ISSN 2644-3260 (online)

Artificial Intelligence in *Medical Imaging*

Artif Intell Med Imaging 2022 February 28; 3(1): 1-20

Published by Baishideng Publishing Group Inc

AIMI

Artificial Intelligence in Medical Imaging

Contents

Bimonthly Volume 3 Number 1 February 28, 2022

MINIREVIEWS

Artificial Intelligence in Medical Imaging

Bimonthly Volume 3 Number 1 February 28, 2022

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Medical Imaging*, Erdal Karavaş, MD, MSc, Associate Professor, Department of Radiology, Erzincan Binali Yildirim University, Erzincan 24100, Turkey. erdalkaravas@hotmail.com

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

NAME OF JOURNAL
Artificial Intelligence in Medical Imaging

ISSN
ISSN 2644-3260 (online)

LAUNCH DATE
June 28, 2020

FREQUENCY
Bimonthly

PUBLICATION DATE
February 28, 2022

Artificial Intelligence in Medical Imaging

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

Artif Intell Med Imaging 2022 February 28; 3(1): 1-7

DOI: 10.35711/aimi.v3.i1.1

ISSN 2644-3260 (online)

MINIREVIEWS

Applications of artificial intelligence in common pulmonary diseases

Saiara Choudhury, Asad Chohan, Rahul Dadhwal, Abhay P Vakil, Rene Franco, Pahnwat Tonya Taweesedt

Specialty type: Respiratory system

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, C
Grade D (Fair): D, D, D
Grade E (Poor): 0

P-Reviewer: Hanada E, Morilla I, Muneer A, Wan XH, Wang P, Zhang W

Received: December 13, 2021
Peer-review started: December 13, 2021
First decision: January 26, 2022
Revised: February 14, 2022
Accepted: February 17, 2022
Article in press: February 17, 2022
Published online: February 28, 2022

Saiara Choudhury, Asad Chohan, Rahul Dadhwal, Abhay P Vakil, Rene Franco, Pahnwat Tonya Taweesedt, Department of Pulmonary Medicine, Corpus Christi Medical Center, Corpus Christi, TX 78411, United States

Corresponding author: Pahnwat Tonya Taweesedt, MD, Academic Fellow, Department of Pulmonary Medicine, Corpus Christi Medical Center, 3315 S Alameda St., Corpus Christi, TX 78411, United States. pahnwatt@gmail.com

Abstract

Artificial intelligence (AI) is a branch of computer science where machines are trained to imitate human-level intelligence and perform well-defined tasks. AI can provide accurate results as well as analyze vast amounts of data that cannot be analyzed *via* conventional statistical methods. AI has been utilized in pulmonary medicine for almost two decades and its utilization continues to expand. AI can help in making diagnoses and predicting outcomes in pulmonary diseases based on clinical data, chest imaging, lung pathology, and pulmonary function testing. AI-based applications enable physicians to use enormous amounts of data and improve their precision in the treatment of pulmonary diseases. Given the growing role of AI in pulmonary medicine, it is important for practitioners caring for patients with pulmonary diseases to understand how AI can work in order to implement it into clinical practices and improve patient care. The goal of this mini-review is to discuss the use of AI in pulmonary medicine and imaging in cases of obstructive lung disease, interstitial lung disease, infections, nodules, and lung cancer.

Key Words: Artificial intelligence; Machine learning; Imaging; Lung; Respiratory; Pulmonary disease; Coronavirus disease 2019

Core Tip: Artificial Intelligence (AI) has the potential to have a tremendous influence when dealing with pulmonary diseases. This review provides a glimpse of AI application in pulmonary medicine and explains how AI uses imaging data to facilitate precision medicine in our data-driven era.

Citation: Choudhury S, Chohan A, Dadhwal R, Vakil AP, Franco R, Taweesedt PT. Applications of artificial intelligence in common pulmonary diseases. *Artif Intell Med Imaging* 2022; 3(1): 1-7
URL: https://www.wjgnet.com/2644-3260/full/v3/i1/1.htm
DOI: https://dx.doi.org/10.35711/aimi.v3.i1.1

INTRODUCTION

Artificial Intelligence (AI) is a branch of computer science that aims to imitate human thinking ability, learning, planning, and reasoning to solve complex problems. In 1956, scientists began theorizing a computer's ability to learn new information by analyzing data which led to the beginning of the field of AI[1]. While the terms AI, machine learning and deep learning are often used similarly, the relationship between them needs to be clarified to avoid confusion. AI is the overall concept of the simulation of human intelligence using computer systems[2]. Meanwhile, machine learning (ML) is a field of AI which provides knowledge or information using its capability of learning and analyzing massive amounts of data from larger datasets including more variables than conventional statistical methods. Machine learning uses various algorithms to process data, such as supervised learning, unsupervised learning and reinforced learning[1]. Supervised learning involves the computer recognizing patterns from data using guidance. Whereas, unsupervised learning involves pattern recognition by the computer without any guidance[2]. Reinforced learning has the ability to recognize and analyze data without any labels, by using incremental positive or negative feedback[3]. Deep learning is a subset of ML that enables the algorithm to learn from a training data set and apply that to fulfill intended tasks to a new data set[2]. As healthcare data has become increasingly complex, AI has the potential to have a significant influence on medical data analysis and medical practice.

AI has been implemented in many fields of medicine to facilitate precision medicine by predicting outcomes, diagnosis, and therapeutic results. AI may assist in diagnosis of different diseases by recognizing the images from different parts of the body, predicting mortality in the critical care unit, classifying skin biopsies, and identifying new genotypes in heart failure. The US Food and Drug Administration (FDA) and Conformité Européenne (CE)-marked have approved more than 300 AI-based software/medical devices[4,6]. Many of them are related to pulmonary imaging (Table 1)[4,6].

In the 1980s, AI was initially introduced into pulmonary medicine to interpret lung function tests[5]. Since then, AI has been applied in various pulmonary diseases, including, but not limited to obstructive lung diseases, pulmonary infections, interstitial lung disease, and malignancy[6]. Given its widespread use in pulmonary medicine, it is important for pulmonologists to have a general understanding of the utilization of AI in this field and how it can aid them in caring for patients. In this narrative mini-review, we provided an overview of the pulmonary diseases that are commonly diagnosed and managed by general pulmonologists for which AI has been applied including obstructive lung disease, interstitial lung disease, pulmonary tuberculosis (TB), coronavirus disease 2019 (COVID-19) pneumonia, lung nodules and lung cancer (Figure 1.).

METHOD

PubMed was searched from inception to November 30, 2021, using keywords: "artificial intelligence, lung disease", " artificial intelligence, pulmonary disease", "artificial intelligence, COPD, asthma", " artificial intelligence, interstitial lung disease", "artificial intelligence, tuberculosis", "artificial intelligence, COVID-19", and "artificial intelligence, lung nodule, lung cancer". All types of published publications were included, *e.g.*, reviews, observational studies, and meta-analyses. We prioritized recent articles within five years in this narrative mini-review.

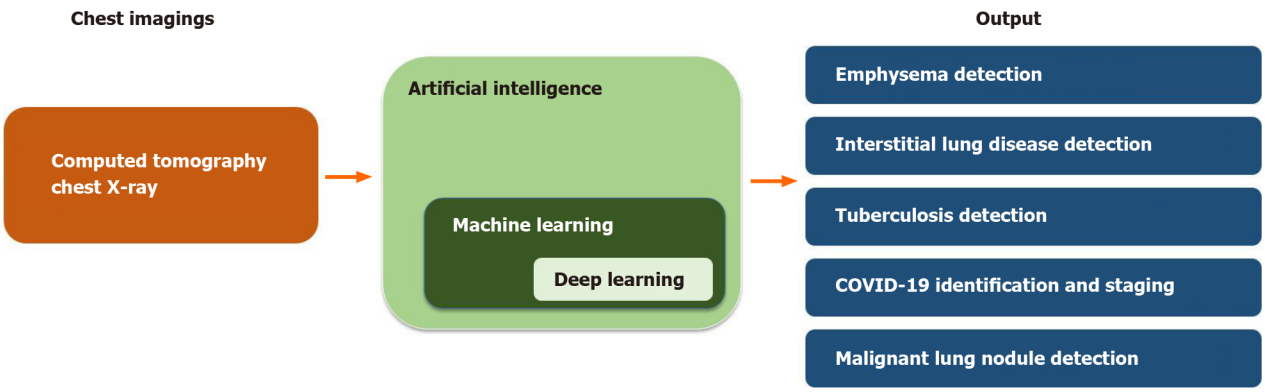
OBSTRUCTIVE LUNG DISEASES

The gold standard of diagnosis in obstructive lung diseases like asthma and chronic obstructive pulmonary disease (COPD) involves a combination of signs, symptoms, and spirometry. While AI cannot replace the clinicians' role, it can complement clinicians' interpretation of the data available at the bedside. A study by Topalovic *et al*[7] compared the accuracy of pulmonologists' interpretation of pulmonary function testing to an AI-based software that used more than 1430 historical patient cases. Both groups were asked to study 50 patient cases and correctly interpret the pulmonary function test while placing them in diagnostic categories. AI-based software was found to outperform the pulmonologist interpretation by a substantial margin[7].

Table 1 Example of Conformité Européenne (CE)-marked, US Food and Drug Administration (FDA)-approved or FDA-permitted artificial intelligence devices

Pulmonary conditions	AI device/algorithm	Imaging	Brief description
Chronic obstructive pulmonary disease	Lung density analysis software	Chest CT	Uses three-dimensional segmentation of the lungs, volumetric analysis and density evaluations from CT images to aid in diagnosis and progression of the disease
	LungQ software	Chest CT	Quantitative analysis of lung volume. Airway morphology analysis
Interstitial lung disease	LungPrint Discovery	Chest CT	Lung tissue and airway evaluation. Quantitative analysis using deep learning to detect interstitial lung disease and chronic obstructive lung disease
	Lung Texture Analysis	Chest CT	Transforms a standard chest CT into a detailed map. Lung textures quantification
Pulmonary infection	Icolung	Non-contrast Chest CT	Detects COVID-19 at an early stage and quantify the extent of lung lesions
	InferRead CT pneumonia	Chest CT	Real-time identification. Alerts of suspected pneumonia cases
Lung nodule	Syngo.CT Lung CAD	Multidetector Chest CT	Computer-aid detection tool designed to detect solid pulmonary nodules using convolutional neural network. To be used as the second reader.
	AI-Rad Companion (Pulmonary)	CT DICOM chest	Quantitative and qualitative analysis using deep learning. Segmentation of lung lobes and identification of lesions
	Temporal Comparison software	Chest X-ray	The new image is superimposed on the old image to detect changes in the lung parenchyma.
	ClearRead CT	CT chest	Lung nodule detection asymptomatic population

COVID-19: Coronavirus disease-2019; CT: Computed tomography; DICOM: Digital imaging and communication.



DOI: 10.35711/aimi.v3.i1.1

Figure 1 Representative diagram showing examples of artificial intelligence applications in pulmonary diseases. COVID-19: Coronavirus disease-2019.

COPD

According to the Global Strategy for Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease (GOLD) reports 2022, COPD is one of the top three causes of death in the world[8]. Moll *et al*[9] also proposed a machine learning mortality prediction model for patients with COPD based on six-minute walk tests, percent predicted of forced expiratory volume in 1 second (FEV1), and age. While the gold standard of diagnosis of COPD is spirometry, studies have suggested that artificial intelligence and deep learning can potentially be utilized to screen patients for COPD. Tang *et al*[10] suggests that low dose computed tomography (CT) screening of the lungs of both smokers and ex-smokers can be examined using deep residual networks to identify patients who may have COPD but remain undiagnosed. AI has also been used to characterize patients already diagnosed with COPD. The Genetic Epidemiology Study (COPDGene) is one of the largest data sets obtained over ten years, consisting of chest imaging, spirometry, and molecular data from patients with COPD. This has been used as the source for multiple studies that have related specific COPD phenotypes to genetic and molecular mechanisms and has led to the prediction of the disease progression of various COPD

subtypes[11]. A study by Fischer *et al*[12] describes an algorithm that can perform lung lobe segmentation and emphysema quantification, which has been shown to correlate with different GOLD stages in patients with COPD per their spirometry data. Furthermore, AI-based applications have also been suggested to help patients identify if they may be having an exacerbation at home and when they should seek help from a medical professional[13]. This can promote patient responsibility and potentially save on resources, including emergency department visits.

Asthma

Asthma is an intermittent and reversible obstructive lung disease with multiple phenotypes. AI may improve diagnosis, phenotype classification, prediction of asthma exacerbations and treatment response [1,15]. Multiple studies have shown good accuracy of ML-based algorithms in screening and diagnosis of asthma in adult patients[1]. In regards to phenotype classification, when using the machine learning approach as well as cluster analysis, the highest corticosteroid-responsiveness phenotype was identified in patients with low pulmonary function, high serum eosinophils, nasal polyps, and late-onset asthma [14]. The least corticosteroid-responsiveness phenotype was also found in young, obese females with early-onset asthma[14]. In another study, Qin *et al*[15] adopted deep learning algorithms-based high-resolution computed tomography (HRCT) chest images to assess small airway thickness with the aim of steroids response evaluation in asthma patients with small airway obstruction. Phenotype identification can help tailor asthma management and possibly improve outcomes.

INTERSTITIAL LUNG DISEASE

Interstitial lung disease (ILD) is an umbrella term that encompasses all disease processes that can cause pleural/parenchymal inflammation and scarring. Deep learning algorithms can help with the diagnosis of ILD using HRCT chest images. In a case-control study by Walsh *et al*[16], a database of 1157 de-identified HRCT images showing evidence of diffuse fibrotic lung disease were classified using the American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/Latin American Thoracic Association (ATS/ERS/JRS/ALAT) idiopathic pulmonary fibrosis guidelines. These images were divided into multiple groups and separately read by a deep learning algorithm and 91 thoracic radiologists. Walsh *et al*[16] found that the algorithm outperformed thoracic radiologists' interpretation of HRCT images with the median accuracy of 73.3% *vs* 70.7%, respectively. This study showed that deep learning algorithms could serve as a valuable tool in the diagnosis of ILD. Similarly, Choe *et al*[17] has revealed that deep learning increases the diagnostic accuracy of chronic hypersensitivity pneumonitis, cryptogenic organizing pneumonia, nonspecific interstitial pneumonia, and usual interstitial pneumonia patterns. Other studies have used AI algorithms to evaluate HRCT images of patients with interstitial pulmonary fibrosis and have successfully been able to quantify airway volumes and parenchymal lesions[17,18].

PULMONARY INFECTIONS

The utilization of AI has also been investigated in multiple pulmonary infections. Here, we briefly review the utilization of AI in pulmonary tuberculosis and COVID-19.

Tuberculosis

Tuberculosis (TB) remains a significant cause of mortality in many parts of the world. Due to the variable presentations of TB in chest radiography, diagnosis remains a challenge. The first conventional computer-aided diagnosis (CAD) was made in 2016 to aid in the detection of TB. Over the years, investigators have also developed multiple CAD algorithms that can detect various radiographic findings in TB, for example, cavitary and focal TB[19]. In addition to diagnosis, AI can be helpful in other aspects of TB care as well. AI has been suggested as an aid to review records, identify symptomatic patterns, surveillance, and factors that may contribute to the treatment and medication adherence failure in TB [20]. Doshi *et al*[21] describe innovative ways in which AI-based software can provide access to care and facilitate the management of TB patients worldwide.

COVID-19

In recent times, COVID-19 has taken the world by storm. Morbidity and mortality around the world have risen as treatment options for COVID-19 remain largely experimental. AI software has been developed to aid in the early diagnosis and prognostication of patients with COVID-19. In a retrospective, multi-center study by Li *et al*[22], a deep learning model, called COVID-19 detection neural network was developed to identify CT findings of COVID-19 infection and differentiate it from CT findings in community-acquired pneumonia. Another study developed a deep learning convolution neural network to effectively stage the severity of COVID-19 infection *via* scoring of various

radiographic features[23]. This can help in early prognostication of the disease, which can lead to making early treatment decisions. Another study by Burdick *et al*[24] used ML algorithm to build a model which uses inputs of diastolic blood pressure, systolic blood pressure, heart rate, temperature, respiratory rate, oxygen saturation, white blood cell, platelet count, lactate, blood urea nitrogen, creatinine, and bilirubin to predict the need for mechanical ventilation. Furthermore, investigators have developed deep learning algorithms which help to identify protein structures and shapes. The data provided using this algorithm has been invaluable in the development of the COVID-19 vaccine[6].

PULMONARY NODULES AND LUNG MALIGNANCY

Despite recent advances in the treatment of pulmonary malignancies, the World Health Organization considers them among the deadliest of all solid malignancies[25]. Early and accurate diagnosis remains paramount in improving patient outcomes. CAD systems use deep learning algorithms as an aid for radiologists to analyze CT images by lung segmentation and provide a more focused analysis that will allow nodule detection and classification. One such state-of-the-art algorithm implemented by Siemen Healthcare uses statistical finite element analysis or three-dimensional lung segmentation in adversarial neural network training[26]. A study by Chauvie *et al*[27] compared different machine learning algorithms and lung-RADs criteria and concluded that neural network algorithms enhanced the positive predictive value in chest digital tomosynthesis in lung cancer detection. One identified disadvantage of deep learning is that it does not provide uniform features for identifying malignant versus benign nodules. This problem has been addressed using a method called Radiomics[28]. Radiomics uses features from one image in order to provide data-characterization algorithms that helps to identify similar features in new data. This tool can help in finding characteristics of malignancies that can be otherwise missed by human experts. The combination of Radiomics and deep learning promises the ability to provide radiologists around the world an advantage in diagnosing pulmonary malignancies. Finally, a study by Afshar *et al*[29] has proposed a deep learning-based Radiomics model to predict the time-to-event outcome prediction, that utilizes raw images of CT and PET (Positron Emission Tomography) scans and can calculate the image-based risk of death or recurrence, for each patient.

LIMITATIONS OF AI IN CLINICAL PRACTICE

Despite the promising outcomes of AI, small or unstructured databases and missing data may result in unsatisfactory AI quality. For example, in the diagnosis of lung nodules and lung malignancy, the software's ability is usually compared to the ability of expert radiologists. However, since the ultimate goal is to diagnose malignancies and not just identify lung nodules, algorithms should be made to focus on identifying malignancies with a different reference standard[30]. Similarly, AI poses other limitations as well. For example, characteristics of CT imaging are being primarily used as an input for AI algorithm to diagnose early COVID-19 infection. However, it should be noted that while CT scan has high sensitivity it does not have very high specificity for COVID-19. So, diagnosing the disease based solely on CT images with the help of AI may be erroneous[31]. Therefore, while AI has many advantages, it is important to keep these limitations in mind. Finally, cooperation between physicians and AI researchers is needed to be able to develop well-structured AI applications that can be validated in real-world study before launching AI models into clinical fields.

CONCLUSION

The implementation of AI and machine learning algorithms is an evolving and relevant topic in pulmonary medicine. Human errors can occur in the medical field. It can be associated with missed, late, and incorrect diagnoses leading to health and economic burden. AI is an efficient tool that can be implemented to prevent this problem by aiding in the fast, accurate, and early diagnosis, prognostication, as well as treatment of pulmonary diseases. Nonetheless, the lack of knowledge and confidence in applying AI into practice may hinder the utilization of AI in the medical field. Moreover, well-performed AI algorithms require a large well quality database. Physician and AI algorithm developers should work closely to minimize these limitations. While AI alone cannot replace clinician expertise, it can add to the armamentarium and improve patient care and healthcare worldwide.

FOOTNOTES

Author contributions: Choudhury S performed the majority of the writing; Chohan AA provided the input in writing the paper; Dadhwal R provided the input in writing the paper; Abhay V provided the input in writing the paper;

Franco R provided the input in writing the paper; Taweesedt PT designed the outline and coordinated the writing of the paper.

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

Country/Territory of origin: United States

ORCID number: Saiara Choudhury 0000-0001-7225-4905; Asad Chohan 0000-0002-1801-9792; Rahul Dadhwal 0000-0001-6963-1466; Abhay P Vakil 0000-0003-4947-0233; Rene Franco 0000-0002-0684-7732; Pahnwat Tonya Taweesedt 0000-0002-5791-6920.

