [PMID: 34573917 DOI: 10.3390/diagnostics11091575]

96 **Repici A**, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, Ferrara E, Spadaccini M, Alkandari A, Fugazza A, Anderloni A, Galtieri PA, Pellegatta G, Carrara S, Di Leo M, Craviotto V, Lamonaca L, Lorenzetti R, Andrealli A, Antonelli G, Wallace M, Sharma P, Rosch T, Hassan C. Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology* 2020; **159**: 512-520.e7 [PMID: 32371116 DOI: 10.1053/j.gastro.2020.04.062]

97 **Liu Y**. Artificial intelligence-assisted endoscopic detection of esophageal neoplasia in early stage: The next step? *World J Gastroenterol* 2021; **27**: 1392-1405 [PMID: 33911463 DOI: 10.3748/wjg.v27.i14.1392]

98 **Jiang L**, Wu Z, Xu X, Zhan Y, Jin X, Wang L, Qiu Y. Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. *J Int Med Res* 2021; **49**: 3000605211000157 [PMID: 33771068 DOI: 10.1177/03000605211000157]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com