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

REFERENCES

1 **Feng Y**, Wang Y, Zeng C, Mao H. Artificial Intelligence and Machine Learning in Chronic Airway Diseases: Focus on Asthma and Chronic Obstructive Pulmonary Disease. *Int J Med Sci* 2021; **18**: 2871-2889 [PMID: 34220314 DOI: 10.7150/ijms.58191]
2 **Kaplan A**, Cao H, FitzGerald JM, Iannotti N, Yang E, Kocks JWH, Kostikas K, Price D, Reddel HK, Tsiligianni I, Vogelmeier CF, Pfister P, Mastoridis P. Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis. *J Allergy Clin Immunol Pract* 2021; **9**: 2255-2261 [PMID: 33618053 DOI: 10.1016/j.jaip.2021.02.014]
3 **Ali I**, Hart GR, Gunabushanam G, Liang Y, Muhammad W, Nartowt B, Kane M, Ma X, Deng J. Lung Nodule Detection *via* Deep Reinforcement Learning. *Front Oncol* 2018; **8**: 108 [PMID: 29713615 DOI: 10.3389/fonc.2018.00108]
4 The Medical Futurist. The Medical Futurist. Available from: https://medicalfuturist.com/fda-approved-ai-based-algorithms/
5 **Aikins JS**, Kunz JC, Shortliffe EH, Fallat RJ. PUFF: an expert system for interpretation of pulmonary function data. *Comput Biomed Res* 1983; **16**: 199-208 [PMID: 6347509 DOI: 10.1016/0010-4809(83)90021-6]
6 **Khemasuwan D**, Sorensen JS, Colt HG. Artificial intelligence in pulmonary medicine: computer vision, predictive model and COVID-19. *Eur Respir Rev* 2020; **29** [PMID: 33004526 DOI: 10.1183/16000617.0181-2020]
7 **Topalovic M**, Das N, Burgel PR, Daenen M, Derom E, Haenebalcke C, Janssen R, Kerstjens HAM, Liistro G, Louis R, Ninane V, Pison C, Schlesser M, Vercauter P, Vogelmeier CF, Wouters E, Wynants J, Janssens W; Pulmonary Function Study Investigators; Pulmonary Function Study Investigators:. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J* 2019; **53** [PMID: 30765505 DOI: 10.1183/13993003.01660-2018]
8 GOLD-REPORT-2022-v1.1-22Nov2021_WMV.pdf. Available from: https://goldcopd.org/wp-content/uploads/2021/11/GOLD-REPORT-2022-v1.1-22Nov2021_WMV.pdf
9 **Moll M**, Qiao D, Regan EA, Hunninghake GM, Make BJ, Tal-Singer R, McGeachie MJ, Castaldi PJ, San Jose Estepar R, Washko GR, Wells JM, LaFon D, Strand M, Bowler RP, Han MK, Vestbo J, Celli B, Calverley P, Crapo J, Silverman EK, Hobbs BD, Cho MH. Machine Learning and Prediction of All-Cause Mortality in COPD. *Chest* 2020; **158**: 952-964 [PMID: 32353417 DOI: 10.1016/j.chest.2020.02.079]
10 **Tang LYW**, Coxson HO, Lam S, Leipsic J, Tam RC, Sin DD. Towards large-scale case-finding: training and validation of residual networks for detection of chronic obstructive pulmonary disease using low-dose CT. *Lancet Digit Health* 2020; **2**: e259-e267 [PMID: 33328058 DOI: 10.1016/S2589-7500(20)30064-9]
11 **Castaldi PJ**, Boueiz A, Yun J, Estepar RSJ, Ross JC, Washko G, Cho MH, Hersh CP, Kinney GL, Young KA, Regan EA, Lynch DA, Criner GJ, Dy JG, Rennard SI, Casaburi R, Make BJ, Crapo J, Silverman EK, Hokanson JE; COPDGene Investigators. Machine Learning Characterization of COPD Subtypes: Insights From the COPDGene Study. *Chest* 2020; **157**: 1147-1157 [PMID: 31887283 DOI: 10.1016/j.chest.2019.11.039]
12 **Fischer AM**, Varga-Szemes A, Martin SS, Sperl JI, Sahbaee P, Neumann D, Gawlitza J, Henzler T, Johnson CM, Nance JW, Schoenberg SO, Schoepf UJ. Artificial Intelligence-based Fully Automated Per Lobe Segmentation and Emphysema-quantification Based on Chest Computed Tomography Compared With Global Initiative for Chronic Obstructive Lung Disease Severity of Smokers. *J Thorac Imaging* 2020; **35** Suppl 1: S28-S34 [PMID: 32235188 DOI: 10.1097/RTI.0000000000000500]
13 **Swaminathan S**, Qirko K, Smith T, Corcoran E, Wysham NG, Bazaz G, Kappel G, Gerber AN. A machine learning approach to triaging patients with chronic obstructive pulmonary disease. *PLoS One* 2017; **12**: e0188532 [PMID: 29166411 DOI: 10.1371/journal.pone.0188532]
14 **Wu W**, Bang S, Bleecker ER, Castro M, Denlinger L, Erzurum SC, Fahy JV, Fitzpatrick AM, Gaston BM, Hastie AT, Israel E, Jarjour NN, Levy BD, Mauger DT, Meyers DA, Moore WC, Peters M, Phillips BR, Phipatanakul W, Sorkness RL, Wenzel SE. Multiview Cluster Analysis Identifies Variable Corticosteroid Response Phenotypes in Severe Asthma. *Am J Respir Crit Care Med* 2019; **199**: 1358-1367 [PMID: 30682261 DOI: 10.1164/rccm.201808-1543OC]
15 **Qin Y**, Wang J, Han Y, Lu L. Deep Learning Algorithms-Based CT Images in Glucocorticoid Therapy in Asthma Children

with Small Airway Obstruction. *J Healthc Eng* 2021; 2021: 5317403 [PMID: 34721824 DOI: 10.1155/2021/5317403]

16 **Walsh SLF**, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; **6**: 837-845 [PMID: 30232049 DOI: 10.1016/S2213-2600(18)30286-8]

17 **Choe J**, Hwang HJ, Seo JB, Lee SM, Yun J, Kim MJ, Jeong J, Lee Y, Jin K, Park R, Kim J, Jeon H, Kim N, Yi J, Yu D, Kim B. Content-based Image Retrieval by Using Deep Learning for Interstitial Lung Disease Diagnosis with Chest CT. *Radiology* 2022; **302**: 187-197 [PMID: 34636634 DOI: 10.1148/radiol.2021204164]

18 **Handa T**, Tanizawa K, Oguma T, Uozumi R, Watanabe K, Tanabe N, Niwamoto T, Shima H, Mori R, Nobashi TW, Sakamoto R, Kubo T, Kurosaki A, Kishi K, Nakamoto Y, Hirai T. Novel Artificial Intelligence-based Technology for Chest Computed Tomography Analysis of Idiopathic Pulmonary Fibrosis. *Ann Am Thorac Soc* 2021 [PMID: 34410886 DOI: 10.1513/AnnalsATS.202101-044OC]

19 **Kulkarni S**, Jha S. Artificial Intelligence, Radiology, and Tuberculosis: A Review. *Acad Radiol* 2020; **27**: 71-75 [PMID: 31759796 DOI: 10.1016/j.acra.2019.10.003]

20 **Chopra KK**, Arora VK. Artificial intelligence and TB management - The way forward. *Indian J Tuberc* 2020; **67**: 1-2 [PMID: 32192602 DOI: 10.1016/j.ijtb.2020.02.002]

21 **Doshi R**, Falzon D, Thomas BV, Temesgen Z, Sadasivan L, Migliori GB, Raviglione M. Tuberculosis control, and the where and why of artificial intelligence. *ERJ Open Res* 2017; **3** [PMID: 28656130 DOI: 10.1183/23120541.00056-2017]

22 **Li L**, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K, Liu D, Wang G, Xu Q, Fang X, Zhang S, Xia J. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* 2020; **296**: E65-E71 [PMID: 32191588 DOI: 10.1148/radiol.2020200905]

23 **Zhu J**, Shen B, Abbasi A, Hoshmand-Kochi M, Li H, Duong TQ. Deep transfer learning artificial intelligence accurately stages COVID-19 Lung disease severity on portable chest radiographs. *PLoS One* 2020; **15**: e0236621 [PMID: 32722697 DOI: 10.1371/journal.pone.0236621]

24 **Burdick H**, Lam C, Mataraso S, Siefkas A, Braden G, Dellinger RP, McCoy A, Vincent JL, Green-Saxena A, Barnes G, Hoffman J, Calvert J, Pellegrini E, Das R. Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. *Comput Biol Med* 2020; **124**: 103949 [PMID: 32798922 DOI: 10.1016/j.compbiomed.2020.103949]

25 **Binczyk F**, Prazuch W, Bozek P, Polanska J. Radiomics and artificial intelligence in lung cancer screening. *Transl Lung Cancer Res* 2021; **10**: 1186-1199 [PMID: 33718055 DOI: 10.21037/tlcr-20-708]

26 **Zhang Y**, Osanlouy M, Clark AR, Kumar H, King C, Wilsher ML, Milne DG, Hoffman EA, Tawhai MH. Pulmonary lobar segmentation from computed tomography scans based on a statistical finite element analysis of lobe shape. Proc. SPIE 10949, Medical Imaging 2019 [DOI: 10.1117/12.2512642]

27 **Chauvie S**, De Maggi A, Baralis I, Dalmasso F, Berchialla P, Priotto R, Violino P, Mazza F, Melloni G, Grosso M; SOS Study team. Artificial intelligence and radiomics enhance the positive predictive value of digital chest tomosynthesis for lung cancer detection within SOS clinical trial. *Eur Radiol* 2020; **30**: 4134-4140 [PMID: 32166491 DOI: 10.1007/s00330-020-06783-z]

28 **Lambin P**, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; **48**: 441-446 [PMID: 22257792 DOI: 10.1016/j.ejca.2011.11.036]

29 **Afshar P**, Mohammadi A, Tyrrell PN, Cheung P, Sigiuk A, Plataniotis KN, Nguyen ET, Oikonomou A. [Formula: see text]: deep learning-based radiomics for the time-to-event outcome prediction in lung cancer. *Sci Rep* 2020; **10**: 12366 [PMID: 32703973 DOI: 10.1038/s41598-020-69106-8]

30 **Schreuder A**, Scholten ET, van Ginneken B, Jacobs C. Artificial intelligence for detection and characterization of pulmonary nodules in lung cancer CT screening: ready for practice? *Transl Lung Cancer Res* 2021; **10**: 2378-2388 [PMID: 34164285 DOI: 10.21037/tlcr-2020-lcs-06]

31 **Belfiore MP**, Urraro F, Grassi R, Giacobbe G, Patelli G, Cappabianca S, Reginelli A. Artificial intelligence to codify lung CT in Covid-19 patients. *Radiol Med* 2020; **125**: 500-504 [PMID: 32367319 DOI: 10.1007/s11547-020-01195-x]

AIMI Artificial Intelligence in Medical Imaging

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

Artif Intell Med Imaging 2022 February 28; 3(1): 8-20

DOI: 10.35711/aimi.v3.i1.8

ISSN 2644-3260 (online)

MINIREVIEWS

Chest ultrasound in neonates: What neonatologists should know

Adel Salah Bediwy, Mohammed Al-Biltagi, Jameel Ahmed Nazeer, Nermin Kamal Saeed

Specialty type: Pediatrics

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): 0
Grade E (Poor): 0

P-Reviewer: Haurylenka D, Trovato G

Received: January 6, 2022
Peer-review started: January 6, 2022
First decision: February 10, 2022
Revised: February 12, 2022
Accepted: February 21, 2022
Article in press: February 21, 2022
Published online: February 28, 2022

Adel Salah Bediwy, Department of Chest Disease, Faculty of Medicine, Tanta University, Tanta 31511, Algharbia, Egypt

Adel Salah Bediwy, Department of Pulmonology, University Medical Center, Arabian Gulf University, Manama 26671, Manama, Bahrain

Mohammed Al-Biltagi, Department of Pediatrics, Faculty of Medicine, Tanta University, Tanta 31511, Algharbia, Egypt

Mohammed Al-Biltagi, Department of Pediatrics, University Medical Center, Arabian Gulf University, Manama 26671, Manama, Bahrain

Jameel Ahmed Nazeer, Department of Radiology, University Medical center, King Abdulla Medical City, Arabian Gulf University, Manama 26671, Manama, Bahrain

Nermin Kamal Saeed, Medical Microbiology Section, Pathology Department, Salmaniya Medical Complex , Ministry of Health, Kingdom of Bahrain , Manama 12, Manama, Bahrain

Nermin Kamal Saeed, Microbiology Section, Pathology Department, Irish Royal College of Surgeon, Bahrain, Busaiteen 15503, Muharraq, Bahrain

Corresponding author: Mohammed Al-Biltagi, MBChB, MD, MSc, PhD, Chairman, Professor, Department of Pediatrics, Faculty of Medicine, Tanta University, Al Bahr street, Tanta 31511, Algharbia, Egypt. mbelrem@hotmail.com

Abstract

For many years, ultrasound was thought to have no indications in pulmonary imaging because lungs are filled with air, creating no acoustic mismatch, as encountered by ultrasound wave beam. Lung ultrasound (LUS) was started in adult critical care settings to detect pleural effusion and acquired more indications over time. In the neonatal intensive care unit (NICU), the use of chest ultrasound has gained more attention during the last two decades. Being a radiation-free, bedside, rapid, and handy tool, LUS started to replace chest X-rays in NICU. Using LUS depends upon understanding the nature of normal lungs and the changes induced by different diseases. With the help of LUS, an experienced neonatologist can detect many of the respiratory problems so fast that interventional therapy can be introduced as early as possible. LUS can diagnose pleural effusion, pneumothorax, pneumonia, transient tachypnoea of the newborn, respiratory distress syndrome, pulmonary atelectasis, meconium aspiration syndrome, bronchopulmonary dysplasia, and some other disorders with very high accuracy. LUS will be helpful in initial diagnosis, follow-up, and predicting

the need for further procedures such as mechanical ventilation, diuretic therapy, surfactant therapy, *etc.* There are some limitations to using LUS in some respiratory disorders such as bullae, interstitial emphysema, and other conditions. This review will highlight the importance of LUS, its uses, and limitations.

Key Words: Lung ultrasound; Neonatal respiratory Disorders; Neonatal chest ultrasound; Meconium; Pneumonia; Pneumothorax

Core Tip: Lung ultrasound is a valuable imaging procedure in neonatal respiratory care. It helps diagnose many respiratory disorders with excellent accuracy and safety. Some limitations are experienced for its use, but its benefits are more.

Citation: Bediwy AS, Al-Biltagi M, Nazeer JA, Saeed NK. Chest ultrasound in neonates: What neonatologists should know. *Artif Intell Med Imaging* 2022; 3(1): 8-20
URL: https://www.wjgnet.com/2644-3260/full/v3/i1/8.htm
DOI: https://dx.doi.org/10.35711/aimi.v3.i1.8

INTRODUCTION

Lung diseases are the most common reasons of respiratory distress in newborn, leading in some instances to respiratory failure; even may end with death. Mortality caused by neonatal respiratory problems was estimated to be 11% in the United States and 32% in China[1,2]. Thus, neonatologists need to identify the etiology and pathology of lung disease causing respiratory problems. Since the sixties of the last century, applying point-of-care ultrasound (POCUS) in neonates was first illustrated, with growing interest with several applications to be used over the past two decades[3,4]. Lung ultrasound started in adult critical care medicine to diagnose various lung and pleural problems. Then in the early nineties, chest ultrasound was suggested to diagnose neonatal respiratory distress syndrome (RDS). Since then, pediatric and neonatal ultrasound of the lung has developed rapidly[5]. After that, several indications were introduced for the lung ultrasound in neonates as transient tachypnoea of the newborn (TTN), neonatal pneumonia, pneumothorax, and meconium aspiration syndrome (MAS) with high specificity and sensitivity[6-10]. Neonatal lung ultrasound (LUS) is an easy bedside procedure with no radiation exposure and can be done serially in neonates[11,12]. LUS can differentiate neonatal respiratory diseases and predict neonatal morbidity[3]. Because of its advantages, LUS aids in distinguishing the various causes of neonatal respiratory failure and guides the management[3,13]. Another advantage of performing LUS in the neonatal intensive care unit (ICU) is the immediate interpretation by the neonatologist with a more accurate diagnosis aiding to start a precise and rapid therapeutic intervention[13]. Although the European Resuscitation Council guidelines recommend utilizing LUS to confirm the placement of the endotracheal tube (ETT) diagnose cardiac tamponade, pneumothorax, and pneumonia, the use of LUS is still not routinely taught in neonatology training programs around the world[14].

There was a notable increase in publications on the use of LUS in both adults and neonates during the last fifteen years. The successful establishment of LUS programs in some neonatal intensive care units (NICU) resulted in a significant reduction in chest radiograms and, subsequently, radiation exposure to patients[12]. One study showed that the risk of cancer occurrence in infants receiving a single small dose of radiation was two to three times higher than the average population and was six to nine times higher than the risk from an exposure of a 60-year-old patient[15]. The POCUS Working Group of the European Society of Paediatric and Neonatal Intensive Care issued evidence-based guidelines on POCUS for neonates and children in 2020[16]. Because it costs less than chest radiology, being radiation-free with higher sensitivity for diagnosing small lesions close to the pleural surface, LUS has been widely used in NICUs. Recently, it has been the most preferred radiological intervention for diagnosing many diseases in the neonatal ICU as RDS, TTN, pneumothorax, MAS, pleural effusions, and neonatal pneumonia than the chest X-ray[17]. LUS is beneficial in the initial diagnosis, follow-up, and assessing the need for further procedures such as mechanical ventilation. Every neonatologist needs to know LUS and get training courses for this unique safe technique.

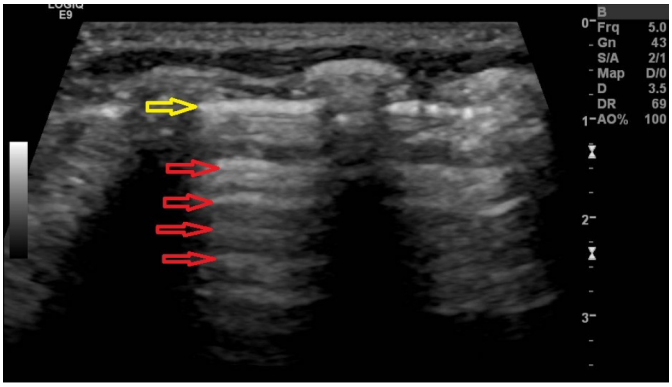
TECHNIQUE OF LUNG ULTRASOUND

Ultrasound imaging uses one principle; an interface reflects the ultrasound wave between the different media with various acoustic absorption and impedance[18]. Ultrasound is of limited use in normal well-aerated lungs as there is no acoustic discrepancy in the ultrasound beam as it confronts air[19]. LUS is very useful in neonates because of the thin chest wall and less ossification of the bony thoracic cage[11,20]. A high-frequency linear probe is preferred to perform LUS in neonates because of the relatively thinner chest walls and smaller thoraxes. This high-frequency probe gives a better image quality and allows visualization of the entire lung surface[21]. The high-frequency probe gives a good resolution with penetration to a superficial depth. We use probes with higher frequencies in preterm neonates, for example micro-linear probes with a small footprint (like a hockey stick). An operator with high experience may use different probe types[22]. Different ultrasound modes can be used for LUS. 2-Dimensional brightness (B-mode) and motion (M-mode), and the color doppler to estimate blood flow [23].

To perform lung ultrasound in neonates we perform it in the lateral, supine, or prone position. Each chest side hemithorax is divided into three areas: Posterior, anterior, and lateral, by the posterior and anterior axillary lines. We can perform longitudinal and transverse scans in all areas to directly identify the ribs, subcutaneous tissue, pleural line, and recognize the lung sliding to indirectly assess the lung tissue[21]. To evaluate or interpret the LUS images, we should understand some terms such as pleural line, A-lines, B-lines, lung sliding and acoustic shadowing artifacts (rib shadow). The pleural line (Figure 1) represents the lung's outer surface, including the visceral and parietal pleura. The pleural line is a regular and smooth hyperechoic line, moving to and fro with respiration. We can clearly visualize the pleural lines in neonates even without pleural or pulmonary pathology. It becomes apparent after birth following the first few breaths[24]. The Bat sign (Figure 2) represents a normal lung surface and is identified by visualizing the bright lateral pleural line (visceral and parietal) and the dark "bat wings" of the two adjacent ribs on each side. In the presence of lung or pleural diseases, the pleural line may become thick and coarse compared to the thin and regular hyperechoic pleural line shape in the healthy lung.

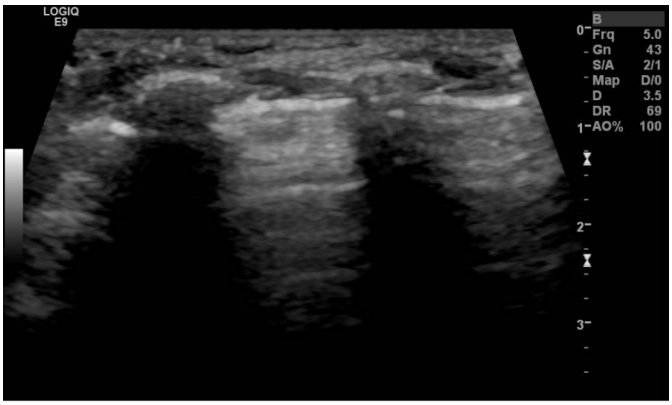
The A-lines are a group of parallel flat lines, occurring at regular distances below and in parallel with the pleural line. They represent a significant alteration in acoustic impedance at the pleuropulmonary line creating horizontal artifacts[25]. A-lines are echo artifacts reflected from the pleural line. They are visualized as hyperechoic, horizontal lines, occurring at equal spaces and extending deeply into the two-Dimensional image. The acoustic shadowing of the ribs represents an artifact arising from the ribs, shown by an anechoic area underneath the ribs and extending deeply into the two-dimensional image and disrupting the A-lines[26]. When the air content of the lung decreases as in subpleural interstitial edema, there will be an acoustic mismatch generated by the ultrasound wave between the fluid interface surrounded by air. This change will be reflected repeatedly at the deeper zones[21,27] and creates vertical artifacts called B-lines. These B-lines correlate with the pulmonary interstitial fluid content. The number of these lines increases with reducing the air content. B-lines or comet tail artifacts represent reverberation artifacts that are laser-like, hyperechoic, shadows that arise from this pleural line extending to the edge of the screen with coinciding movement with respiration. They can be caused by interstitial edema or interlobar septal pulmonary scarring[11,20]. The presence of multiple B-lines indicates alveolar interstitial edema[28,29]. Proof of compact coalesced B-lines in the lung denotes a serious form of the alveolar-interstitial syndrome, called "white lung". It is normal to visualize B-lines in healthy neonatal lungs. Their number will decrease with the baby's growth until being non-visualized at the age of 6 mo in a healthy infant[30,31]. Serial ultrasound imaging is advised to differentiate between standard B lines visualized during the neonatal period from pathological B-lines. If B lines increase, being more compact and coalesced, they will be more pathological. The denser the B-lines are, the more likely they are due to underlying lung pathology.

Lung sliding (Figure 3) represents the to-and-fro movement of the parietal and visceral pleura (pleural line) with respiratory movements and could be seen in B-mode and M-mode. Lung sliding visualized in B-mode is known as the movement of marching ants alongside the pleural line with respiration while, in M-mode, we can see lung sliding as the seashore sign in which the non-moving structures above the pleural line correspond to the sea, and the movement underneath the pleural line induces some irregularities simulating a sandy shore (Figure 4)[21,26]. Sometimes, the lung sliding is absent, which indicates a problem in the pleuropulmonary interface that can be observed in pneumothorax, complete atelectasis, pleuropulmonary pathology, and severe hyperinflation that could be seen in cases of foreign body aspiration[32]. Neonatal LUS scores provide a standardized approach to assess pulmonary pathology in the neonate, and evaluation of the disease progression is a semi-quantitative way[3,33-36]. Practically, the score of LUS is frequently assessed by six chest regions over the anterior and lateral zones of the chest. Early after birth, gravity plays a significant role, giving a slight distinction between the dependent and non-dependent lung zones[37]. For each zone, the score will range from 0 to 3. Thus, the total score will be between 0 and 18. Different neonatal pulmonary and pleural diseases have different numbers of B-lines and subpleural lung consolidations per each zone, which can help distinguish each of them[33]. A recent study proved that using more lung zones (10 or even 12 zones) in the first few days after birth did not result in better accuracy for diagnosis and management of



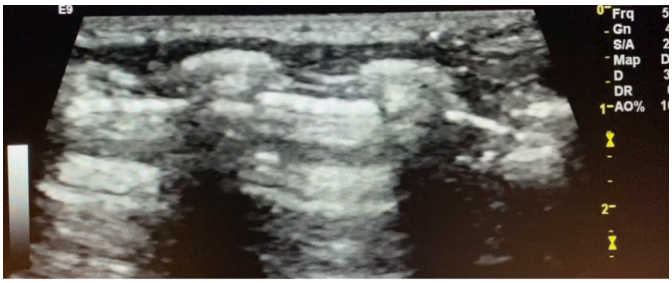
DOI: 10.35711/aimi.v3.i1.8 **Copyright** ©The Author(s) 2022.

Figure 1 Pleural line and A-lines in normal lung. The A-lines (red arrows) are horizontal artifactual repetitions of the pleural line (yellow lines) displayed at regular intervals.



DOI: 10.35711/aimi.v3.i1.8 **Copyright** ©The Author(s) 2022.

Figure 2 Bat sign created by the pleural line and ribs on either side. This view represents a normal lung surface, where the bright lateral line is the visceral and parietal interface, and the dark "bat wings" are rib shadows.



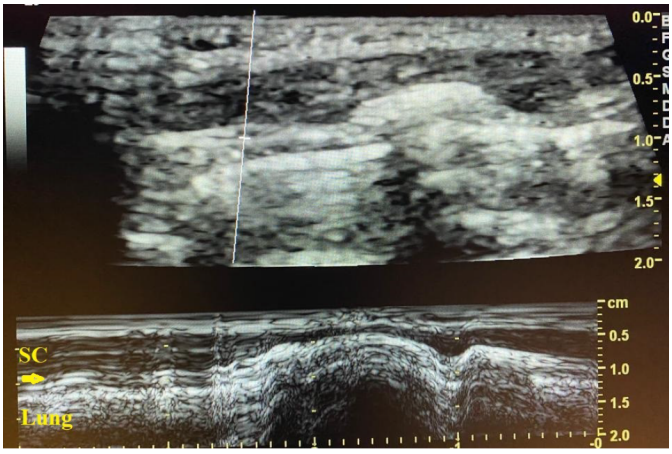
DOI: 10.35711/aimi.v3.i1.8 **Copyright** ©The Author(s) 2022.

Figure 3 Lung sliding and a shimmering appearance of the pleura. Lung sliding refers to a to-and-fro movement of the visceral pleura in contact with the parietal pleura due to shimmering/glimmering (or twinkling) of the pleural line on 2-Dimensional ultrasound.

bronchopulmonary dysplasia when compared to the standard six zones approach[38].

CLINICAL USES OF NEONATAL LUS

Neonatal LUS has a broad spectrum of clinical uses nowadays. The guidelines made by the POCUS working group of the European Society of Paediatric and Neonatal Intensive Care in 2020[16] stated that there was reasonable evidence (level B evidence) for neonatal LUS use in cases of transient tachypnoea of the newborn (TTN), respiratory distress syndrome (RDS), pneumothorax, and pleural effusions (with the advantage of guiding the thoracentesis). In some other diseases, the level of evidence was less (level



DOI: 10.35711/aimi.v3.i1.8

Figure 4 M-mode of the normal lung shows "Sand on the Beach" appearance or Seashore sign. The movement of the lung during respiration creates a speckled appearance like grains of sand (the shore) beneath the bright pleural line (Yellow arrow). In contrast, the soft tissues (Subcutaneous fat tissues) above the pleural line do not move with respiration and do not change with time and thus have a linear appearance (Sea appearance).

C), such as pulmonary edema and atelectasis. Different algorithms were suggested for neonatal LUS, *e.g.*, evaluation of life-threatening situations[3,39], the neonatal respiratory pathologies algorithm [20], the neonatal LUS protocol[40,41], and SAFE-R protocol (which also include assessment of cardiac tamponade, myocardial function, pleural effusion, and pneumothorax) in the decompensating neonate [39]. These algorithms require more controlled studies on many patients with different pathologies. Some limitations for using LUS in neonates will be discussed separately.

PLEURAL EFFUSION

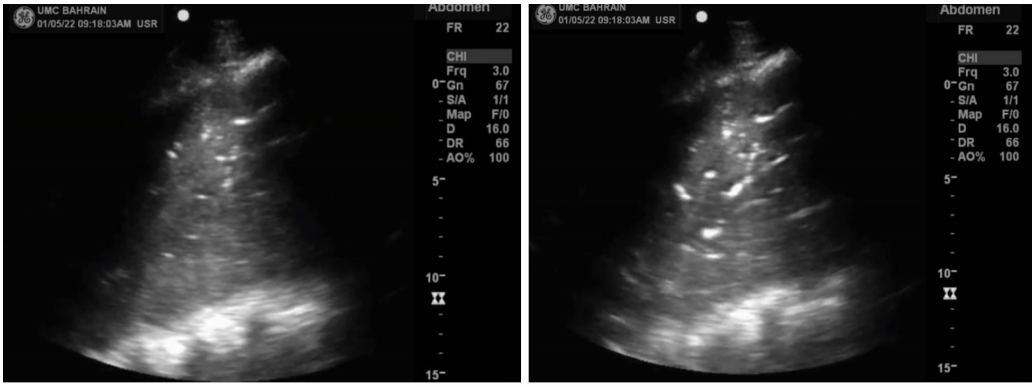
LUS in the neonate can detect even small volumes of pleural effusion very efficiently and can be used to guide pleural fluid aspiration[42]. In the B-mode, fluid is usually anechoic, sometimes with hepatization of the lung parenchyma. We can see the sinusoid sign in-M mode with the visceral line moving towards the pleural line during respiration. Colour doppler is not commonly used in these cases but can differentiate between echogenic and solid collections inside the effusion[21,26].

PNEUMONIA

Pneumonia is a severe neonatal disease that carries a high risk of morbidity and mortality, with about one million neonatal deaths yearly and about 10% of the worldwide child mortality[43]. Many pathogens are causing pneumonia in the neonates, such as bacteria, fungi, and viruses. Pneumonia can be acquired after birth or even during the intrauterine period[44]. The pathology includes epithelial injury of airways and alveoli, leakage of protein fluid (exudate), and interstitial edema of the alveoli. Clinical presentations are usually nonspecific and can be indistinguishable from RDS or TTN. Besides the laboratory workup, LUS can help in diagnosis. LUS in neonatal pneumonia cases shows pulmonary consolidation areas with irregular margins surrounding multiple B-lines. Other LUS findings that could present in pneumonia include an invisible pleural line on the affected part of the lung and absent lung sliding. Sometimes we can observe a dynamic air bronchogram, moving with respiration (Figure 5), especially in extensive areas of consolidation, which indicates the patency of airways (thus excluding atelectasis)[45]. In one study on forty cases of neonatal pneumonia *vs* forty neonates without pulmonary diseases, the authors found that LUS was a reliable method to diagnose neonatal pneumonia. They recommended routine use of LUS in the NICU[46]. A meta-analysis reviewed eight studies found that LUS has excellent sensitivity (96%) and specificity (93%) for the diagnosis of pneumonia in children, and the study recommended LUS as an alternative tool in such cases with no radiation exposure[47].

RDS

RDS or hyaline membrane disease is a significant reason for NICU admission and neonatal death. It primarily happens in preterm babies as about 70% of cases are seen in neonates born before 28 wk of



DOI: 10.35711/aimi.v3.i1.8

Figure 5 Dynamic air bronchogram.

pregnancy, and 15%-30% of cases occur in neonates 32-36 wk of gestation[48]. Pulmonary surfactant deficiency is significant in the pathogenesis of RDS. Type II pneumocytes produce pulmonary surfactants. One of their essential functions is to reduce the surface tension in the alveoli preventing the end-expiratory collapse of the alveoli, which requires more work of breathing to re-open in the next respiratory cycle. Affected patients present with respiratory distress and failure within 4-6 h postpartum and, in many cases, require mechanical ventilation[21,26]. LUS in RDS cases shows compact B-lines that coalesce together, giving the appearance of an echographic white lung, a thickened and irregular pleural line, and multiple areas of subpleural pulmonary consolidation (reflecting the presence of alveolar collapse). In one study, these ultrasonic features showed both sensitivity and specificity of 100% for RDS diagnosis[49]. In another study involving 59 neonates having clinical features suggestive of RDS, only 23 of them had actual RDS. In that study, the sensitivity of LUS was 95.6% (in comparison to 91.3 for chest X-ray), and the specificity was 94.4% (it was 84.2% in chest X-ray)[50]. LUS appearance of RDS is, sometimes, not symmetrical in the same or both lungs. Due to gravity issues, these features are usually found in the posterior parts of the chest because of the supine position acquired by the baby most of the time. So, it is crucial to examine the posterior chest in neonates not to miss these signs[51].

The treatment of choice in cases of RDS is the administration of surfactant and supported ventilation as needed. Neonatal LUS is able to expect the requirement for giving surfactant therapy and possibility of mechanical ventilation. One study showed that the presence of white lung signs in neonatal respiratory distress anticipated the need for intubation and mechanical ventilation with good sensitivity and specificity (88.9%, 100%, respectively)[52]. Another study showed that the lung ultrasound score in the first few hours after birth significantly correlates with the oxygenation condition (oxygen indices) in neonates and revealed adequate reliability to predict the requirement for surfactant therapy in premature infants[37]. Two more studies showed that the accuracy of LUS was higher than the fraction of inspired oxygen (FiO_2) in predicting the need for surfactant administration in premature babies[53,54]. A recently published trial showed a significant ability of LUS in predicting the need for surfactant compared to FiO_2 as a guide for that[55]. The concept of Echography-guided Surfactant THERapy, which uses the LUS score to direct surfactant therapy, resulted in an earlier intake of surfactant, which reduced the duration of invasive mechanical ventilation without any additional cost[56-58].

ATELECTASIS

Atelectasis is a collapse of a part of the lung parenchyma causing impairment of gas exchange. It can be caused by either airway obstruction, lung compression (by pulmonary or extrapulmonary lesion), or alveolar collapse due to increased surface tension of the alveolar wall. The most common mechanism of atelectasis in neonates is airway obstruction by thick mucus, meconium, or foreign particles. Atelectasis is usually associated with some other respiratory disorders[59]. Additionally, right upper lobe collapse in an intubated and mechanically ventilated baby can occur because of traumatic damage to the airway mucosa of the right-sided bronchi[60].

LUS can demonstrate atelectasis as an area of consolidation with the anechoic border and A-lines disruption[20,61]. Complete collapse leads to the absence of lung sliding and lung hepatization[11,61]. In severe atelectasis, lung pulse signs can be noticed in LUS, in which the collapsed part of the lung is pulsating with heartbeats[61]. Static air bronchogram can be observed with atelectasis, and this is different from dynamic air bronchograms (in pneumonia) that move with respiration, although differentiating between them is often challenging and requires an experienced sonographer[46,62]. Also, atelectasis in many cases is indistinguishable from pleural effusion in chest X-ray, but with LUS, it is easy to distinguish. One study showed that the sensitivity of LUS for diagnosing lung atelectasis was

100% *vs* 75% of chest X-rays (CT was the reference procedure in this study)[61]. Another study showed that the accuracy of LUS for diagnosis of post-anesthesia atelectasis in children was 88%, with a sensitivity of 89% and specificity of 88% (using magnetic resonance imaging as reference)[63].

PNEUMOTHORAX

The incidence of pneumothorax in neonates is about 1%-2%, but this rate is much more in neonates on mechanical ventilation, reaching up to 30%[64]. Tension pneumothorax is mainly encountered in neonates on mechanical ventilation either due to the original disease (as meconium aspiration or ball-valve obstruction of airways causing air trapping and rupture of alveoli) or due to iatrogenic causes such as birth trauma or improper suctioning techniques[65]. LUS signs of a pneumothorax include absent lung sliding, absent B-lines, and the existence of lung point. The absence of lung sliding and B-lines can be explained by accumulation of air in the pleural cavity, preventing the movement of the visceral pleura. It is worth noting that any disease that interrupts the visceral and parietal pleural interface will also cause absent lung sliding. The lung point sign is an area identified where parietal and visceral pleura separate[66]. This sign may be lacking in large tension pneumothorax[67,68].

Many studies showed the usefulness of LUS to detect pneumothorax. One study showed sensitivity and specificity to be 96.7% and 100%, respectively[69]. Another study showed the superiority of LUS over chest X-rays in diagnosing pneumothorax[66]. Another large multi-center study found that LUS is a safe and effective tool to identify serious pneumothorax and assist to manage chest drainage without doing chest X-rays. That study also showed that LUS has sensitivity, specificity, positive predictive value, and negative predictive value reaching up to 100% in diagnosing pneumothorax[8]. Another study compared three imaging techniques for the diagnosis of pneumothorax. It showed that LUS had 100% sensitivity and specificity, chest X-ray had 96% sensitivity and 100% specificity, while chest transillumination had 87% sensitivity and 96% specificity[67].

TTN

TTN, or the so-called "wet lung", is considered the most common reason of neonatal respiratory distress. TTN is usually a mild disorder, caused by a delay in the fetal lung fluid clearance (most of the fluid is removed by vaginal squeezing of the chest during labor, while the lymphatics system and pulmonary circulation clear the remaining fluid after being transported to lung interstitium)[70]. So, prematurity and elective cesarean sections are the main precipitating factors[72]. The condition usually resolves spontaneously within 24 h after birth but in a few cases may persist for several days. LUS can distinguish TTN from RDS by identifying B-lines' number and site[6,7]. In TTN cases, there are bilateral symmetric B-lines with a regular pleural line. Severe TTN presents as a white lung. LUS has high specificity but low sensitivity for the diagnosis of TTN. The double lung point sign represents the area between the upper and lower lung zones at which we can distinguish spaced-out B-lines next to confluent B-lines. So, double lung point can be considered a demarcation point of echogenic differences in the lung field[6,7,68].

The double lung point additionally occurs during the diseases recovery phase, such as severe TTN, RDS, and pneumonia[6] Sometimes a mixed RDS/TTN pattern can be identified when the baby has reduced reabsorption of the lung fluid and relative surfactant deficiency. This pattern can be recognized using the LUS score[72]. One prospective cohort study on 59 neonates with respiratory distress found that sensitivity and specificity of LUS for TTN diagnosis were 93.3% and 96.5%, respectively. These values were better than those observed in chest X-rays (89.4% and 91.3%, respectively)[50]. Another recent meta-analysis concluded that LUS has excellent specificity and sensitivity for diagnosing TTN [73]. Studies also showed that LUS could diagnose TTN and predict which neonate may need a higher level of care[74].

BRONCHOPULMONARY DYSPLASIA

Bronchopulmonary dysplasia (BPD) is a common complication related to prematurity and is one of the common complications of RDS. BPD is associated with required respiratory support and/or oxygen supplement at 36 wk corrected gestation. It associates with long-term morbidity and even mortality in some cases[75]. In BPD, structural lung abnormalities, immature biochemical pathways, and oxidant injuries are associated with repeated pulmonary infections and poor nutrition, leading to impaired cardiopulmonary function[76]. LUS features of BPD include thickened coarse pleural lines, subpleural consolidations, and B-lines. According to the severity of inter-lobar septal scarring and interstitial edema, B-lines can be scattered or diffuse. LUS score can help diagnose BPD severity[77] and guide the management, including diuretics use[36].

LUS score can predict the development of BPD in some studies. In a multi-center cohort study, authors found that LUS score on day seven and day fourteen correlates with the oxygenation indices and predicts BPD occurrence when adjusted for gestation and sex[38]. In another cohort study, LUS was done on days 3, 7, and 14 in neonates born before 29 wk gestation[78]. This study showed that the LUS score was higher in neonates who later developed BPD on all-time points, with an LUS score of more than ten on day seven having the highest sensitivity and specificity.

MAS

MAS is due to intra-uterine aspiration of meconium-contaminated amniotic fluid into the newborn airways due to fetal hypoxia, acidosis, or infection[79]. Meconium obstructs the airways and induces surfactant dysfunction, chemical pneumonitis, and secondary infection. These will lead to hypoxia due to ventilation/perfusion mismatch[80]. Neonates with MAS have yellowish greenish (meconium stained) skin, umbilical cord, and nails, and signs of respiratory distress. It may develop immediately after birth. MAS is a specific type of pneumonia. So, its LUS features are like pneumonia, giving the features of irregular subpleural consolidations with coalescent B-lines. These features are usually unilateral[81]. Some studies showed the usefulness of LUS for diagnosing MAS in neonates[9,82]. However, LUS should be correlated to the clinical circumstances and physical examination.

Table 1 summarizes lung ultrasound appearance in different neonatal lung diseases compared to chest X-rays.

OTHER USES OF LUNG ULTRASOUND IN NEONATES

LUS can be used to assess lung recruitment with positive end-expiratory pressure without the need for exposure to ionizing radiation by doing CT chest[83]. LUS can also effectively monitor bronchoalveolar lavage in neonates with atelectasis, with an efficacy approaching 93%[84]. Another application of interest that was seen in some studies is the use of LUS to assess the position of the ETT in the trachea by measuring the space between the ETT distal end and the aortic arch apex[85] or the space between ETT distal end and the superior edge of the right pulmonary artery[86]. We can achieve this technique by utilising either a phase array probe (while doing the high parasternal view) or a linear probe (in the midsagittal view). Another study reported the use of LUS to immediately confirm the proper ETT position during neonatal resuscitation. This study used a linear probe in the transverse position[87].

Another critical use of ultrasounds is evaluation of vocal cord function. One study displayed that utilising high-frequency linear hockey stick probe in a transverse position over the middle of the neck could identify the presence of vocal cord paresis post-operatively (after aortic arch repair) with high sensitivity and specificity is compared to flexible fibreoptic endoscopy[88]. LUS has also been utilized to evaluate the diaphragm[89,90]. A recent study used LUS and diaphragmatic shortening fraction, a known way of assessing adult diaphragm function, to evaluate diaphragm in neonates. This study found that the diaphragmatic shortening fraction could be assessed in neonates[91]. LUS has also been suggested as a modality to follow asymptomatic CPAMs, but more studies are needed to stabilize this indication[92,93].

LIMITATIONS OF LUNG ULTRASOUND USE FOR NEONATAL RESPIRATORY PROBLEMS

Although LUS is a very effective and safe imaging technique in neonates, we should consider the clinical finding of each case. Moreover, according to the application of LUS, and some problems in actual clinical practice, LUS has some limitations in some pulmonary conditions. For example, as mentioned above, the diagnosis of CPAMs using LUS is still not standardized, and many studies must be done in this context. Some cases of CPAMs can be detected in utero using ultrasound as the fetal lung is filled with fluid. On the contrary, due to air-filled neonatal lungs, their diagnosis by LUS in the neonatal period seems to be difficult because these lesions are usually away from the chest wall. Thus, lesions that are located away from the pleura could not be visualized by LUS[92].

LUS cannot identify some specific lesions because of the influence of gas in front of the lesion. When the acoustic beam of ultrasound encounters gas, it will be reflected ultimately. So, cases of pulmonary bullae cannot be visualized by LUS because of the large amount of gas in the bulla reflecting the acoustic beam of ultrasound. Similarly, the presence of subcutaneous emphysema or pneumomediastinum will affect the results of LUS due to the same reasons described above. Although LUS is a handy tool to diagnose pneumothorax, it cannot measure the size due to the total reflection caused by the gas[66].

Table 1 Lung ultrasound appearance in different neonatal lung diseases compared to chest X-ray

Disease	Chest X-ray	Lung ultrasound
Pleural effusion	Homogenous opacity obliterating costophrenic and cardiophrenic angles	B-mode: Fluid is anechoic, sometimes ± hepatization of the lung parenchyma. M-mode: The sinusoid sign with the visceral line moving towards the pleural line during respiration
Pneumonia	Homogeneous opacities that can be patchy or lobar in distribution	Consolidation areas with irregular margins surrounding multiple B-lines. Invisible pleural line on the affected area. Sometimes: Dynamic air bronchogram
RDS	Alveolar shadowing (ground glass) with air bronchogram	Compact coalescent B-lines (white lung). Thickened, irregular pleural line. Multiple areas of sub-pleural consolidation
Atelectasis	Area of opacity in the lung with features of volume loss as shifting of mediastinum to the same side, pulled fissure, *etc.*	Area of consolidation with anechoic clear border and disrupted A-lines. Static air bronchogram. Complete collapse leads to the absence of lung sliding, lung hepatization, and lung pulse signs
Pneumothorax	Jet black translucency with collapsed lung and sometimes mediastinal shift to the other side	Absent lung sliding, absent B-lines, and the presence of lung point
TTN	Interstitial oedema predominantly in the peri-hilar region (wet silhouette)	Double lung point sign. B-lines. In severe cases: (white lung)
BPD	Ill-defined diffuse reticular markings with circular lucent areas in between and hyperinflated lung	Thickened coarse pleural linesSubpleural areas of consolidation. B-lines
MAS	Patchy consolidation	Same as pneumonia

BPD: Bronchopulmonary dysplasia; MAS: Meconium aspiration syndrome; RDS: respiratory distress syndrome; TTN: Transient tachypnoea of the newborn.

Consequently, we need more studies to quantify the size of pneumothorax using LUS. Pulmonary interstitial emphysema is another condition that LUS cannot diagnose. In a published case study, the authors used LUS to follow-up localized interstitial emphysema. The infant presented again with tachypnoea after being treated with continuous positive airway pressure for three days. The chest computed tomography revealed localized interstitial emphysema of the left upper lobe, whereas LUS did not show this lesion[94]. We emphasized that using LUS is potentially harmful without adequate expertise. It may not provide definite diagnostic information and may allow over trust in the procedure, which could have profound legal implications and not address the underlying lesions. The misuse of artifacts as a diagnostic tool should be abandoned. Lung ultrasound imaging is advantageous when definite imaging is possible, even in the newborn.

CONCLUSION

Lung ultrasound is a valuable imaging tool frequently used in neonatal respiratory care. It helps diagnose many respiratory disorders with excellent accuracy and safety with no radiation risk. Lung ultrasound is operator dependent and needs adequate experience to achieve good results. Some limitations are encountered for its use, but its benefits are more.

ACKNOWLEDGEMENTS

We thank the anonymous referees for their useful suggestions.

FOOTNOTES

Author contributions: Bediwy AS, Al-Biltagi M, Nazeer JA, and Saeed NK collected the data and wrote and revised the manuscript.

Conflict-of-interest statement: The authors declare no conflicts of interest for this article.

Country/Territory of origin: Egypt

ORCID number: Adel Salah Bediwy 0000-0002-0281-0010; Mohammed Al-Biltagi 0000-0002-7761-9536; Jameel Ahmed Nazeer 0000-0002-0085-9822; Nermin Kamal Saeed 0000-0001-7875-8207.

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

REFERENCES

1 **Angus DC**, Linde-Zwirble WT, Clermont G, Griffin MF, Clark RH. Epidemiology of neonatal respiratory failure in the United States: projections from California and New York. *Am J Respir Crit Care Med* 2001; **164**: 1154-1160 [PMID: 11673202 DOI: 10.1164/ajrccm.164.7.2012126]
2 **Qian L**, Liu C, Zhuang W, Guo Y, Yu J, Chen H, Wang S, Lin Z, Xia S, Ni L, Liu X, Chen C, Sun B; Chinese Collaborative Study Group for Neonatal Respiratory Diseases. Neonatal respiratory failure: a 12-month clinical epidemiologic study from 2004 to 2005 in China. *Pediatrics* 2008; **121**: e1115-e1124 [PMID: 18450855 DOI: 10.1542/peds.2006-2426]
3 **Raimondi F**, Yousef N, Migliaro F, Capasso L, De Luca D. Point-of-care lung ultrasound in neonatology: classification into descriptive and functional applications. *Pediatr Res* 2021; **90**: 524-531 [PMID: 30127522 DOI: 10.1038/s41390-018-0114-9]
4 **Mongodi S**, Santangelo E, De Luca D, Rovida S, Corradi F, Volpicelli G, Gargani L, Bouhemad B, Mojoli F. Quantitative Lung Ultrasound: Time for a Consensus? *Chest* 2020; **158**: 469-470 [PMID: 32768066 DOI: 10.1016/j.chest.2020.03.080]
5 **Avni EF**, Braude P, Pardou A, Matos C. Hyaline membrane disease in the newborn: diagnosis by ultrasound. *Pediatr Radiol* 1990; **20**: 143-146 [PMID: 2191263 DOI: 10.1007/BF02012957]
6 **Liu J**, Chen XX, Li XW, Chen SW, Wang Y, Fu W. Lung Ultrasonography to Diagnose Transient Tachypnea of the Newborn. *Chest* 2016; **149**: 1269-1275 [PMID: 26836942 DOI: 10.1016/j.chest.2015.12.024]
7 **Raimondi F**, Yousef N, Rodriguez Fanjul J, De Luca D, Corsini I, Shankar-Aguilera S, Dani C, Di Guardo V, Lama S, Mosca F, Migliaro F, Sodano A, Vallone G, Capasso L. A Multicenter Lung Ultrasound Study on Transient Tachypnea of the Neonate. *Neonatology* 2019; **115**: 263-268 [PMID: 30731475 DOI: 10.1159/000495911]
8 **Raimondi F**, Rodriguez Fanjul J, Aversa S, Chirico G, Yousef N, De Luca D, Corsini I, Dani C, Grappone L, Orfeo L, Migliaro F, Vallone G, Capasso L; Lung Ultrasound in the Crashing Infant (LUCI) Protocol Study Group. Lung Ultrasound for Diagnosing Pneumothorax in the Critically Ill Neonate. *J Pediatr* 2016; **175**: 74-78.e1 [PMID: 27189678 DOI: 10.1016/j.jpeds.2016.04.018]
9 **Piastra M**, Yousef N, Brat R, Manzoni P, Mokhtari M, De Luca D. Lung ultrasound findings in meconium aspiration syndrome. *Early Hum Dev* 2014; **90** Suppl 2: S41-S43 [PMID: 25220126 DOI: 10.1016/S0378-3782(14)50011-4]
10 **Corsini I**, Parri N, Gozzini E, Coviello C, Leonardi V, Poggi C, Giacalone M, Bianconi T, Tofani L, Raimondi F, Dani C. Lung Ultrasound for the Differential Diagnosis of Respiratory Distress in Neonates. *Neonatology* 2019; **115**: 77-84 [PMID: 30304736 DOI: 10.1159/000493001]
11 **Ammirabile A**, Buonsenso D, Di Mauro A. Lung Ultrasound in Pediatrics and Neonatology: An Update. *Healthcare (Basel)* 2021; **9** [PMID: 34442152 DOI: 10.3390/healthcare9081015]
12 **Escourrou G**, De Luca D. Lung ultrasound decreased radiation exposure in preterm infants in a neonatal intensive care unit. *Acta Paediatr* 2016; **105**: e237-e239 [PMID: 26880491 DOI: 10.1111/apa.13369]
13 **Okbay Gunes A**, Karadag N, Cakir H, Hakyemez Toptan H, Karatekin G. The Associations Between Lung Ultrasonography Scores in the First Day of Life and Clinical Outcomes: Authors' Reply. *J Ultrasound Med* 2021 [PMID: 34196035 DOI: 10.1002/jum.15771]
14 **Van de Voorde P**, Turner NM, Djakow J, de Lucas N, Martinez-Mejias A, Biarent D, Bingham R, Brissaud O, Hoffmann F, Johannesdottir GB, Lauritsen T, Maconochie I. European Resuscitation Council Guidelines 2021: Paediatric Life Support. *Resuscitation* 2021; **161**: 327-387 [PMID: 33773830 DOI: 10.1016/j.resuscitation.2021.02.015]
15 **Hall EJ**. Lessons we have learned from our children: cancer risks from diagnostic radiology. *Pediatr Radiol* 2002; **32**: 700-706 [PMID: 12244457 DOI: 10.1007/s00247-002-0774-8]
16 **Singh Y**, Tissot C, Fraga MV, Yousef N, Cortes RG, Lopez J, Sanchez-de-Toledo J, Brierley J, Colunga JM, Raffaj D, Da Cruz E, Durand P, Kenderessy P, Lang HJ, Nishisaki A, Kneyber MC, Tissieres P, Conlon TW, De Luca D. International evidence-based guidelines on Point of Care Ultrasound (POCUS) for critically ill neonates and children issued by the POCUS Working Group of the European Society of Paediatric and Neonatal Intensive Care (ESPNIC). *Crit Care* 2020; **24**: 65 [PMID: 32093763 DOI: 10.1186/s13054-020-2787-9]
17 Retraction. Immunization with pseudotype baculovirus expressing envelope protein of Japanese encephalitis virus elicits protective immunity in mice by Yaoming Li, Jing Ye, Shengbo Cao, Shaobo Xiao, Qian Zhao, Xueqin Liu, Meilin Jin, Huanchun Chen. *J Gene Med* 2009; **11**: 454 [PMID: 19382283 DOI: 10.1002/jgm.1330]
18 **Schiller NB**, Shah PM, Crawford M, DeMaria A, Devereux R, Feigenbaum H, Gutgesell H, Reichek N, Sahn D, Schnittger I. Recommendations for quantitation of the left ventricle by two-dimensional echocardiography. American Society of Echocardiography Committee on Standards, Subcommittee on Quantitation of Two-Dimensional Echocardiograms. *J Am Soc Echocardiogr* 1989; **2**: 358-367 [PMID: 2698218 DOI: 10.1016/s0894-7317(89)80014-8]
19 **Gargani L**. Lung ultrasound: a new tool for the cardiologist. *Cardiovasc Ultrasound* 2011; **9**: 6 [PMID: 21352576 DOI: 10.1186/1476-7120-9-6]
20 **Kurepa D**, Zaghloul N, Watkins L, Liu J. Neonatal lung ultrasound exam guidelines. *J Perinatol* 2018; **38**: 11-22 [PMID:

29144490 DOI: 10.1038/jp.2017.140]
21 **Liang HY**, Liang XW, Chen ZY, Tan XH, Yang HH, Liao JY, Cai K, Yu JS. Ultrasound in neonatal lung disease. *Quant Imaging Med Surg* 2018; **8**: 535-546 [PMID: 30050788 DOI: 10.21037/qims.2018.06.01]
22 **Gomond-Le Goff C**, Vivalda L, Foligno S, Loi B, Yousef N, De Luca D. Effect of Different Probes and Expertise on the Interpretation Reliability of Point-of-Care Lung Ultrasound. *Chest* 2020; **157**: 924-931 [PMID: 31785252 DOI: 10.1016/j.chest.2019.11.013]
23 **Demi L**, van Hoeve W, van Sloun RJG, Soldati G, Demi M. Determination of a potential quantitative measure of the state of the lung using lung ultrasound spectroscopy. *Sci Rep* 2017; **7**: 12746 [PMID: 28986558 DOI: 10.1038/s41598-017-13078-9]
24 **Blank DA**, Rogerson SR, Kamlin COF, Fox LM, Lorenz L, Kane SC, Polglase GR, Hooper SB, Davis PG. Lung ultrasound during the initiation of breathing in healthy term and late preterm infants immediately after birth, a prospective, observational study. *Resuscitation* 2017; **114**: 59-65 [PMID: 28249708 DOI: 10.1016/j.resuscitation.2017.02.017]
25 **Lichtenstein DA**, Mezière GA, Lagoueyte JF, Biderman P, Goldstein I, Gepner A. A-lines and B-lines: lung ultrasound as a bedside tool for predicting pulmonary artery occlusion pressure in the critically ill. *Chest* 2009; **136**: 1014-1020 [PMID: 19809049 DOI: 10.1378/chest.09-0001]
26 **Ruoss JL**, Bazacliu C, Cacho N, De Luca D. Lung Ultrasound in the Neonatal Intensive Care Unit: Does It Impact Clinical Care? *Children (Basel)* 2021; **8** [PMID: 34943297 DOI: 10.3390/children8121098]
27 **Jambrik Z**, Monti S, Coppola V, Agricola E, Mottola G, Miniati M, Picano E. Usefulness of ultrasound lung comets as a nonradiologic sign of extravascular lung water. *Am J Cardiol* 2004; **93**: 1265-1270 [PMID: 15135701 DOI: 10.1016/j.amjcard.2004.02.012]
28 **Lichtenstein D**, Mézière G, Biderman P, Gepner A, Barré O. The comet-tail artifact. An ultrasound sign of alveolar-interstitial syndrome. *Am J Respir Crit Care Med* 1997; **156**: 1640-1646 [PMID: 9372688 DOI: 10.1164/ajrccm.156.5.96-07096]
29 **Volpicelli G**, Mussa A, Garofalo G, Cardinale L, Casoli G, Perotto F, Fava C, Frascisco M. Bedside lung ultrasound in the assessment of alveolar-interstitial syndrome. *Am J Emerg Med* 2006; **24**: 689-696 [PMID: 16984837 DOI: 10.1016/j.ajem.2006.02.013]
30 **Aichhorn L**, Küng E, Habrina L, Werther T, Berger A, Urlesberger B, Schwaberger B. The Role of Lung Ultrasound in the Management of the Critically Ill Neonate-A Narrative Review and Practical Guide. *Children (Basel)* 2021; **8** [PMID: 34438519 DOI: 10.3390/children8080628]
31 **Buonsenso D**, Soldati G, Curatola A, Morello R, De Rose C, Vacca ME, Lazzareschi I, Musolino AM, Valentini P. Lung Ultrasound Pattern in Healthy Infants During the First 6 Months of Life. *J Ultrasound Med* 2020; **39**: 2379-2388 [PMID: 32468627 DOI: 10.1002/jum.15347]
32 **Lovrenski J**, Vilotijević Dautović G, Lovrenski A. Reduced or Absent "Lung Sliding" - A Novel Lung Ultrasound Sign of Pediatric Foreign Body Aspiration. *J Ultrasound Med* 2019; **38**: 3079-3082 [PMID: 30892735 DOI: 10.1002/jum.14988]
33 **Raimondi F**, Migliaro F, Corsini I, Meneghin F, Dolce P, Pierri L, Perri A, Aversa S, Nobile S, Lama S, Varano S, Savoia M, Gatto S, Leonardi V, Capasso L, Carnielli VP, Mosca F, Dani C, Vento G, Lista G. Lung Ultrasound Score Progress in Neonatal Respiratory Distress Syndrome. *Pediatrics* 2021; **147** [PMID: 33688032 DOI: 10.1542/peds.2020-030528]
34 **Oulego-Erroz I**, Alonso-Quintela P, Terroba-Seara S, Jiménez-González A, Rodríguez-Blanco S. Early assessment of lung aeration using an ultrasound score as a biomarker of developing bronchopulmonary dysplasia: a prospective observational study. *J Perinatol* 2021; **41**: 62-68 [PMID: 32665687 DOI: 10.1038/s41372-020-0724-z]
35 **Alonso-Ojembarrena A**, Serna-Guerediaga I, Aldecoa-Bilbao V, Gregorio-Hernández R, Alonso-Quintela P, Concheiro-Guisán A, Ramos-Rodríguez A, de Las Heras-Martín M, Rodeño-Fernández L, Oulego-Erroz I. The Predictive Value of Lung Ultrasound Scores in Developing Bronchopulmonary Dysplasia: A Prospective Multicenter Diagnostic Accuracy Study. *Chest* 2021; **160**: 1006-1016 [PMID: 33689782 DOI: 10.1016/j.chest.2021.02.066]
36 **Alonso-Ojembarrena A**, Lechuga-Sancho AM, Morales-Arandojo P, Acuñas-Soto S, López-de-Francisco R, Lubián-López SP. Lung ultrasound score and diuretics in preterm infants born before 32 weeks: A pilot study. *Pediatr Pulmonol* 2020; **55**: 3312-3318 [PMID: 32986302 DOI: 10.1002/ppul.25098]
37 **Brat R**, Yousef N, Klifa R, Reynaud S, Shankar Aguilera S, De Luca D. Lung Ultrasonography Score to Evaluate Oxygenation and Surfactant Need in Neonates Treated With Continuous Positive Airway Pressure. *JAMA Pediatr* 2015; **169**: e151797 [PMID: 26237465 DOI: 10.1001/jamapediatrics.2015.1797]
38 **Loi B**, Vigo G, Baraldi E, Raimondi F, Carnielli VP, Mosca F, De Luca D. Lung Ultrasound to Monitor Extremely Preterm Infants and Predict Bronchopulmonary Dysplasia. A Multicenter Longitudinal Cohort Study. *Am J Respir Crit Care Med* 2021; **203**: 1398-1409 [PMID: 33352083 DOI: 10.1164/rccm.202008-3131OC]
39 **Yousef N**, Singh Y, De Luca D. "Playing it SAFE in the NICU" SAFE-R: a targeted diagnostic ultrasound protocol for the suddenly decompensating infant in the NICU. *Eur J Pediatr* 2022; **181**: 393-398 [PMID: 34223967 DOI: 10.1007/s00431-021-04186-w]
40 **Liu J**, Copetti R, Sorantin E, Lovrenski J, Rodriguez-Fanjul J, Kurepa D, Feng X, Cattaross L, Zhang H, Hwang M, Yeh TF, Lipener Y, Lodha A, Wang JQ, Cao HY, Hu CB, Lyu GR, Qiu XR, Jia LQ, Wang XM, Ren XL, Guo JY, Gao YQ, Li JJ, Liu Y, Fu W, Wang Y, Lu ZL, Wang HW, Shang LL. Protocol and Guidelines for Point-of-Care Lung Ultrasound in Diagnosing Neonatal Pulmonary Diseases Based on International Expert Consensus. *J Vis Exp* 2019 [PMID: 30907892 DOI: 10.3791/58990]
41 **Liu J**, Guo G, Kurepa D, Volpicelli G, Sorantin E, Lovrenski J, Alonso-Ojembarrena A, Hsieh KS, Lodha A, Yeh TF, Jagła M, Shah H, Yan W, Hu CB, Zhou XG, Guo RJ, Cao HY, Wang Y, Zong HF, Shang LL, Ma HR, Liu Y, Fu W, Shan RY, Qiu RX, Ren XL, Copetti R, Rodriguez-Fanjul J, Feletti F; Society of Pediatrics, Asia-Pacific Health Association; the Division of Critical Ultrasound, Pediatric Society of Asia-Pacific Health Association; the Critical Ultrasound Group of Neonatal Specialty Committee, the Cross-Straits Medicine Exchange Association as well as the World Interactive Network Focused On Critical Ultrasound China Branch. Specification and guideline for technical aspects and scanning parameter settings of neonatal lung ultrasound examination. *J Matern Fetal Neonatal Med* 2022; **35**: 1003-1016 [PMID: 34182870 DOI: 10.1080/14767058.2021.1940943]

42 **Soni NJ**, Franco R, Velez MI, Schnobrich D, Dancel R, Restrepo MI, Mayo PH. Ultrasound in the diagnosis and management of pleural effusions. *J Hosp Med* 2015; **10**: 811-816 [PMID: 26218493 DOI: 10.1002/jhm.2434]

43 **Duke T**. Neonatal pneumonia in developing countries. *Arch Dis Child Fetal Neonatal Ed* 2005; **90**: F211-F219 [PMID: 15846010 DOI: 10.1136/adc.2003.048108]

44 **Nissen MD**. Congenital and neonatal pneumonia. *Paediatr Respir Rev* 2007; **8**: 195-203 [PMID: 17868917 DOI: 10.1016/j.prrv.2007.07.001]

45 **Copetti R**, Cattarossi L. Ultrasound diagnosis of pneumonia in children. *Radiol Med* 2008; **113**: 190-198 [PMID: 18386121 DOI: 10.1007/s11547-008-0247-8]

46 **Liu J**, Liu F, Liu Y, Wang HW, Feng ZC. Lung ultrasonography for the diagnosis of severe neonatal pneumonia. *Chest* 2014; **146**: 383-388 [PMID: 24833216 DOI: 10.1378/chest.13-2852]

47 **Pereda MA**, Chavez MA, Hooper-Miele CC, Gilman RH, Steinhoff MC, Ellington LE, Gross M, Price C, Tielsch JM, Checkley W. Lung ultrasound for the diagnosis of pneumonia in children: a meta-analysis. *Pediatrics* 2015; **135**: 714-722 [PMID: 25780071 DOI: 10.1542/peds.2014-2833]

48 **Hjalmarson O**. Epidemiology and classification of acute, neonatal respiratory disorders. A prospective study. *Acta Paediatr Scand* 1981; **70**: 773-783 [PMID: 7324931 DOI: 10.1111/j.1651-2227.1981.tb06228.x]

49 **Copetti R**, Cattarossi L, Macagno F, Violino M, Furlan R. Lung ultrasound in respiratory distress syndrome: a useful tool for early diagnosis. *Neonatology* 2008; **94**: 52-59 [PMID: 18196931 DOI: 10.1159/000113059]

50 **Vergine M**, Copetti R, Brusa G, Cattarossi L. Lung ultrasound accuracy in respiratory distress syndrome and transient tachypnea of the newborn. *Neonatology* 2014; **106**: 87-93 [PMID: 24819542 DOI: 10.1159/000358227]

51 **Lovrenski J**. Lung ultrasonography of pulmonary complications in preterm infants with respiratory distress syndrome. *Ups J Med Sci* 2012; **117**: 10-17 [PMID: 22283442 DOI: 10.3109/03009734.2011.643510]

52 **Raimondi F**, Migliaro F, Sodano A, Ferrara T, Lama S, Vallone G, Capasso L. Use of neonatal chest ultrasound to predict noninvasive ventilation failure. *Pediatrics* 2014; **134**: e1089-e1094 [PMID: 25180278 DOI: 10.1542/peds.2013-3924]

53 **Badurdeen S**, Kamlin COF, Rogerson SR, Kane SC, Polglase GR, Hooper SB, Davis PG, Blank DA. Lung ultrasound during newborn resuscitation predicts the need for surfactant therapy in very- and extremely preterm infants. *Resuscitation* 2021; **162**: 227-235 [PMID: 33548362 DOI: 10.1016/j.resuscitation.2021.01.025]

54 **De Martino L**, Yousef N, Ben-Ammar R, Raimondi F, Shankar-Aguilera S, De Luca D. Lung Ultrasound Score Predicts Surfactant Need in Extremely Preterm Neonates. *Pediatrics* 2018; **142** [PMID: 30108142 DOI: 10.1542/peds.2018-0463]

55 **Raimondi F**, Migliaro F, Corsini I, Meneghin F, Pierri L, Salomè S, Perri A, Aversa S, Nobile S, Lama S, Varano S, Savoia M, Gatto S, Leonardi V, Capasso L, Carnielli VP, Mosca F, Dani C, Vento G, Dolce P, Lista G. Neonatal Lung Ultrasound and Surfactant Administration: A Pragmatic, Multicenter Study. *Chest* 2021; **160**: 2178-2186 [PMID: 34293317 DOI: 10.1016/j.chest.2021.06.076]

56 **De Luca D**, Autilio C, Pezza L, Shankar-Aguilera S, Tingay DG, Carnielli VP. Personalized Medicine for the Management of RDS in Preterm Neonates. *Neonatology* 2021; **118**: 127-138 [PMID: 33735866 DOI: 10.1159/000513783]

57 **Spila-Alegiani S**, Da Cas R, Giambi C, Raschetti R, Salmaso S. [Human papillomavirus vaccine register]. *Recenti Prog Med* 2013; **104**: 262-266 [PMID: 23801230 DOI: 10.1701/1295.14327]

58 **De Luca D**, Yousef N. Pharmaceutical Expenditure Is Unchanged with Ultrasound-Guided Surfactant Administration. *Am J Perinatol* 2020 [PMID: 32819020 DOI: 10.1055/s-0040-1715821]

59 **Peroni DG**, Boner AL. Atelectasis: mechanisms, diagnosis and management. *Paediatr Respir Rev* 2000; **1**: 274-278 [PMID: 12531090 DOI: 10.1053/prrv.2000.0059]

60 **Whitfield JM**, Jones MD Jr. Atelectasis associated with mechanical ventilation for hyaline membrane disease. *Crit Care Med* 1980; **8**: 729-731 [PMID: 7449403 DOI: 10.1097/00003246-198012000-00006]

61 **Liu J**, Chen SW, Liu F, Li QP, Kong XY, Feng ZC. The diagnosis of neonatal pulmonary atelectasis using lung ultrasonography. *Chest* 2015; **147**: 1013-1019 [PMID: 25341049 DOI: 10.1378/chest.14-1306]

62 **Öktem A**, Zenciroğlu A, Üner Ç, Aydoğan S, Dilli D, Okumuş N. Efficiency of Lung Ultrasonography in the Diagnosis and Follow-up of Viral Pneumonia in Newborn. *Am J Perinatol* 2021 [PMID: 34044459 DOI: 10.1055/s-0041-1729880]

63 **Acosta CM**, Maidana GA, Jacovitti D, Belaunzarán A, Cereceda S, Rae E, Molina A, Gonorazky S, Bohm SH, Tusman G. Accuracy of transthoracic lung ultrasound for diagnosing anesthesia-induced atelectasis in children. *Anesthesiology* 2014; **120**: 1370-1379 [PMID: 24662376 DOI: 10.1097/ALN.0000000000000231]

64 **Apiliogullari B**, Sunam GS, Ceran S, Koc H. Evaluation of neonatal pneumothorax. *J Int Med Res* 2011; **39**: 2436-2440 [PMID: 22289564 DOI: 10.1177/147323001103900645]

65 **Wyatt TH**. Pneumothorax in the neonate. *J Obstet Gynecol Neonatal Nurs* 1995; **24**: 211-216 [PMID: 7782953 DOI: 10.1111/j.1552-6909.1995.tb02465.x]

66 **Fei Q**, Lin Y, Yuan TM. Lung Ultrasound, a Better Choice for Neonatal Pneumothorax: A Systematic Review and Meta-analysis. *Ultrasound Med Biol* 2021; **47**: 359-369 [PMID: 33341304 DOI: 10.1016/j.ultrasmedbio.2020.11.011]

67 **Cattarossi L**, Copetti R, Brusa G, Pintaldi S. Lung Ultrasound Diagnostic Accuracy in Neonatal Pneumothorax. *Can Respir J* 2016; **2016**: 6515069 [PMID: 27445558 DOI: 10.1155/2016/6515069]

68 **Copetti R**, Cattarossi L. The 'double lung point': an ultrasound sign diagnostic of transient tachypnea of the newborn. *Neonatology* 2007; **91**: 203-209 [PMID: 17377407 DOI: 10.1159/000097454]

69 **Dahmarde H**, Parooie F, Salarzaei M. Accuracy of Ultrasound in Diagnosis of Pneumothorax: A Comparison between Neonates and Adults-A Systematic Review and Meta-Analysis. *Can Respir J* 2019; **2019**: 5271982 [PMID: 31933707 DOI: 10.1155/2019/5271982]

70 **Yurdakök M**. Transient tachypnea of the newborn: what is new? *J Matern Fetal Neonatal Med* 2010; **23** Suppl 3: 24-26 [PMID: 20807157 DOI: 10.3109/14767058.2010.507971]

71 **Guglani L**, Lakshminrusimha S, Ryan RM. Transient tachypnea of the newborn. *Pediatr Rev* 2008; **29**: e59-e65 [PMID: 18977854 DOI: 10.1542/pir.29-11-e59]

72 **Machado AL**, Bochio BC, Wady AF, Jorge JH, Canevarolo SV Jr, Vergani CE. Impact strength of denture base and reline acrylic resins: An in vitro study. *J Dent Biomech* 2012; **3**: 1758736012459535 [PMID: 22977461 DOI:

10.1177/1758736012459535]
73 **He L**, Sun Y, Sheng W, Yao Q. Diagnostic performance of lung ultrasound for transient tachypnea of the newborn: A meta-analysis. *PLoS One* 2021; **16**: e0248827 [PMID: 33780485 DOI: 10.1371/journal.pone.0248827]
74 **Raimondi F**, Migliaro F, Sodano A, Umbaldo A, Romano A, Vallone G, Capasso L. Can neonatal lung ultrasound monitor fluid clearance and predict the need of respiratory support? *Crit Care* 2012; **16**: R220 [PMID: 23151314 DOI: 10.1186/cc11865]
75 **Kewitz G**, Wudel S, Hopp H, Hopfenmüller W, Vogel M, Roots I. Below median birth weight in appropriate-for-gestational-age preterm infants as a risk factor for bronchopulmonary dysplasia. *J Perinat Med* 2008; **36**: 359-364 [PMID: 18598128 DOI: 10.1515/JPM.2008.056]
76 **McEvoy CT**, Jain L, Schmidt B, Abman S, Bancalari E, Aschner JL. Bronchopulmonary dysplasia: NHLBI Workshop on the Primary Prevention of Chronic Lung Diseases. *Ann Am Thorac Soc* 2014; **11** Suppl 3: S146-S153 [PMID: 24754823 DOI: 10.1513/AnnalsATS.201312-424LD]
77 **Liu J**, Chen SW, Liu F, Wang Y, Kong XY, Li QP, Huang JJ. BPD, Not BPD, or iatrogenic BPD: findings of lung ultrasound examinations. *Medicine (Baltimore)* 2014; **93**: e133 [PMID: 25415666 DOI: 10.1097/MD.0000000000000133]
78 **Mohamed A**, Mohsen N, Diambomba Y, Lashin A, Louis D, Elsayed Y, Shah PS. Lung Ultrasound for Prediction of Bronchopulmonary Dysplasia in Extreme Preterm Neonates: A Prospective Diagnostic Cohort Study. *J Pediatr* 2021; **238**: 187-192.e2 [PMID: 34237347 DOI: 10.1016/j.jpeds.2021.06.079]
79 **Poggi SH**, Ghidini A. Pathophysiology of meconium passage into the amniotic fluid. *Early Hum Dev* 2009; **85**: 607-610 [PMID: 19836908 DOI: 10.1016/j.earlhumdev.2009.09.011]
80 **Chettri S**, Bhat BV, Adhisivam B. Current Concepts in the Management of Meconium Aspiration Syndrome. *Indian J Pediatr* 2016; **83**: 1125-1130 [PMID: 27206687 DOI: 10.1007/s12098-016-2128-9]
81 **Cattarossi L**. Lung ultrasound: its role in neonatology and pediatrics. *Early Hum Dev* 2013; **89** Suppl 1: S17-S19 [PMID: 23809341 DOI: 10.1016/S0378-3782(13)70006-9]
82 **Liu J**, Cao HY, Fu W. Lung ultrasonography to diagnose meconium aspiration syndrome of the newborn. *J Int Med Res* 2016; **44**: 1534-1542 [PMID: 27807253 DOI: 10.1177/0300060516663954]
83 **Sameshima YT**, Lourenço de Almeida JF, Silva MM, Remondini R, Haddad LB, Neto MJ, Buarque de Gusmão Funari M. Ultrasound-guided lung recruitment in a 3-month-old infant with acute respiratory distress syndrome. *Ultrasound Q* 2014; **30**: 301-305 [PMID: 25364957 DOI: 10.1097/RUQ.0000000000000072]
84 **Liu J**, Ren XL, Fu W, Liu Y, Xia RM. Bronchoalveolar lavage for the treatment of neonatal pulmonary atelectasis under lung ultrasound monitoring. *J Matern Fetal Neonatal Med* 2017; **30**: 2362-2366 [PMID: 27756159 DOI: 10.1080/14767058.2016.1248935]
85 **Chowdhry R**, Dangman B, Pinheiro JM. The concordance of ultrasound technique versus X-ray to confirm endotracheal tube position in neonates. *J Perinatol* 2015; **35**: 481-484 [PMID: 25611791 DOI: 10.1038/jp.2014.240]
86 **Zaytseva A**, Kurepa D, Ahn S, Weinberger B. Determination of optimal endotracheal tube tip depth from the gum in neonates by X-ray and ultrasound. *J Matern Fetal Neonatal Med* 2020; **33**: 2075-2080 [PMID: 30332898 DOI: 10.1080/14767058.2018.1538350]
87 **Takeuchi S**, Arai J. Ultrasonographic confirmation of tracheal intubation for congenital chylothorax. *Pediatr Int* 2018; **60**: 308-310 [PMID: 29480539 DOI: 10.1111/ped.13493]
88 **Lee MGY**, Millar J, Rose E, Jones A, Wood D, Luitingh TL, Zannino D, Brink J, Konstantinov IE, Brizard CP, d'Udekem Y. Laryngeal ultrasound detects a high incidence of vocal cord paresis after aortic arch repair in neonates and young children. *J Thorac Cardiovasc Surg* 2018; **155**: 2579-2587 [PMID: 29510943 DOI: 10.1016/j.jtcvs.2017.12.133]
89 **Pitt B**, Filippatos G, Agarwal R, Anker SD, Bakris GL, Rossing P, Joseph A, Kolkhof P, Nowack C, Schloemer P, Ruilope LM; FIGARO-DKD Investigators. Cardiovascular Events with Finerenone in Kidney Disease and Type 2 Diabetes. *N Engl J Med* 2021; **385**: 2252-2263 [PMID: 34449181 DOI: 10.1056/NEJMoa2110956]
90 **Llamas-Álvarez AM**, Tenza-Lozano EM, Latour-Pérez J. Diaphragm and Lung Ultrasound to Predict Weaning Outcome: Systematic Review and Meta-Analysis. *Chest* 2017; **152**: 1140-1150 [PMID: 28864053 DOI: 10.1016/j.chest.2017.08.028]
91 **Alonso-Ojembarrena A**, Ruiz-González E, Estepa-Pedregosa L, Armenteros-López AI, Segado-Arenas A, Lubián-López SP. Reproducibility and reference values of diaphragmatic shortening fraction for term and premature infants. *Pediatr Pulmonol* 2020; **55**: 1963-1968 [PMID: 32458563 DOI: 10.1002/ppul.24866]
92 **Yousef N**, Mokhtari M, Durand P, Raimondi F, Migliaro F, Letourneau A, Tissières P, De Luca D. Lung Ultrasound Findings in Congenital Pulmonary Airway Malformation. *Am J Perinatol* 2018; **35**: 1222-1227 [PMID: 29715700 DOI: 10.1055/s-0038-1645861]
93 **Adin ME**. Ultrasound as a screening tool in the follow-up of asymptomatic congenital cystic adenomatoid malformation. *Ultrasound* 2016; **24**: 175-179 [PMID: 27867411 DOI: 10.1177/1742271X16657120]
94 **Balcells C**, Del Río R, Riaza L, Rebollo M, Rodriguez-Fanjul J, Camprubí M. Lung ultrasound: a useful tool for the follow-up of neonatal localized interstitial emphysema. *J Pediatr* 2015; **166**: 1543 [PMID: 25799192 DOI: 10.1016/j.jpeds.2015.02.016]

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

ISSN 2644-3260 (online)

Artificial Intelligence in *Medical Imaging*

Artif Intell Med Imaging 2022 April 28; 3(2): 21-54

Published by Baishideng Publishing Group Inc

AIMI

Artificial Intelligence in Medical Imaging

Contents

Bimonthly Volume 3 Number 2 April 28, 2022

MINIREVIEWS

SYSTEMATIC REVIEWS

Artificial Intelligence in Medical Imaging

Bimonthly Volume 3 Number 2 April 28, 2022

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Medical Imaging*, Drozdstoy Stoyanov, MD, PhD, DSc, IDFAPA, Professor, Head Department of Psychiatry and Medical Psychology, Scientific Secretary of the Research Institute at Medical University Plovdiv; Vice President of the Bulgarian Academy of Sciences and Arts; Vassil Aprilov 15 a, Plovdiv 4002, Bulgaria. drozdstoy.stoyanov@mu-plovdiv.bg

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
April 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

Artificial Intelligence in Medical Imaging

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

Artif Intell Med Imaging 2022 April 28; 3(2): 21-32

DOI: 10.35711/aimi.v3.i2.21

ISSN 2644-3260 (online)

MINIREVIEWS

Role of artificial intelligence in early detection and screening for pancreatic adenocarcinoma

Kenneth Weicong Lin, Tiing Leong Ang, James Weiquan Li

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): C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Nazari N, Iran; Zhang JX, China

Received: December 16, 2021
Peer-review started: December 16, 2021
First decision: January 26, 2022
Revised: February 12, 2022
Accepted: March 16, 2022
Article in press: March 16, 2022
Published online: April 28, 2022

Kenneth Weicong Lin, Tiing Leong Ang, James Weiquan Li, Department of Gastroenterology and Hepatology, Changi General Hospital, Singapore 529889, Singapore

Corresponding author: James Weiquan Li, FRCPE, MBBS, MMed, Assistant Professor, Doctor, Department of Gastroenterology and Hepatology, Changi General Hospital, 2 Simei Street 3, Singapore 529889, Singapore. james.li.w.q@singhealth.com.sg

Abstract

Pancreatic adenocarcinoma remains to be one of the deadliest malignancies in the world despite treatment advancement over the past few decades. Its low survival rates and poor prognosis can be attributed to ambiguity in recommendations for screening and late symptom onset, contributing to its late presentation. In the recent years, artificial intelligence (AI) as emerged as a field to aid in the process of clinical decision making. Considerable efforts have been made in the realm of AI to screen for and predict future development of pancreatic ductal adenocarcinoma. This review discusses the use of AI in early detection and screening for pancreatic adenocarcinoma, and factors which may limit its use in a clinical setting.

Key Words: Artificial intelligence; Pancreatic cancer; Pancreatic adenocarcinoma; screening; Early detection

Core Tip: Pancreatic adenocarcinoma has poor survival rate and high morbidity. Artificial intelligence is a potential tool to screen for high risk individuals and for early detection of pancreatic adenocarcinoma. Despite advances made in artificial intelligence research in pancreatic adenocarcinoma, it faces a number of challenges before it can be generalised and applied in a clinical setting.

Citation: Lin KW, Ang TL, Li JW. Role of artificial intelligence in early detection and screening for pancreatic adenocarcinoma. *Artif Intell Med Imaging* 2022; 3(2): 21-32
URL: https://www.wjgnet.com/2644-3260/full/v3/i2/21.htm
DOI: https://dx.doi.org/10.35711/aimi.v3.i2.21

INTRODUCTION

The global incidence of pancreatic cancer is increasing, and it remains as one of the leading causes of cancer-related death, with 495773 new cases of pancreatic cancer diagnosed and accounting for 466003 deaths in 2020[1]. Although the 5-year survival rates for pancreatic ductal adenocarcinoma (PDAC) have improved, it remains low at approximately 9%[2,3], and the overall prognosis of PDAC is poor. This is partly due to the late stage of presentation of PDAC, which is largely dependent on patient symptoms for suspicion of the disease[4,5]. Early cases are asymptomatic and there is a lack of a simple screening tool for clinical use unlike the case of colorectal cancer screening where screening can be performed in the primary care setting with the use of fecal immunohistochemical test. In the case of PDAC, cross-sectional imaging tests such as computed tomography (CT) or magnetic resonance imaging (MRI) are needed for detection, making widespread population screening unfeasible. Germline mutations and a family history of PDAC have been identified as the strongest risk factors for the disease [6,7]. As such, efforts in screening programmes have focused their attention on this group of patients[8]. Pancreatic cysts, increased age, and smoking are also known risk factors for PDAC[5,9,10], although it may not be practical to conduct routine surveillance for patients with these risk factors. There is an interest in selecting higher risk patients for screening, as the appropriate use biomarkers and imaging may result in detection of early-stage PDAC amenable to curative resection[2,3,11-15].

Artificial intelligence (AI) is a branch in computer science where computer systems are designed to perform tasks which would require human intelligence. It is recognised as a potential tool as part of the screening efforts and building predictive models[16]. Most progress for AI in endoscopy has been made in the field of colonoscopy, where polyp detection and characterisation has been studied[17]. Computer-aided diagnosis has also been extended to detection and screening of PDAC[18] in endoscopic ultrasound (EUS)[19,20], MRI[21] and cytology from fine needle sampling[22]. In recent years, various groups have harnessed the potential of AI in creating prediction models. These include The Felix Project [23], the Pancreatic-Cancer Collective[24], and the Early Detection Research Network[25] effort.

This mini-review aims to study the role of AI in the early detection and screening for pancreatic cancer, as well as factors which may limit its use.

METHODS

A comprehensive literature search was performed in the PubMed, MEDLINE and EMBASE electronic databases from the inception of the databases up to and including 30 November 2021. The key words used were "artificial intelligence", "pancreatic cancer", "pancreatic adenocarcinoma", "pancreatic ductal adenocarcinoma", "pancreatic carcinoma", "screening", and "early detection". These were supplemented with manual searches of references from retrieved articles. Publications in English were considered for this mini-review.

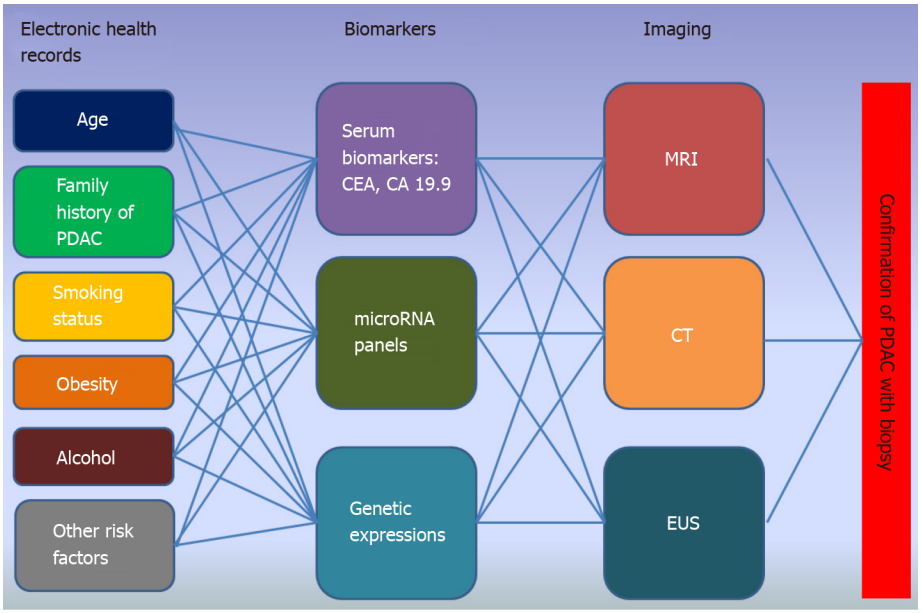
AI BASIC PRINCIPLES AND TERMINOLOGIES

AI is a term that refers to the ability of a computer programme to imitate the human mind to perform tasks such as problem solving and learning[26,27].

Machine learning (ML) is the commonest branch of AI used in medicine and refers to a mathematical model that aims to generate a prediction based on a set of data provided[28,29]. In supervised learning, the data points are labelled and the ML model "learns" from these labels and identifies new data points. In contrast, labels are not provided in unsupervised learning, and the model recognises the patterns of the data by learning its unknown properties and identifying crucial data checkpoints. This is especially important when the gold standard is not available[29].

Deep learning (DL) is subset of ML that employs the use of Artificial Neural Networks (ANN). Like the human brain, ANN consists of layers of artificial neurons that are interlinked. Each layer receives a weighted signal from the previous layer(s) and these signals will be propagated to the next layer when a specific threshold is exceeded[29]. In the setting of a pancreatic lesion or cancer, DL first identifies the basics of the lesion (*e.g.*, location) in its initial layers before moving on to next layer for further characterisation (*e.g.*, size, shape, colour). A final prediction of the pancreatic lesion is made after a systematic assessment *via* multiple layers of neural network[29].

ANNs are first trained using the training data set, where the model learns to identify specific patterns to obtain a relationship between the input and the output. Hyperparameters refer to all settings that are pre-determined by the investigator and are used to construct the model for optimal execution of a particular task or on a specific dataset. The validation data set involves a different data set that is used to fine-tune the hyperparameters of the model. Finally, the test data set refers to a data set whose purpose is to evaluate the performance of the model against unseen data and determine its generalizability[29]. This set needs to be unseen by the model during training and validation. However in certain studies, the test set is sometimes a subset of the training or validation data set, which many result in



DOI: 10.35711/aimi.v3.i2.21

Figure 1 Schematic diagram showing the workflow and neural network to be designed for an early detection protocol. CT: Computed tomography; CEA: Carcinoembryonic antigen; PDAC: Pancreatic ductal adenocarcinoma; MRI: Magnetic resonance imaging.

overfitting of the model. This may lead to a discrepancy in the performance of the model when tested in the same centre and a decline in performance when validated externally.

MODEL FOR SCREENING FOR AND EARLY IDENFICATION OF DEVELOPING PDAC

Early detection of pancreatic cancer requires a step wise approach in order to systematically screen for risk factors and identify high-risk groups. Figure 1 is a schematic diagram showing the workflow and neural network to be designed for an early detection protocol. It represents the complex interplay between each of the input(s) to be processed for the next neural layer(s) until a final output is obtained. We will be discussing the role of AI in early detection of pancreatic cancer based on this model.

AI IN CLINICAL DECISION MAKING USING HEALTH RECORDS

The identification of risk factors for pancreatic cancer is essential in identifying the specific population which would benefit from screening[18,30,31]. Factors such as diabetes, hemoglobin A1C (HbA1c) value, weight, body mass index (BMI), blood type, smoking status, alcohol use and family history of pancreatic cancer influence the age of onset of screening for an individual[13,32]. These factors are easily available in the primary care setting and could potentially predict the development of pancreatic cancer within 5 years, even before any changes to the pancreas can be detected on imaging[30]. However, most of the data is stored in health records, which are often proprietary or internet-separated to protect patient data. The retrieval and subsequent integration of data from different platforms remains a manual and laborious process for physicians[30]. Even after retrieval, there are no validated scoring systems to assess these risk factors and stratify patients. On the other hand, AI, with the aid of Natural Language Processing, can facilitate this process[33-38]. In a case-control study, Malhotra *et al*[33] created an algorithm based on electronic health records (EHR) obtained from primary care to identify 41.3% of patients (≤ 60 years old) who had significant risk of developing pancreatic cancer up to 20 mo prior to diagnosis with a sensitivity, specificity, area under the receiver operating characteristic (AUROC) curve of 72.5%, 59.0% and 0.66%, respectively. Similarly, Appelbaum *et al*[35] was able to train an ANN using 101381 EHRs to predict the development of PDAC one year before the diagnosis in a population of high-risk patients (AUROC 0.68, confidence interval (CI): 0.65-0.71).

Despite its potential benefits, research in AI for the above purpose is still preliminary as they are mostly based on retrospective data from single institutions or registries, and hence not ready for use in a wider clinical setting[33-38]. One of the major limitations would be the lack validation in the real-world setting or at least in populations derived from different centres to overcome the risk of bias and overfitting.

The use of AI in EHR faces other challenges. Various institutions' medical records are built on different healthcare systems and encoding systems, making the task of harmonising them difficult[30]. There is also a lack of standardised clinical research data collection models. To overcome this, efforts are made to build a model of processing and integrating data across institutions. The i2b2 was created to review medical records, retrieve specific data of interest and repurpose it for research[39]. The Observational Health Data Sciences and Informatics was developed from the Observational Medical Outcomes Partnership, an initiative that develops the Common Data Model aiming to gather information from different data sets or medical repositories and systemically analyse them in a common platform[40]. Similarly, the National Patient-centered Clinical research network is another example which was developed in United States to access millions of EHR and create a common data set for research purposes[41]. A common dataset with a standardised format for input of data relevant to PDAC would enable AI systems to leverage on big data to identify changing risk profiles in PDAC, enabling the clinician to channel resources for screening to the appropriate cohorts of patients depending on the population from which this data has been derived.

While these are upcoming and promising initiatives, concerns surrounding restrictions in data sharing, privacy issues, and maintenance costs could hinder data collection efforts[18]. EHRs are also stored in different languages in different regions of the world, making the integration of data difficult. Besides, once data sets are gathered, obtaining IRB approval from the various sites for research may be difficult.

AI AND THE USE OF NON-INVASIVE BIOMARKERS

Carbohydrate Antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA) are the most widely used markers for screening of PDAC, but have also been proven to lack the specificity when applied individually and without clinical context[42,43]. On the other hand, a combined measurement can potentially increase its sensitivity and specificity up to 1 year before the diagnosis of PDAC[44-46]. Capitalising on this concept, Yang *et al*[47], developed an algorithm (with 658 subjects in its training set) to diagnose pancreatic cancer by using ANN to combine CA19-9, CA125 and CEA values. This model was subsequently evaluated against the test set and was able to yield an AUROC of 0.905 (95%CI = 0.868-0.942) and a high diagnostic accuracy of 83.5% for pancreatic cancer.

New biomarkers for PDAC such as MicroRNAs and gene expressions have generated much interest in the recent years[45,48-52]. MicroRNAs are non-coding RNAs that are involved in the regulation of biological pathways, and when altered, could lead to the development of PDAC[53]. MicroRNAs can potentially predict future PDAC[54] or detect early stage pancreatic cancer. However, they have the same limitations in sensitivity and specificity when applied without clinical context and as independent test[55,56]. A combination of the commonly used biomarkers and newer biomarkers may address the problem of low sensitivity and specificity[56], and in particular can be combined with clinical and demographic information as described earlier to increase its usefulness.

While AI is able to make use of plasma microRNA panels and specific gene expressions to diagnose pancreatic cancer[57,58], studies on their use on predicting future pancreatic cancer are not available [55]. By integrating Particle Swarm Optimization, ANN and Neighborhood Component Analysis iterations on a list of microRNAs that are most commonly expressed by pancreatic cancer, Alizadeh *et al* [59] created a model consisting of 5 MicroRNAs (miR-663a, miR-1469, miR-92a-2-5p, miR-125b-1-3p and miR-532-5p) to diagnose pancreatic cancer (Accuracy: 0.93, Sensitivity: 93%, and Specificity: 92%). Similarly in a multicentre study by Cao *et al*[57], a machine learning approach was able to identify 2 panels of microRNAs to differentiate pancreatic cancer from chronic pancreatitis with an accuracy of above 80%.

Gene expressions have gained popularity in diagnosing pancreatic cancer[13,60]. Using a machine learning approach, Khatri *et al*[61] analysed the results from transcriptomics-based meta-analysis to create a nine-gene panel to diagnose pancreatic cancer. This panel was able to differentiate PDAC from chronic pancreatitis with a specificity of 89%, sensitivity of 78%, and accuracy of 83% and an AUROC of 0.95. As compared to a normal pancreas, it was also used to identify stage I and II PDACs with a sensitivity of 74%, specificity of 75%, and an AUROC of 0.82. In another study, a machine learning algorithm was formulated based on the biochemical differences in the serum of 2 groups of subjects (PDAC group and High risk group) detected *via* the use of Probe Electrospray Ionization Mass Spectrometry (PESI-MS) to identify early stages of pancreatic cancer[62]. It was able to differentiate healthy controls from subjects with earlier stage of PDAC with sensitivity of 81.2% and specificity of 96.8% respectively and an accuracy of 92.9%.

At present, these studies have shown that AI can offer the advantage of identifying specific microRNA and genetic combinations to identifying pancreatic cancer at a faster speed, making this process less laborious. However, these studies lack external validation, limiting their application in modern practice. Besides, studies utilising AI to formulate specific sequences to accurately predict future pancreatic cancer development are still lacking. More studies are required to analyse its ability in predicting future pancreatic cancer for high risk groups especially during the latency period.

Table 1 Studies on artificial intelligence using computed tomography or MRI imaging to diagnose pancreatic ductal adenocarcinoma

Ref.	Clinical question	Training set (number of subjects)	Validation set (number of subjects)	AI instrument	AUROC	Accuracy	Sensitivity	Specificity
Watson *et al* [66], 2021	Detection of pancreatic cystic neoplasms (including PDAC) *vs* benign cysts	18	9	CNN	NA	NA	NA	NA
Si *et al* [65], 2021	Detection of pancreatic cancer (including PDAC, IPMN, PNET)	319	347	DL	0.871	87.6% for PDAC	86.8% for pancreatic cancer	69.5% for pancreatic cancer
Park *et al*[64], 2020	Distinguishing pancreatic cancer tissue from autoimmune pancreatitis	120	62	Random forest machine learning	0.975	95.2%	89.7%	100%
Ma *et al* [63], 2020	Differentiate pancreatic cancer from benign tissue	330	41	CNN	0.9653 (plain scan)	95.47% (plain scan),95.76% (arterial scan), 95.15% (venous phase)	91.58% (plain scan), 94.08% (arterial scan), 92.28% (venous phase)	98.3% (plain scan), 97.6% (arterial scan), 97.9% (venous phase)
Zhang *et al* [67], 2020	Detection of pancreatic cancer	2650 images	240 images	CNN	0.9455	90.2%	83.8%	91.8%
Liu *et al* [69], 2020	Differentiating pancreatic cancer tissue from non-cancerous pancreatic tissue	412	139	CNN	0.92	83.2%	79.0%	97.6%
Gao *et al*[71], 2020	To differentiate pancreatic diseases in pancreatic lesions	398	106	CNN	0.9035 (includes PDAC, adenosquamous carcinoma, acinar cell carcinoma, colloid carcinoma, myoepithelial carcinoma, undifferentiated carcinoma with osteoclast-like giant cells, mucinous cystadenocarcinoma, pancreatoblastoma, pancreatic neuroendocrine carcinoma and metastatic carcinoma)	NA	NA	NA
Chu *et al*[70], 2019	Differentiating PDAC from normal pancreas	255	125	Random forest	NA	93.6%	95%	92.3%
Zhu *et al*[72], 2019	Detecting PDAC from normal pancreas	205	234	CNN	NA	57.3%	94.1%	98.5%
Liu *et al* [73], 2019	Diagnosis of pancreatic cancer	238	100	CNN	0.9632	NA	NA	NA
Corral *et al* [21], 2019	Identify and stratify IPMN lesions	139		DL	0.783	NA	75% (for PDAC or high grade dysplasia)	78% (for PDAC or high grade dysplasia)
Chu *et al*[74], 2019	Differentiating PDAC from normal pancreas	456		DL	NA	NA	94.1%	98.5%
Fu *et al* [75], 2018	Pancreas segmentation (including PDAC, IPMN, Pancreatic Neuroendocrine	59		CNN	NA	NA	82.5%	76.22 (PPV)

	Tumors, Serous Cyst Adenoma, and Solid Pseudopapillary Tumour of the pancreas)				

AUROC: Area under the receiver operating characteristic; AI: Artificial intelligence; CNN: Convolutional neural network; DL: Deep learning; NA: Not available; IPMN: Intraductal papillary mucinous neoplasm; PNET: Pancreatic neuroendocrine tumour; PDAC: Pancreatic ductal adenocarcinoma.

CURRENT EVIDENCE IN PREDICTING THE DEVELOPMENT OF PANCREATIC LESIONS INTO PDAC IN THE FUTURE

Various studies have been conducted using AI to diagnose pancreatic cancer and yielded promising results. Table 1 summarises the studies to date[21,63-75]. In a retrospective study, Liu *et al*[69] was able to train a convolutional neural network (CNN) to identify pancreatic cancer on contrast-enhanced CT and achieve an AUROC of 0.9, with more than 90% for its sensitivity and specificity for its test set. It maintained good sensitivity of 91.3%, specificity of 84.5%, an accuracy of 85.6% and AUROC of 0.955 (95%CI 0.955-0.956) with the validation set. Further analysis revealed that with CNN, radiologists missed 7% of the pancreatic cancers, of which majority were accurately diagnosed by CNN[69]. By enhancing the CNN, Liu *et al*[73] was able to process the CT images and obtain the diagnosis faster than the radiologists (3 s for CNN *vs* 8 mins for a radiologist) with an AUROC of 0.9632, proving that AI is comparable to radiologists.

Besides CT, EUS has been frequently utilised to diagnosed pancreatic cancer. Table 2 summaries these studies[19,20,76-86]. The EUS-CAD based CNN was developed in a retrospective study by Tonozuka *et al*[83] to identify lesions harbouring pancreatic cancer in patients with chronic pancreatitis with a sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of 90.2%, 74.9%, 80.1%, and 88.7%, respectively, and an AUROC of 0.924. Similar findings were also echoed in Zhu *et al*[86] who utilised SVM to obtain a sensitivity, specificity, PPV and NPV of over 90% for diagnosis of pancreatic cancer in chronic pancreatitis.

Despite numerous studies looking at using AI to diagnose pancreatic cancer (as shown in Tables 1 and 2), only a few attempted to predict the development to pancreatic cancer. On average, CT changes for early pancreatic cancer starts approximately 12 to 18 mo before diagnosis[87]. Yet, pancreatic cancer can advance from being undetectable to metastatic in a short period of time even before the next surveillance imaging[88,89]. AI-based imaging itself cannot be used to predict pancreatic cancer and should be combined with other markers.

An ideal AI model for predicting pancreatic cancer is one that integrates multiple biochemical, radiological and clinical data[90]. In a retrospective proof-of-concept study, Springer *et al*[91] developed a supervised machine learning-based approach (CompCyst) based on a combination of patient-reported symptoms, imaging results (including CT, MRI and EUS images), cyst fluid and molecular characteristics to calculate its malignant potential and subsequently determine the management of pancreatic cyst(s). When tested against the validation set, CompCyst outperformed the current standard of care (accuracy 56%) in its ability to identify patients who required surgery, close monitoring or can be discharged (accuracy 69%). CompCyst correctly identified 60% of the surgeries that were not warranted and could have been avoided, while not compromising on its ability to identifying those who truly require surgery. With CompCyst, 71% of the pancreatic lesions were correctly identified as PDAC as compared to 58% based on clinical suspicion[91].

While this study has proven that AI has the potential to incorporate various clinical characteristics, biomarkers, and imaging characteristics to assess for the malignant potential of a pancreatic lesion, it has a number of limitations. Firstly, the imaging characteristics and molecular biomarkers that were identified as high risk features were obtained at the time of surgery and not during screening. These features may not be present early enough to be identified by routine screening. Secondly, important risk factors (including age and diabetes) that were crucial in the early detection of PDAC (as shown in Figure 1) were not included in its learning process, representing a missed step in the screening process. Finally, CompCyst is yet to be externally validated and cannot be applied to the clinical setting currently.

While CompCyst is a potential tool to aid in clinical decision making, future studies aiming at early detection of PDAC face a myriad of challenges. Firstly, the pancreas is a complex organ. Unlike the other organs, the pancreas can be highly variable in its anatomy and location. Moreover, the training data set is highly dependent on the quality of the images provided. Hence, automated segmentation of the pancreas *via* a deep learning approach remains challenging[92]. Secondly, the lack of databases limits the ability to develop new training sets. There are currently only a few open-access databases[93], and there are issues regarding sharing of images across various institutions as pointed out by the Alliance of Pancreatic Cancer Consortia imaging working group[90]. Finally, the algorithm for early detection of PDAC will have to evaluate images of pancreatic lesion(s) across different time points of

Table 2 Studies on artificial intelligence using endoscopic ultrasound to diagnose pancreatic ductal adenocarcinoma

Ref.	Clinical question	Training set (number of subjects)	Validation set (number of subjects)	AI instrument	AUROC	Accuracy	Sensitivity	Specificity
Udristoiu *et al*[84], 2021	Detecting focal pancreatic masses in four EUS imaging modalities	65		CNN and Long Short-term Memory models	0.97	97.6%	98.1%	96.7%
Tonozuka *et al*[83], 2021	Detecting PDAC in patients with normal pancreas/Chronic pancreatitis	92		CNN	0.924	NA	90.2%	74.9%
Marya *et al* [78], 2021	Differentiate AIP from PDAC, chronic pancreatitis and other pancreatic diseases	336	124	CNN	0.976	NA	95%	90%
Kuwahara *et al*[77], 2019	Predicting malignancy in IPMN	50		CNN	0.98	94%	95.7%	92.6%
Ozkan *et al* [80], 2016	Differentiating pancreatic cancer from healthy pancreas	260 images	72 images	ANN	NA	87.5%	83.3%	93.3%
Saftoiu *et al* [81], 2015	Differentiate pancreatic cancer from chronic pancreatitis	117	25	ANN	NA	NA	94.6%	94.4%
Zhu *et al* [86], 2013	Differentiating pancreatic cancer from chronic pancreatitis.	194	194	SVM	NA	94.2%	96.3%	93.4%
Saftoiu *et al* [82], 2012	Diagnosis of focal pancreatic lesions	258 patients		ANN	0.94	84.27%	87.59%	82.94%
Zhang *et al* [85], 2010	Differentiate pancreatic cancer from non-tumorous tissue	108	108	SVM	NA	97.98%	94.3%	99.45%
Saftoiu *et al* [20], 2008 cancer	Differentiate normal pancreas, chronic pancreatitis, pancreatic cancer, and neuroendocrine tumors	68		Neural network	0.847 (for PDAC *vs* chronic pancreatitis)	86.1% (for PDAC *vs* chronic pancreatitis)	93.8% (for PDAC *vs* chronic pancreatitis)	63.6% (for PDAC *vs* chronic pancreatitis)
Das *et al* [19], 2008	Differentiating pancreatic adenocarcinoma from non-neoplastic tissue (includes normal pancreas and chronic pancreatitis)	160	159	ANN	0.93	NA	93%	92%
Norton *et al* [79], 2001	Differentiate malignancy from pancreatitis	35		ML	NA	80%	100%	50%

AUROC: Area under the receiver operating characteristic; AI: Artificial intelligence; CNN: Convolutional neural network; EUS: Endoscopic ultrasound; SVM: Support vector machines; ML: Machine learning; NA: Not available; IPMN: Intraductal papillary mucinous neoplasm; PDAC: Pancreatic ductal adenocarcinoma.

surveillance and from different 3 imaging modalities (namely CT, MRI, and EUS). Unlike CompCyst which looks at images at one time point (*i.e.* at surgery), combining multiple images obtained from periodical surveillance *via* these 3 imaging modalities will require a very large database and multiple layers.

There is a major gap that needs to be bridged before AI systems for early detection of pancreatic cancer can be developed. Given sufficient training data and cooperation, AI-based image analyzers could match or even outperform physicians in image classification and lesion detection[90].

CONCLUSION

Despite the recent advances to predict future PDAC, the use of AI in screening for pancreatic cancer

remains limited in the clinical setting. Much of the efforts are made in the research setting and lack external validation and generalisability. However, this field remains promising as we recognise the challenges ahead to bridge the necessary gaps. The hope to develop an integrated AI model to screen for PDAC remains a reality, and it will play a complementary role in assisting physicians in their clinical decision making process but not replace it.

FOOTNOTES

Author contributions: Lin KW performed the literature search and drafted the manuscript; Ang TL performed the literature search and was involved in the drafting of the manuscript; Li JW conceptualised the project, performed literature search and was involved in the drafting of the manuscript; all authors vetted and approved the final manuscript.

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

Country/Territory of origin: Singapore

ORCID number: Kenneth Weicong Lin 0000-0002-0490-4002; Tiing Leong Ang 0000-0001-9993-8549; James Weiquan Li 0000-0002-5241-4278.

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

REFERENCES

1 **GLOBOCAN**. International Agency for Research on Cancer. 2020
2 **Blackford AL**, Canto MI, Klein AP, Hruban RH, Goggins M. Recent Trends in the Incidence and Survival of Stage 1A Pancreatic Cancer: A Surveillance, Epidemiology, and End Results Analysis. *J Natl Cancer Inst* 2020; **112**: 1162-1169 [PMID: 31958122 DOI: 10.1093/jnci/djaa004]
3 **Huang L**, Jansen L, Balavarca Y, Babaei M, van der Geest L, Lemmens V, Van Eycken L, De Schutter H, Johannesen TB, Primic-Žakelj M, Zadnik V, Besselink MG, Schrotz-King P, Brenner H. Stratified survival of resected and overall pancreatic cancer patients in Europe and the USA in the early twenty-first century: a large, international population-based study. *BMC Med* 2018; **16**: 125 [PMID: 30126408 DOI: 10.1186/s12916-018-1120-9]
4 **Gobbi PG**, Bergonzi M, Comelli M, Villano L, Pozzoli D, Vanoli A, Dionigi P. The prognostic role of time to diagnosis and presenting symptoms in patients with pancreatic cancer. *Cancer Epidemiol* 2013; **37**: 186-190 [PMID: 23369450 DOI: 10.1016/j.canep.2012.12.002]
5 **Ryan DP**, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. *N Engl J Med* 2014; **371**: 1039-1049 [PMID: 25207767 DOI: 10.1056/NEJMra1404198]
6 **Abe T**, Blackford AL, Tamura K, Ford M, McCormick P, Chuidian M, Almario JA, Borges M, Lennon AM, Shin EJ, Klein AP, Hruban RH, Canto MI, Goggins M. Deleterious Germline Mutations Are a Risk Factor for Neoplastic Progression Among High-Risk Individuals Undergoing Pancreatic Surveillance. *J Clin Oncol* 2019; **37**: 1070-1080 [PMID: 30883245 DOI: 10.1200/JCO.18.01512]
7 **Klein AP**. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. *Nat Rev Gastroenterol Hepatol* 2021; **18**: 493-502 [PMID: 34002083 DOI: 10.1038/s41575-021-00457-x]
8 **Capurso G**, Paiella S, Carrara S, Butturini G, Secchettin E, Frulloni L, Zerbi A, Falconi M. Italian registry of families at risk of pancreatic cancer: AISP Familial Pancreatic Cancer Study Group. *Dig Liver Dis* 2020; **52**: 1126-1130 [PMID: 32819857 DOI: 10.1016/j.dld.2020.07.027]
9 **Crippa S**, Bassi C, Salvia R, Malleo G, Marchegiani G, Rebours V, Levy P, Partelli S, Suleiman SL, Banks PA, Ahmed N, Chari ST, Fernández-Del Castillo C, Falconi M. Low progression of intraductal papillary mucinous neoplasms with worrisome features and high-risk stigmata undergoing non-operative management: a mid-term follow-up analysis. *Gut* 2017; **66**: 495-506 [PMID: 26743012 DOI: 10.1136/gutjnl-2015-310162]
10 **Weissman S**, Takakura K, Eibl G, Pandol SJ, Saruta M. The Diverse Involvement of Cigarette Smoking in Pancreatic Cancer Development and Prognosis. *Pancreas* 2020; **49**: 612-620 [PMID: 32433397 DOI: 10.1097/MPA.0000000000001550]
11 **Canto MI**, Almario JA, Schulick RD, Yeo CJ, Klein A, Blackford A, Shin EJ, Sanyal A, Yenokyan G, Lennon AM, Kamel IR, Fishman EK, Wolfgang C, Weiss M, Hruban RH, Goggins M. Risk of Neoplastic Progression in Individuals at High Risk for Pancreatic Cancer Undergoing Long-term Surveillance. *Gastroenterology* 2018; **155**: 740-751.e2 [PMID: 29803839 DOI: 10.1053/j.gastro.2018.05.035]

12 **Aslanian HR**, Lee JH, Canto MI. AGA Clinical Practice Update on Pancreas Cancer Screening in High-Risk Individuals: Expert Review. *Gastroenterology* 2020; **159**: 358-362 [PMID: 32416142 DOI: 10.1053/j.gastro.2020.03.088]

13 **Gonda TA**, Everett JN, Wallace M, Simeone DM; PRECEDE Consortium. Recommendations for a More Organized and Effective Approach to the Early Detection of Pancreatic Cancer From the PRECEDE (Pancreatic Cancer Early Detection) Consortium. *Gastroenterology* 2021; **161**: 1751-1757 [PMID: 34454916 DOI: 10.1053/j.gastro.2021.08.036]

14 **Kenner BJ**, Chari ST, Maitra A, Srivastava S, Cleeter DF, Go VL, Rothschild LJ, Goldberg AE. Early Detection of Pancreatic Cancer-a Defined Future Using Lessons From Other Cancers: A White Paper. *Pancreas* 2016; **45**: 1073-1079 [PMID: 27518362 DOI: 10.1097/MPA.0000000000000701]

15 **Overbeek KA**, Levink IJM, Koopmann BDM, Harinck F, Konings ICAW, Ausems MGEM, Wagner A, Fockens P, van Eijck CH, Groot Koerkamp B, Busch ORC, Besselink MG, Bastiaansen BAJ, van Driel LMJW, Erler NS, Vleggaar FP, Poley JW, Cahen DL, van Hooft JE, Bruno MJ; Dutch Familial Pancreatic Cancer Surveillance Study Group. Long-term yield of pancreatic cancer surveillance in high-risk individuals. *Gut* 2021 [PMID: 33820756 DOI: 10.1016/j.pan.2019.05.303]

16 **Richter AN**, Khoshgoftaar TM. A review of statistical and machine learning methods for modeling cancer risk using structured clinical data. *Artif Intell Med* 2018; **90**: 1-14 [PMID: 30017512 DOI: 10.1016/j.artmed.2018.06.002]

17 **Li JW**, Ang TL. Colonoscopy and artificial intelligence: Bridging the gap or a gap needing to be bridged? *AIGE* 2021; **2**: 36-49 [DOI: 10.37126/aige.v2.i2.36]

18 **Kenner B**, Chari ST, Kelsen D, Klimstra DS, Pandol SJ, Rosenthal M, Rustgi AK, Taylor JA, Yala A, Abul-Husn N, Andersen DK, Bernstein D, Brunak S, Canto MI, Eldar YC, Fishman EK, Fleshman J, Go VLW, Holt JM, Field B, Goldberg A, Hoos W, Iacobuzio-Donahue C, Li D, Lidgard G, Maitra A, Matrisian LM, Poblete S, Rothschild L, Sander C, Schwartz LH, Shalit U, Srivastava S, Wolpin B. Artificial Intelligence and Early Detection of Pancreatic Cancer: 2020 Summative Review. *Pancreas* 2021; **50**: 251-279 [PMID: 33835956 DOI: 10.1097/MPA.0000000000001762]

19 **Das A**, Nguyen CC, Li F, Li B. Digital image analysis of EUS images accurately differentiates pancreatic cancer from chronic pancreatitis and normal tissue. *Gastrointest Endosc* 2008; **67**: 861-867 [PMID: 18179797 DOI: 10.1016/j.gie.2007.08.036]

20 **Săftoiu A**, Vilmann P, Gorunescu F, Gheonea DI, Gorunescu M, Ciurea T, Popescu GL, Iordache A, Hassan H, Iordache S. Neural network analysis of dynamic sequences of EUS elastography used for the differential diagnosis of chronic pancreatitis and pancreatic cancer. *Gastrointest Endosc* 2008; **68**: 1086-1094 [PMID: 18656186 DOI: 10.1016/j.gie.2008.04.031]

21 **Corral JE**, Hussein S, Kandel P, Bolan CW, Bagci U, Wallace MB. Deep Learning to Classify Intraductal Papillary Mucinous Neoplasms Using Magnetic Resonance Imaging. *Pancreas* 2019; **48**: 805-810 [PMID: 31210661 DOI: 10.1097/MPA.0000000000001327]

22 **Momeni-Boroujeni A**, Yousefi E, Somma J. Computer-assisted cytologic diagnosis in pancreatic FNA: An application of neural networks to image analysis. *Cancer Cytopathol* 2017; **125**: 926-933 [PMID: 28885766 DOI: 10.1002/cncy.21915]

23 **The Lustgarten Foundation**. Deep Learning for Radiologists: A Beginner's Guide 2021. Available from: https://www.ctisus.com/responsive/deep-learning/felix.asp

24 **The Lustgarten Foundation**. The Pancreatic Cancer Collective 2021. Available from: https://pancreaticcancercollective.org/

25 **National Institutes of Health**. Early Detection Research Network 2021. Available from: https://edrn.nci.nih.gov/

26 **Nakaura T**, Higaki T, Awai K, Ikeda O, Yamashita Y. A primer for understanding radiology articles about machine learning and deep learning. *Diagn Interv Imaging* 2020; **101**: 765-770 [PMID: 33121910 DOI: 10.1016/j.diii.2020.10.001]

27 **Nakata N**. Recent technical development of artificial intelligence for diagnostic medical imaging. *Jpn J Radiol* 2019; **37**: 103-108 [PMID: 30706381 DOI: 10.1007/s11604-018-0804-6]

28 **Shalev-Shwartz S**, Ben-David S. Understanding machine learning. From theory to algorithms. Understanding Machine Learning: From Theory to Algorithms. 2013 [DOI: 10.1017/cbo9781107298019.022]

29 **van der Sommen F**, de Groof J, Struyvenberg M, van der Putten J, Boers T, Fockens K, Schoon EJ, Curvers W, de With P, Mori Y, Byrne M, Bergman JJGHM. Machine learning in GI endoscopy: practical guidance in how to interpret a novel field. *Gut* 2020; **69**: 2035-2045 [PMID: 32393540 DOI: 10.1136/gutjnl-2019-320466]

30 **Kenner BJ**, Abrams ND, Chari ST, Field BF, Goldberg AE, Hoos WA, Klimstra DS, Rothschild LJ, Srivastava S, Young MR, Go VLW. Early Detection of Pancreatic Cancer: Applying Artificial Intelligence to Electronic Health Records. *Pancreas* 2021; **50**: 916-922 [PMID: 34629446 DOI: 10.1097/MPA.0000000000001882]

31 **Poruk KE**, Firpo MA, Mulvihill SJ. Screening for pancreatic cancer. *Adv Surg* 2014; **48**: 115-136 [PMID: 25293611 DOI: 10.1016/j.yasu.2014.05.004]

32 **US Preventive Services Task Force**, Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, Caughey AB, Curry SJ, Doubeni CA, Epling JW Jr, Kubik M, Landefeld CS, Mangione CM, Pbert L, Silverstein M, Simon MA, Tseng CW, Wong JB. Screening for Pancreatic Cancer: US Preventive Services Task Force Reaffirmation Recommendation Statement. *JAMA* 2019; **322**: 438-444 [PMID: 31386141 DOI: 10.1001/jama.2019.10232]

33 **Malhotra A**, Rachet B, Bonaventure A, Pereira SP, Woods LM. Can we screen for pancreatic cancer? *PLoS One* 2021; **16**: e0251876 [PMID: 34077433 DOI: 10.1371/journal.pone.0251876]

34 **Chen Q**, Cherry DR, Nalawade V, Qiao EM, Kumar A, Lowy AM, Simpson DR, Murphy JD. Clinical Data Prediction Model to Identify Patients With Early-Stage Pancreatic Cancer. *JCO Clin Cancer Inform* 2021; **5**: 279-287 [PMID: 33739856 DOI: 10.1200/CCI.20.00137]

35 **Appelbaum L**, Cambronero JP, Stevens JP, Horng S, Pollick K, Silva G, Haneuse S, Piatkowski G, Benhaga N, Duey S, Stevenson MA, Mamon H, Kaplan ID, Rinard MC. Development and validation of a pancreatic cancer risk model for the general population using electronic health records: An observational study. *Eur J Cancer* 2021; **143**: 19-30 [PMID: 33278770 DOI: 10.1016/j.ejca.2020.10.019]

36 **Appelbaum L**, Berg A, Cambronero JP, Dang THY, Jin CC, Zhang L, Kundrot S, Palchuk M, EvansLA, KaplanID, Rinard M. Development of a pancreatic cancer prediction model using a multinational medical records database. *JCO* 2021; **39**: 394 [DOI: 10.1200/jco.2021.39.3_suppl.394]

37 **Muhammad W**, Hart GR, Nartowt B, Farrell JJ, Johung K, Liang Y, Deng J. Pancreatic Cancer Prediction Through an Artificial Neural Network. *Front Artif Intell* 2019; **2**: 2 [PMID: 33733091 DOI: 10.3389/frai.2019.00002]

38 **Placido D**, Yuan B, Hjaltelin JX, Haue AD, Yuan C, Kim J, Umeton R, Antell G, Chowdhury A, Franz A, Brais L, Andrews E, Regev A, Kraft P, WolpinBM, Rosenthal M, Brunak S, Sander C. Pancreatic cancer risk predicted from disease trajectories using deep learning. *BioRxiv* 2021 [DOI: 10.1101/2021.06.27.449937]

39 **Murphy SN**, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010; **17**: 124-130 [PMID: 20190053 DOI: 10.1136/jamia.2009.000893]

40 **Informatics OHDSA**. Data Standardization 2021. Available from: https://ohdsi.org/data-standardization/

41 **Network TNP-CCR**. Accelerating Data Value Across a National Community Health Center (ADVANCE) Network 2021. Available from: http://advancecollaborative.org/

42 **Locker GY**, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, Somerfield MR, Hayes DF, Bast RC Jr; ASCO. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol* 2006; **24**: 5313-5327 [PMID: 17060676 DOI: 10.1200/JCO.2006.08.2644]

43 **Sekiguchi M**, Matsuda T. Limited usefulness of serum carcinoembryonic antigen and carbohydrate antigen 19-9 levels for gastrointestinal and whole-body cancer screening. *Sci Rep* 2020; **10**: 18202 [PMID: 33097814 DOI: 10.1038/s41598-020-75319-8]

44 **Kriz D**, Ansari D, Andersson R. Potential biomarkers for early detection of pancreatic ductal adenocarcinoma. *Clin Transl Oncol* 2020; **22**: 2170-2174 [PMID: 32447642 DOI: 10.1007/s12094-020-02372-0]

45 **Brezgyte G**, Shah V, Jach D, Crnogorac-Jurcevic T. Non-Invasive Biomarkers for Earlier Detection of Pancreatic Cancer-A Comprehensive Review. *Cancers (Basel)* 2021; **13** [PMID: 34072842 DOI: 10.3390/cancers13112722]

46 **Liao Q**, Zhao YP, Yang YC, Li LJ, Long X, Han SM. Combined detection of serum tumor markers for differential diagnosis of solid lesions located at the pancreatic head. *Hepatobiliary Pancreat Dis Int* 2007; **6**: 641-645 [PMID: 18086633]

47 **Yang Y**, Chen H, Wang D, Luo W, Zhu B, Zhang Z. Diagnosis of pancreatic carcinoma based on combined measurement of multiple serum tumor markers using artificial neural network analysis. *Chin Med J (Engl)* 2014; **127**: 1891-1896 [PMID: 24824251]

48 **Yu J**, Ploner A, Kordes M, Löhr M, Nilsson M, de Maturana MEL, Estudillo L, Renz H, Carrato A, Molero X, Real FX, Malats N, Ye W. Plasma protein biomarkers for early detection of pancreatic ductal adenocarcinoma. *Int J Cancer* 2021; **148**: 2048-2058 [PMID: 33411965 DOI: 10.1002/ijc.33464]

49 **Ray K**. Biomarkers for the early detection of PDAC. *Nat Rev Gastroenterol Hepatol* 2017; **14**: 505

50 **Young MR**, Wagner PD, Ghosh S, Rinaudo JA, Baker SG, Zaret KS, Goggins M, Srivastava S. Validation of Biomarkers for Early Detection of Pancreatic Cancer: Summary of The Alliance of Pancreatic Cancer Consortia for Biomarkers for Early Detection Workshop. *Pancreas* 2018; **47**: 135-141 [PMID: 29346214 DOI: 10.1097/MPA.0000000000000973]

51 **Hasan S**, Jacob R, Manne U, Paluri R. Advances in pancreatic cancer biomarkers. *Oncol Rev* 2019; **13**: 410 [PMID: 31044028 DOI: 10.4081/oncol.2019.410]

52 **Tarasiuk A**, Mackiewicz T, Małecka-Panas E, Fichna J. Biomarkers for early detection of pancreatic cancer - miRNAs as a potential diagnostic and therapeutic tool? *Cancer Biol Ther* 2021; **22**: 347-356 [PMID: 34224317 DOI: 10.1080/15384047.2021.1941584]

53 **Schultz NA**, Dehlendorff C, Jensen BV, Bjerregaard JK, Nielsen KR, Bojesen SE, Calatayud D, Nielsen SE, Yilmaz M, Holländer NH, Andersen KK, Johansen JS. MicroRNA biomarkers in whole blood for detection of pancreatic cancer. *JAMA* 2014; **311**: 392-404 [PMID: 24449318 DOI: 10.1001/jama.2013.284664]

54 **Duell EJ**, Lujan-Barroso L, Sala N, Deitz McElyea S, Overvad K, Tjonneland A, Olsen A, Weiderpass E, Busund LT, Moi L, Muller D, Vineis P, Aune D, Matullo G, Naccarati A, Panico S, Tagliabue G, Tumino R, Palli D, Kaaks R, Katzke VA, Boeing H, Bueno-de-Mesquita HBA, Peeters PH, Trichopoulou A, Lagiou P, Kotanidou A, Travis RC, Wareham N, Khaw KT, Ramon Quiros J, Rodríguez-Barranco M, Dorronsoro M, Chirlaque MD, Ardanaz E, Severi G, Boutron-Ruault MC, Rebours V, Brennan P, Gunter M, Scelo G, Cote G, Sherman S, Korc M. Plasma microRNAs as biomarkers of pancreatic cancer risk in a prospective cohort study. *Int J Cancer* 2017; **141**: 905-915 [PMID: 28542740 DOI: 10.1002/ijc.30790]

55 **Khan IA**, Rashid S, Singh N, Singh V, Gunjan D, Das P, Dash NR, Pandey RM, Chauhan SS, Gupta S, Saraya A. Panel of serum miRNAs as potential non-invasive biomarkers for pancreatic ductal adenocarcinoma. *Sci Rep* 2021; **11**: 2824 [PMID: 33531550 DOI: 10.1038/s41598-021-82266-5]

56 **Shams R**, Saberi S, Zali M, Sadeghi A, Ghafouri-Fard S, Aghdaei HA. Identification of potential microRNA panels for pancreatic cancer diagnosis using microarray datasets and bioinformatics methods. *Sci Rep* 2020; **10**: 7559 [PMID: 32371926 DOI: 10.1038/s41598-020-64569-1]

57 **Cao Z**, Liu C, Xu J, You L, Wang C, Lou W, Sun B, Miao Y, Liu X, Wang X, Zhang T, Zhao Y. Plasma microRNA panels to diagnose pancreatic cancer: Results from a multicenter study. *Oncotarget* 2016; **7**: 41575-41583 [PMID: 27223429 DOI: 10.18632/oncotarget.9491]

58 **Almeida PP**, Cardoso CP, de Freitas LM. PDAC-ANN: an artificial neural network to predict pancreatic ductal adenocarcinoma based on gene expression. *BMC Cancer* 2020; **20**: 82 [PMID: 32005189 DOI: 10.1186/s12885-020-6533-0]

59 **Alizadeh Savareh B**, Asadzadeh Aghdaie H, Behmanesh A, Bashiri A, Sadeghi A, Zali M, Shams R. A machine learning approach identified a diagnostic model for pancreatic cancer through using circulating microRNA signatures. *Pancreatology* 2020; **20**: 1195-1204 [PMID: 32800647 DOI: 10.1016/j.pan.2020.07.399]

60 **Yang J**, Xu R, Wang C, Qiu J, Ren B, You L. Early screening and diagnosis strategies of pancreatic cancer: a comprehensive review. *Cancer Commun (Lond)* 2021; **41**: 1257-1274 [PMID: 34331845 DOI: 10.1002/cac2.12204]

61 **Khatri I**, Bhasin MK. A Transcriptomics-Based Meta-Analysis Combined With Machine Learning Identifies a Secretory Biomarker Panel for Diagnosis of Pancreatic Adenocarcinoma. *Front Genet* 2020; **11**: 572284 [PMID: 33133160 DOI: 10.3389/fgene.2020.572284]

62 **Chung WY**, Correa E, Yoshimura K, Chang MC, Dennison A, Takeda S, Chang YT. Using probe electrospray ionization

mass spectrometry and machine learning for detecting pancreatic cancer with high performance. *Am J Transl Res* 2020; **12**: 171-179 [PMID: 32051746]

63 **Ma H**, Liu ZX, Zhang JJ, Wu FT, Xu CF, Shen Z, Yu CH, Li YM. Construction of a convolutional neural network classifier developed by computed tomography images for pancreatic cancer diagnosis. *World J Gastroenterol* 2020; **26**: 5156-5168 [PMID: 32982116 DOI: 10.3748/wjg.v26.i34.5156]

64 **Park S**, Chu LC, Hruban RH, Vogelstein B, Kinzler KW, Yuille AL, Fouladi DF, Shayesteh S, Ghandili S, Wolfgang CL, Burkhart R, He J, Fishman EK, Kawamoto S. Differentiating autoimmune pancreatitis from pancreatic ductal adenocarcinoma with CT radiomics features. *Diagn Interv Imaging* 2020; **101**: 555-564 [PMID: 32278586 DOI: 10.1016/j.diii.2020.03.002]

65 **Si K**, Xue Y, Yu X, Zhu X, Li Q, Gong W, Liang T, Duan S. Fully end-to-end deep-learning-based diagnosis of pancreatic tumors. *Theranostics* 2021; **11**: 1982-1990 [PMID: 33408793 DOI: 10.7150/thno.52508]

66 **Watson MD**, Lyman WB, Passeri MJ, Murphy KJ, Sarantou JP, Iannitti DA, Martinie JB, Vrochides D, Baker EH. Use of Artificial Intelligence Deep Learning to Determine the Malignant Potential of Pancreatic Cystic Neoplasms With Preoperative Computed Tomography Imaging. *Am Surg* 2021; **87**: 602-607 [PMID: 33131302 DOI: 10.1177/0003134820953779]

67 **Zhang Z**, Li S, Wang Z, Lu Y. A Novel and Efficient Tumor Detection Framework for Pancreatic Cancer *via* CT Images. *Annu Int Conf IEEE Eng Med Biol Soc* 2020; **2020**: 1160-1164 [PMID: 33018193 DOI: 10.1109/EMBC44109.2020.9176172]

68 **Batts KP**, Ludwig J. Chronic hepatitis. An update on terminology and reporting. *Am J Surg Pathol* 1995; **19**: 1409-1417 [PMID: 7503362 DOI: 10.1097/00000478-199512000-00007]

69 **Liu KL**, Wu T, Chen PT, Tsai YM, Roth H, Wu MS, Liao WC, Wang W. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. *Lancet Digit Health* 2020; **2**: e303-e313 [PMID: 33328124 DOI: 10.1016/S2589-7500(20)30078-9]

70 **Chu LC**, Park S, Kawamoto S, Fouladi DF, Shayesteh S, Zinreich ES, Graves JS, Horton KM, Hruban RH, Yuille AL, Kinzler KW, Vogelstein B, Fishman EK. Utility of CT Radiomics Features in Differentiation of Pancreatic Ductal Adenocarcinoma From Normal Pancreatic Tissue. *AJR Am J Roentgenol* 2019; **213**: 349-357 [PMID: 31012758 DOI: 10.2214/AJR.18.20901]

71 **Gao X**, Wang X. Performance of deep learning for differentiating pancreatic diseases on contrast-enhanced magnetic resonance imaging: A preliminary study. *Diagn Interv Imaging* 2020; **101**: 91-100 [PMID: 31375430 DOI: 10.1016/j.diii.2019.07.002]

72 **Zhu Z**, Xia Y, Xie L, Fishman EK, Yuille AL. Multi-Scale Coarse-to-Fine Segmentation for Screening Pancreatic Ductal Adenocarcinoma. MICCAI; 2019

73 **Liu SL**, Li S, Guo YT, Zhou YP, Zhang ZD, Lu Y. Establishment and application of an artificial intelligence diagnosis system for pancreatic cancer with a faster region-based convolutional neural network. *Chin Med J (Engl)* 2019; **132**: 2795-2803 [PMID: 31856050 DOI: 10.1097/CM9.0000000000000544]

74 **Chu LC**, Park S, Kawamoto S, Wang Y, Zhou Y, Shen W, Zhu Z, Xia Y, Xie L, Liu F, Yu Q, Fouladi DF, Shayesteh S, Zinreich E, Graves JS, Horton KM, Yuille AL, Hruban RH, Kinzler KW, Vogelstein B, Fishman EK. Application of Deep Learning to Pancreatic Cancer Detection: Lessons Learned From Our Initial Experience. *J Am Coll Radiol* 2019; **16**: 1338-1342 [PMID: 31492412 DOI: 10.1016/j.jacr.2019.05.034]

75 **Fu M**, Wu W, Hong X, Liu Q, Jiang J, Ou Y, Zhao Y, Gong X. Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images. *BMC Syst Biol* 2018; **12**: 56 [PMID: 29745840 DOI: 10.1186/s12918-018-0572-z]

76 **Angulo P**, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, Enders F, Saksena S, Burt AD, Bida JP, Lindor K, Sanderson SO, Lenzi M, Adams LA, Kench J, Therneau TM, Day CP. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007; **45**: 846-854 [PMID: 17393509 DOI: 10.1002/hep.21496]

77 **Kuwahara T**, Hara K, Mizuno N, Okuno N, Matsumoto S, Obata M, Kurita Y, Koda H, Toriyama K, Onishi S, Ishihara M, Tanaka T, Tajika M, Niwa Y. Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Clin Transl Gastroenterol* 2019; **10**: 1-8 [PMID: 31117111 DOI: 10.14309/ctg.0000000000000045]

78 **Marya NB**, Powers PD, Chari ST, Gleeson FC, Leggett CL, Abu Dayyeh BK, Chandrasekhara V, Iyer PG, Majumder S, Pearson RK, Petersen BT, Rajan E, Sawas T, Storm AC, Vege SS, Chen S, Long Z, Hough DM, Mara K, Levy MJ. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. *Gut* 2021; **70**: 1335-1344 [PMID: 33028668 DOI: 10.1136/gutjnl-2020-322821]

79 **Norton ID**, Zheng Y, Wiersema MS, Greenleaf J, Clain JE, Dimagno EP. Neural network analysis of EUS images to differentiate between pancreatic malignancy and pancreatitis. *Gastrointest Endosc* 2001; **54**: 625-629 [PMID: 11677484 DOI: 10.1067/mge.2001.118644]

80 **Ozkan M**, Cakiroglu M, Kocaman O, Kurt M, Yilmaz B, Can G, Korkmaz U, Dandil E, Eksi Z. Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. *Endosc Ultrasound* 2016; **5**: 101-107 [PMID: 27080608 DOI: 10.4103/2303-9027.180473]

81 **Săftoiu A**, Vilmann P, Dietrich CF, Iglesias-Garcia J, Hocke M, Seicean A, Ignee A, Hassan H, Streba CT, Ioncică AM, Gheonea DI, Ciurea T. Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with videos). *Gastrointest Endosc* 2015; **82**: 59-69 [PMID: 25792386 DOI: 10.1016/j.gie.2014.11.040]

82 **Săftoiu A**, Vilmann P, Gorunescu F, Janssen J, Hocke M, Larsen M, Iglesias-Garcia J, Arcidiacono P, Will U, Giovannini M, Dietrich CF, Havre R, Gheorghe C, McKay C, Gheonea DI, Ciurea T; European EUS Elastography Multicentric Study Group. Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. *Clin Gastroenterol Hepatol* 2012; **10**: 84-90.e1 [PMID: 21963957 DOI: 10.1016/j.cgh.2011.09.014]

83 **Tonozuka R**, Itoi T, Nagata N, Kojima H, Sofuni A, Tsuchiya T, Ishii K, Tanaka R, Nagakawa Y, Mukai S. Deep learning analysis for the detection of pancreatic cancer on endosonographic images: a pilot study. *J Hepatobiliary Pancreat Sci*

2021; **28**: 95-104 [PMID: 32910528 DOI: 10.1002/jhbp.825]

84 **Udriștoiu AL**, Cazacu IM, Gruionu LG, Gruionu G, Iacob AV, Burtea DE, Ungureanu BS, Costache MI, Constantin A, Popescu CF, Udriștoiu Ș, Săftoiu A. Real-time computer-aided diagnosis of focal pancreatic masses from endoscopic ultrasound imaging based on a hybrid convolutional and long short-term memory neural network model. *PLoS One* 2021; **16**: e0251701 [PMID: 34181680 DOI: 10.1371/journal.pone.0251701]

85 **Zhang MM**, Yang H, Jin ZD, Yu JG, Cai ZY, Li ZS. Differential diagnosis of pancreatic cancer from normal tissue with digital imaging processing and pattern recognition based on a support vector machine of EUS images. *Gastrointest Endosc* 2010; **72**: 978-985 [PMID: 20855062 DOI: 10.1016/j.gie.2010.06.042]

86 **Zhu M**, Xu C, Yu J, Wu Y, Li C, Zhang M, Jin Z, Li Z. Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images: a diagnostic test. *PLoS One* 2013; **8**: e63820 [PMID: 23704940 DOI: 10.1371/journal.pone.0063820]

87 **Singh DP**, Sheedy S, Goenka AH, Wells M, Lee NJ, Barlow J, Sharma A, Kandlakunta H, Chandra S, Garg SK, Majumder S, Levy MJ, Takahashi N, Chari ST. Computerized tomography scan in pre-diagnostic pancreatic ductal adenocarcinoma: Stages of progression and potential benefits of early intervention: A retrospective study. *Pancreatology* 2020; **20**: 1495-1501 [PMID: 32950386 DOI: 10.1016/j.pan.2020.07.410]

88 **Yu J**, Blackford AL, Dal Molin M, Wolfgang CL, Goggins M. Time to progression of pancreatic ductal adenocarcinoma from low-to-high tumour stages. *Gut* 2015; **64**: 1783-1789 [PMID: 25636698 DOI: 10.1136/gutjnl-2014-308653]

89 **Overbeek KA**, Goggins MG, Dbouk M, Levink IJM, Koopmann BDM, Chuidian M, Konings ICAW, Paiella S, Earl J, Fockens P, Gress TM, Ausems MGEM, Poley JW, Thosani NC, Half E, Lachter J, Stoffel EM, Kwon RS, Stoita A, Kastrinos F, Lucas AL, Syngal S, Brand RE, Chak A, Carrato A, Vleggaar FP, Bartsch DK, van Hooft JE, Cahen DL, Canto MI, Bruno MJ; International Cancer of the Pancreas Screening Consortium. Timeline of Development of Pancreatic Cancer and Implications for Successful Early Detection in High-Risk Individuals. *Gastroenterology* 2022; **162**: 772-785.e4 [PMID: 34678218 DOI: 10.1053/j.gastro.2021.10.014]

90 **Young MR**, Abrams N, Ghosh S, Rinaudo JAS, Marquez G, Srivastava S. Prediagnostic Image Data, Artificial Intelligence, and Pancreatic Cancer: A Tell-Tale Sign to Early Detection. *Pancreas* 2020; **49**: 882-886 [PMID: 32675784 DOI: 10.1097/MPA.0000000000001603]

91 **Springer S**, Masica DL, Dal Molin M, Douville C, Thoburn CJ, Afsari B, Li L, Cohen JD, Thompson E, Allen PJ, Klimstra DS, Schattner MA, Schmidt CM, Yip-Schneider M, Simpson RE, Fernandez-Del Castillo C, Mino-Kenudson M, Brugge W, Brand RE, Singhi AD, Scarpa A, Lawlor R, Salvia R, Zamboni G, Hong SM, Hwang DW, Jang JY, Kwon W, Swan N, Geoghegan J, Falconi M, Crippa S, Doglioni C, Paulino J, Schulick RD, Edil BH, Park W, Yachida S, Hijioka S, van Hooft J, He J, Weiss MJ, Burkhart R, Makary M, Canto MI, Goggins MG, Ptak J, Dobbyn L, Schaefer J, Sillman N, Popoli M, Klein AP, Tomasetti C, Karchin R, Papadopoulos N, Kinzler KW, Vogelstein B, Wolfgang CL, Hruban RH, Lennon AM. A multimodality test to guide the management of patients with a pancreatic cyst. *Sci Transl Med* 2019; **11** [PMID: 31316009 DOI: 10.1126/scitranslmed.aav4772]

92 **Yang Z**, Zhang L, Zhang M, Feng J, Wu Z, Ren F, Lv Y. Pancreas Segmentation in Abdominal CT Scans using Inter-/Intra-Slice Contextual Information with a Cascade Neural Network. *Annu Int Conf IEEE Eng Med Biol Soc* 2019; **2019**: 5937-5940 [PMID: 31947200 DOI: 10.1109/EMBC.2019.8856774]

93 **Barat M**, Chassagnon G, Dohan A, Gaujoux S, Coriat R, Hoeffel C, Cassinotto C, Soyer P. Artificial intelligence: a critical review of current applications in pancreatic imaging. *Jpn J Radiol* 2021; **39**: 514-523 [PMID: 33550513 DOI: 10.1007/s11604-021-01098-5]

Artificial Intelligence in Medical Imaging

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

Artif Intell Med Imaging 2022 April 28; 3(2): 33-41

DOI: 10.35711/aimi.v3.i2.33

ISSN 2644-3260 (online)

MINIREVIEWS

Artificial intelligence: Advances and new frontiers in medical imaging

Marc R Fromherz, Mina S Makary

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, C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Dabbakuti JRKKK, India; Liu Y, China; Wan XH, China

Received: December 20, 2021
Peer-review started: December 20, 2021
First decision: January 26, 2022
Revised: February 20, 2022
Accepted: April 21, 2022
Article in press: April 21, 2022
Published online: April 28, 2022

Marc R Fromherz, Department of Radiology, The Ohio State University Wexner Medical Center, Columbus, OH 43210, United States

Mina S Makary, Division of Vascular and Interventional Radiology, Department of Radiology, The Ohio State University Wexner Medical Center, Columbus, OH 43210, United States

Corresponding author: Mina S Makary, MD, Assistant Professor, Attending Doctor, Director, Division of Vascular and Interventional Radiology, Department of Radiology, The Ohio State University Wexner Medical Center, 395 W. 12th Ave 4th Floor Faculty Office Tower, Columbus, OH 43210, United States. mina.makary@osumc.edu

Abstract

Artificial intelligence (AI) has been entwined with the field of radiology ever since digital imaging began replacing films over half a century ago. These algorithms, ranging from simplistic speech-to-text dictation programs to automated interpretation neural networks, have continuously sought to revolutionize medical imaging. With the number of imaging studies outpacing the amount of trained of readers, AI has been implemented to streamline workflow efficiency and provide quantitative, standardized interpretation. AI relies on massive amounts of data for its algorithms to function, and with the wide-spread adoption of Picture Archiving and Communication Systems (PACS), imaging data is accumulating rapidly. Current AI algorithms using machine-learning technology, or computer aided-detection, have been able to successfully pool this data for clinical use, although the scope of these algorithms remains narrow. Many systems have been developed to assist the workflow of the radiologist through PACS optimization and imaging study triage, however interpretation has generally remained a human responsibility for now. In this review article, we will summarize the current successes and limitations of AI in radiology, and explore the exciting prospects that deep-learning technology offers for the future.

Key Words: Artificial intelligence; Machine-learning; Deep-learning; Radiology workflow; Image interpretation

Core Tip: Artificial intelligence (AI) has been an increasingly publicized subject in the field of radiology. This review will attempt to summarize the evolving philosophy and mechanisms behind the AI movement as well as the current applications, limitations, and future directions of the field.

Citation: Fromherz MR, Makary MS. Artificial intelligence: Advances and new frontiers in medical imaging. *Artif Intell Med Imaging* 2022; 3(2): 33-41
URL: https://www.wjgnet.com/2644-3260/full/v3/i2/33.htm
DOI: https://dx.doi.org/10.35711/aimi.v3.i2.33

INTRODUCTION

Advancements in artificial intelligence (AI) technology have created a stir of excitement—and trepidation—amongst professionals in radiology. With the advent of concepts such as machine learning and artificial neural networks promising instantaneous and accurate image interpretation, AI has been heralded as the next step in radiology evolution[1,2]. The ability to reduce image interpretation time and increase detection to levels beyond what is possible for the human eye could create a revolutionary, and increasingly necessary, impact on patient care across all medical disciplines.

AI in radiology has focused on improving three broad principles attributed to human limitations; efficiency, objectivity, and standardization[1,2,3]. Over the past few years there has been a continual increase in imaging orders, and it has been estimated that a radiologist must interpret an image every 3-4 s to match the demand[3,4] This demand, combined with declining reimbursement, has put more pressure on radiologists to increase productivity[5]. Additionally, human and health system variability has long been seen as a potential target to improve standardization across the field. Depending on who the reader is, what hospital system they work for, the time of day, and the number of scans the radiologist has read can result in measurable discrepancies in accuracy and timeliness of image interpretation[3,6,7].

Despite the exciting potential of AI utilization, the fear of algorithms replacing radiologists is ever present. AI companies have grown at an astonishing rate, with 60 new Food and Drug Administration (FDA) approved products in 2020, however the once foreseen AI takeover has not yet manifested[8-10]. Nonetheless, AI is making an impact, just not in the way it was originally planned. A fundamental shift has occurred in recent years in AI implementation, scope, and underlying philosophy. The idea of "replacing radiologists" is not a viable next step in AI evolution, at least for now, and the new philosophy of "working with radiologists" is one that is rapidly gaining traction[11,12]. By examining the current utilizations and limitations of AI in radiology, we can recognize the importance of this fast-rising technology and where the interaction between human and machine may be headed in the future.

CURRENT AI UTILIZATION IN RADIOLOGY

The current state of AI utilization in the field of radiology is variable based on institution, although there are several widely-adopted systems. Aligning with the newer philosophy of "working with radiologists", many of the current AI systems are being used in a limited capacity as tools to enhance the radiologist's workflow. Many of these AI systems fall under the category of "micro-optimizations" [13].

The primary goal for micro-optimization algorithms is to assist the radiologist in his or her daily tasks rather than fully automating the radiologic process. Micro-optimizations can be broken down into two categories; nonpixel-based optimizations and pixel-based optimizations. By using AI to streamline the efficiency and standardization of time-consuming, mundane, or non-interpretive tasks, radiologists can better allocate their time and energy to further focus on image interpretation, consultation, and overall patient care[3,4,14]. Table 1 provides a summary of AI applications for both nonpixel-based and pixel-based optimizations.

Nonpixel-based optimizations

Nonpixel-based optimizations refers to AI assistance in tasks that are not directly related to image interpretation. Some of these tasks include triaging patients, Picture Archiving and Communication Systems (PACS) optimizations, and standardized reporting. As an example, to better triage patients for immediate interpretation AI systems are currently being tested for risk stratification in patients with possible aortic dissection or aneurysm rupture[15,16]. As a different example, through big data analysis, AI algorithms have started to tackle the issue of automated image protocol creations. By reviewing imaging study requests, AI can determine if the study is appropriate, if another study may be more appropriate, or if contrast is necessary or not. With the ability to automatically mine the electronic

Table 1 Areas of radiology workflow with current artificial intelligence implementation

Workflow target	Application examples
Nonpixel-based	
Triage	Risk stratification for aortic pathology and generation of 'aortic calcification score' to assess for disease severity[15,16]
PACS display	Automated hanging protocol and comparison image generation[11]
Order verification	Patient medical record mining with built-in appropriateness criteria guidelines to approve or flag study orders[17,18,19,20,21]
Reporting	Automated data insertion into templates for standardized reporting of chest radiograph findings[23,24]
Pixel-based	
Segmentation	Segmentation of simple lung nodules on chest CT images[43]
Disease registration	PI-RADS lesion classification based on MRI image characteristics[25,26]
Screening	Algorithmic interpretation and classification of screening mammograms[27,28]

PACS: Picture Archiving and Communication Systems; CT: Computed tomography; MRI: magnetic resonance imaging.

medical record system and compare it to established guidelines, the system can then make the appropriate recommendation[17-19]. With an estimated 10% of all imaging studies being ordered in error, these nonpixel-based algorithms can automatically detect and eliminate erroneous study orders [20,21].

The automatic generation of hanging protocols and standardized screen display is another target for optimization. Before data interpretation can commence, a radiologist can spend 10-60 s selecting the appropriate images for comparison[11]. By having the appropriate hanging protocol and display automatically generate, image interpretation can commence instantaneously. What may at first seem like an insignificant amount of time, the elimination of manual protocol selection can significantly improve efficiency and allow for the redirection of the radiologist's brain power toward actual diagnostic interpretation[11].

The standardization of reporting is one of the final areas for optimization, and one that is becoming increasingly necessary among all medical specialties in order to efficiently navigate and report in the electronic medical systems. Reporting is the final step in the radiologist's workflow, and it is also one of the most error-prone[22]. Many micro-optimization AI algorithms are working on increasing the efficiency of reporting through the creation of automatic report generation tools including pre-selected formats specific for the study and automatic annotation. Automating and standardizing reporting can optimize radiologists' reimbursements and save time, as demonstrated by one current chest x-ray reporting algorithm that saved radiologists an average of 8.5 h per month[23,24].

Pixel-based optimizations

While the importance of these nonpixel-based micro-optimizations cannot be understated, the prospect of instantaneous image interpretation is the ultimate ambition of AI. Although AI technology has not yet achieved this ability in a broad sense, the development of pixel-based micro-optimizations have been paramount in maximizing a radiologist's workflow efficiency[14]. Some example applications of these systems involve image segmentation, reconstruction, and disease registration.

AI segmentation has the ability to automatically delineate structures and provide measurements such as organ volume or the surface area of a tumor. Taken a step further, these AI algorithms can be specialized to stage tumors and provide pre-interpreted read-outs such as PI-RADS scores for prostate cancer staging[25,26]. A study by Sanford *et al*[25] demonstrated a modest 40% agreement between an AI algorithm and an expert radiologist when assigning PI-RADS scores based on magnetic resonance imaging (MRI). This result was comparable with previous human inter-reader agreements. Automated segmentation and pre-interpreted read-outs may be maximally utilized in areas that have the most amount of data, such as screening imaging studies.

Utilizing AI for screening processes helps to reduce the workload for radiologists while not over-extending the abilities of AI. As the typical screen produces categorically "positive", "negative", or "inconclusive" results, the complexity of the AI reads can be minimized. Using machine learning for screening detection is referred to as computer aided detection (CADe). CADe is currently being used in screening mammography, where there is an abundance of imaging studies and a relatively disproportionate amount of mammography trained readers[1,2,27]. CADe highlights the area of interest, and it is then determined whether an additional diagnostic study is indicated. CADe for mammography has been around since 1998 and its implementation into clinical workflow has continued to increase allowing radiologists to read more screening studies in less time. Along with the decreased read-time, it should be noted that several studies comparing the accuracy of CADe mammography to traditional radiologist-read mammograms have shown no discernable difference[26]. In one such study, an

ensemble of top-performing AI algorithms combined with a single radiologist reader achieved an area under the curve (AUC) of 0.942, with 92% specificity, outperforming the radiologists' specificity of 90.5%[28]. This is a representative example of new AI algorithms geared toward instantaneous, automatic interpretation.

LIMITATIONS

Despite the constant development of new AI companies, advanced algorithms, and enhanced learning technology, AI has not yet become mainstream in the radiology world due to a combination of both logistical and clinical challenges. The ease of which AI programs can be implemented varies widely based on the scope and technicalities of the clinical problem they aim to solve, as well as the mechanism by which they solve them. In general terminology, there are two main types of AI systems, machine-learning and deep-learning, each of with have some specific limitations of their own[1,29].

Machine-learning AI

Machine-learning functions largely on the principal of pattern recognition. If the machine is able to "see" enough example image characteristics of a certain disease, it can then look at new images and be able to recognize them based on those previously defined features. The caveat here, is that these "pre-defined features", such as tumor volume, density, *etc*., must be hand-fed into each specific machine-learning classifier[3]. In this way the AI does not actually learn, but rather applies the specifics of its pre-engineered programming. Consequently, machine-learning AI is intrinsically limited by these specific characteristics which can reduce its ability to recognize image features, such as rare or unusual disease presentations[30,31]. Figure 1 demonstrates a schematic example of how machine-learning AI systems utilize these pre-defined features for classification. Furthermore, as the breadth of medical knowledge continues to expand, previous CAD systems may become outdated, and therefore obsolete[30]. The theoretical solution to these hard-wired restrictions is the use of AI algorithms that do not rely on pre-engineered feature recognition, but rather one that can learn and adapt in a manner similar to the human brain.

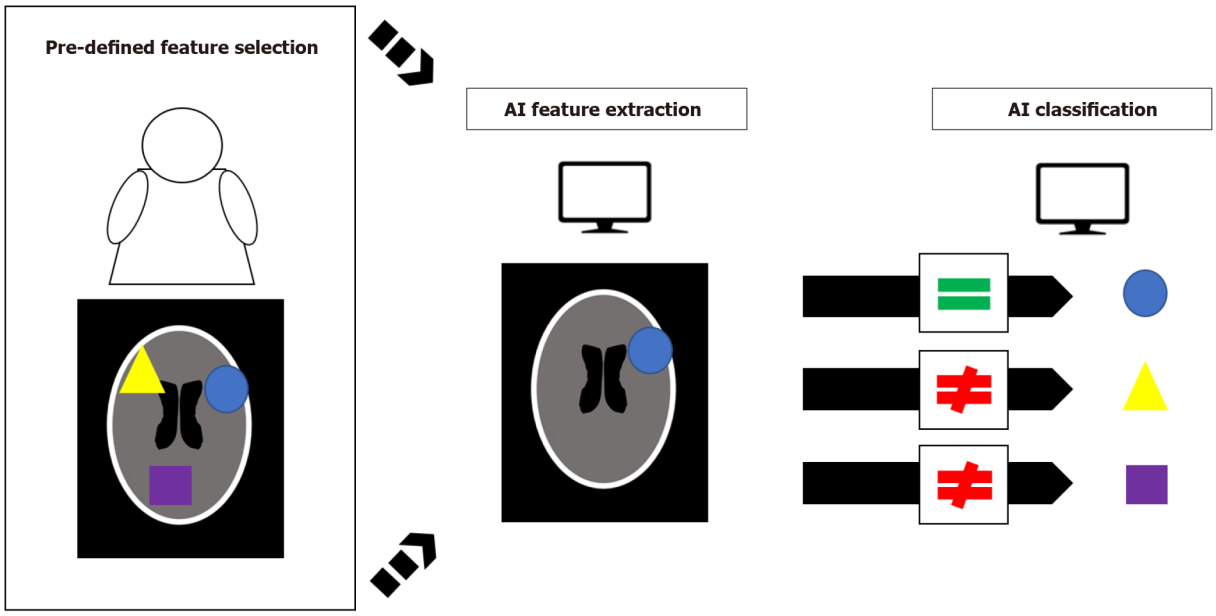
Deep-learning AI

Deep-learning is programmed to mimic the pattern of neural networks such as those in the human brain, referred to in the literature as convolutional neural networks (CNNs). The principal mechanism behind AI algorithms relies on a vast quantity of data, and through this data the AI can develop its own pattern of feature recognition without the need for pre-programming from human experts. Deep-learning AI uses these features to create connections and draw conclusions in a way similar to the human brain, and allowing it to operate freely from human input thus increasing its automaticity and decreasing restrictions[3,32,33]. While in theory this method appears to be a step-up from classical machine-learning technology, the reliance on data and complexity of the mechanism has its limitations.

With the wide-implementation of PACS and an ever-increasing number of medical images, there is no shortage of data for AI algorithms to mine[34]. The issue is not quantity—but quality. Different PACS, different imaging machine manufacturers, and different protocols can all effect the generalizability of an AI algorithm. These variations in image reconstruction, segmentation, and labelling can have adverse effects on the AI's ability to learn, and the process of standardization across these variables would be a time-consuming and expensive task. This is one of the reasons for the current narrow use of AI in clinical practice. Currently approved AI programs only function with specific computed tomography (CT) imager models, specific PAC systems, and specific disease processes. With such a narrow clinical window, AI in its current form is limited in scope[30,31]. If multiple different AI systems are needed for each specific pathology the process of creating and implementing these systems may not be fiscally feasible[35]. Even with implementation, a lapse in the detection of rare diseases would still exist.

Industry acceptance

Questions regarding the mechanism of how deep-learning functions can also create additional limitations, specifically regarding FDA approval and the accuracy of the AI's results[8,36]. The mechanism is extremely complex, and in many instances, the exact way in which the AI forms these CNNs is either unknown or proprietary. If the way the AI algorithm functions to produce its results is not well understood, this begs the question of whether or not its results can be trusted[8,36,37]. This question has haunted AI since its inception, and the answer of whether or not health professionals and patients would be willing to put their faith in the recommendation of a 100% computer-controlled radiologic study is not an easy one to answer. A variety of comparison studies have been conducted to determine whether AI accuracy is comparable to that of human readers, and the results have been mixed.



DOI: 10.35711/aimi.v3.i2.33

Figure 1 Machine-learning requires pre-defined feature inputs which are then extracted in order to classify target image characteristics.
AI: Artificial intelligence.

In the previously mentioned Schaffter *et al*[28] study on breast cancer detection, no single AI algorithm was able to outperform the radiologists, with a specificity of 66.1% for the top-performing algorithm compared to 90.5% for the radiologists. In a breast cancer detection study using a different AI system, the AI outperformed the radiologists with an AUC of 0.740 compared to the radiologists' AUC of 0.625[38]. In a study comparing chest radiograph interpretation, AI outperformed the radiologists on detection of pulmonary edema, underperformed on detection of consolidation, and had comparable performance for detection of pleural effusions[39]. These studies collectively demonstrate that AI systems have mixed performance compared to human radiologists.

The utilization of different algorithms, training datasets, and radiologist experience in these studies makes drawing conclusions about AI's general trustworthiness difficult. Concerns such as these are why the shift toward micro-optimizations has been an attractive one for the interim, however as new technologies are developed and deep-learning systems are polished the future of AI continues to push the boundaries of possibility.

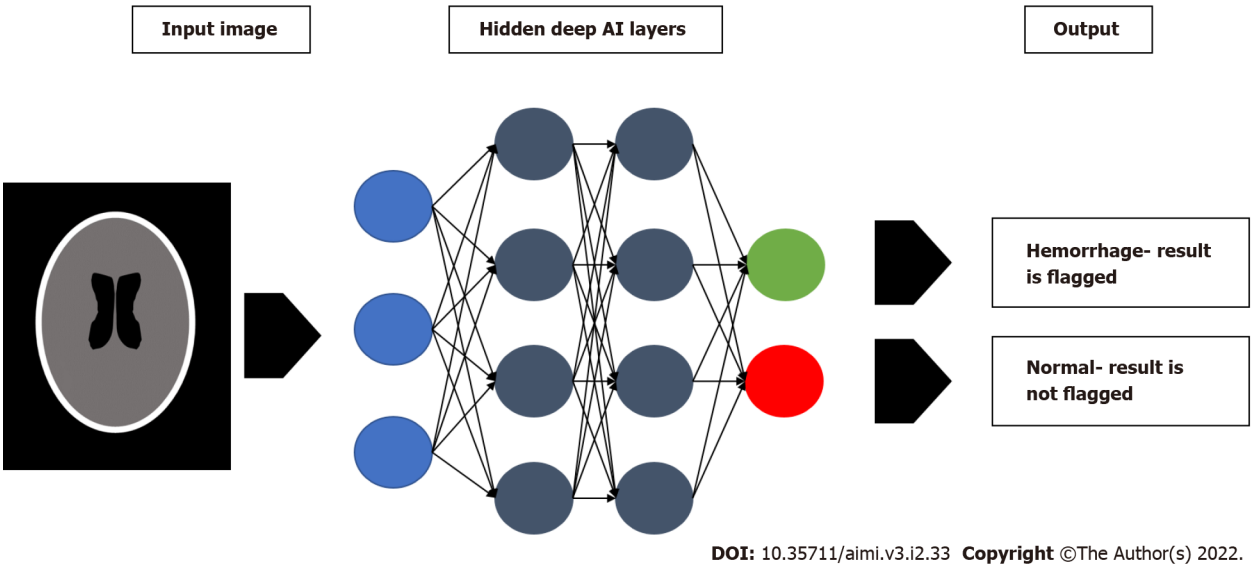
FUTURE DIRECTIONS

The future of AI in radiology is constantly evolving, and with new computer systems, implementation targets, and algorithms being developed seemingly by the day there is no discernable end to what is possible[8-10]. Within PACS, the utilization of deep learning AI could theoretically be implemented wherever large quantities of data are available, although as previously stated there are several limitations to deep learning technology. With the interconnectivity, digitization, and increasing data pool in modern radiology, the limitations of deep-learning may slowly start to be overcome, and the use of micro-optimization may ramp up in scale.

The next phase in AI utilization will likely continue the trend of micro-optimization, but with increased efficiency. As hospital systems become more integrated, with imaging devices and PACS being able to directly communicate with each other, it would only make sense that the AI algorithms within these systems do the same. With AI's current narrow clinical usage, each system excels at only one specific task[30,31]. By combining these systems, the scope of each can be summated into a larger, more efficient system. For example a lung cancer screening CT reconstruction algorithm could be used alongside a hanging protocol algorithm, with CADe for detection, and another algorithm for report generation[40]. Until a more encompassing system is created, combining existing micro-optimizations can scale efficiency in clinical workflow.

Disease recognition and triage

Despite the profound promise of deep learning, it has yet to have seen wide-spread clinical utilization. That being said, the power behind deep learning is data and the amount of available data is continuously growing. As we gather more high-quality data, the deep learning systems should become



DOI: 10.35711/aimi.v3.i2.33

Figure 2 Deep-learning artificial intelligence application in triaging head computed tomography images. The input image characteristics are extracted and analyzed by the convolutional neural network to create an output. The output is then flagged or not flagged depending on the algorithm's interpretation. AI: Artificial intelligence.

more powerful, increasing their usage potential. The full potential of deep learning is still unknown, however there are several promising applications in detection and automated disease monitoring. One of these applications is in the identification of incidental findings. When a radiologist is examining a trauma study, the AI system can detect incidental pulmonary nodules, allowing the radiologist to focus on the primary clinical issue without overlooking other findings[41,42,43]. Looking to improve upon current CAD systems, utilizing deep learning AI for triage is another attractive target, where the urgency of a given study is prioritized and then sent to a radiologist for final interpretation. These algorithms pool hundreds of thousands of imaging studies along with their subsequent reports, and use this information to train their CNNs. In a study of one such algorithm on assigning priority to adult chest radiographs, AI was able to assign priority with a sensitivity of 71% and a specificity of 95%. Importantly, the time taken to report critical findings was reduced significantly from 11.2 to 2.7[32]. Another study on triaging patients based on head CT findings produced similar results, with an AUC of 0.92 for accurately detecting intracranial hemorrhage[44]. Figure 2 is schematic example demonstrating this type of AI triage system. The ability for the system to distinguish between 'normal' and 'abnormal' accurately, and then further stratify 'abnormal' into severity categories, is a promising step toward automated interpretation[32,44].

Disease monitoring

The prospect of monitoring disease progression is a more complicated one, but the ability of the deep learning system to accumulate and track data changes over time makes this an attractive target. These systems may also have the ability to automatically adjust for changes in patient position or body habitus at the times the studies were conducted[3]. One of the obvious applications for this is oncology, with AI models already demonstrating their ability to accurately measure therapeutic response and tumor recurrence[45,46]. Throughout the coronavirus disease 2019 (COVID-19) pandemic, the ability to track disease progression has been crucial for medical decision making. Unfortunately, the wide variability in an individual's disease course has been difficult to predict. To solve this problem, several deep learning systems have been tested to identify minute chest CT changes based on quantitative pixel analysis, giving us a more sophisticated look into the pathophysiology of the disease[47-49]. Not only does this present the potential to make educated decisions for COVID-19 patients regarding the need for hospitalization and allocation of resources, but the pandemic in general has further stressed the need of increased efficiency in radiology during times of unprecedented volume.

CONCLUSION

As the role of AI in radiology continues to advance and diversify, the potential for revolutionary clinical impact persists. One of the most important factors for the continued development of AI in radiology is achieving wide-spread implementation, and to achieve this AI must be embraced by radiologists. Currently, only an estimated 30% of radiologists use AI in day-to-day workflow[50]. With the shift of AI philosophy away from replacing radiologists, the view of AI as a threat to fear may be replaced with its

view as a tool to exploit. As more algorithms are approved, more studies published, and more systems implemented into clinical practice, radiologists and trainees alike need to educate themselves on what AI can do for them and their patients. When radiologists and AI learn to work together, the potential clinical benefits of a human-machine symbiosis can be fully realized.

FOOTNOTES

Author contributions: All authors contributed to the writing and preparation of the manuscript.

Conflict-of-interest statement: The authors declare they have no conflicts of interest.

Country/Territory of origin: United States

ORCID number: Marc R Fromherz 0000-0002-0458-0749; Mina S Makary 0000-0002-2498-7132.

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

REFERENCES

1 **Oakden-Rayner L**. The Rebirth of CAD: How Is Modern AI Different from the CAD We Know? *Radiol Artif Intell* 2019; **1**: e180089 [PMID: 33937793 DOI: 10.1148/ryai.2019180089]
2 **Mun SK**, Wong KH, Lo SB, Li Y, Bayarsaikhan S. Artificial Intelligence for the Future Radiology Diagnostic Service. *Front Mol Biosci* 2020; **7**: 614258 [PMID: 33585563 DOI: 10.3389/fmolb.2020.614258]
3 **Hosny A**, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer* 2018; **18**: 500-510 [PMID: 29777175 DOI: 10.1038/s41568-018-0016-5]
4 **Zha N**, Patlas MN, Duszak R Jr. Radiologist Burnout Is Not Just Isolated to the United States: Perspectives From Canada. *J Am Coll Radiol* 2019; **16**: 121-123 [PMID: 30236858 DOI: 10.1016/j.jacr.2018.07.010]
5 **Stempniak M**. "Grave Concern": Radiologist Reimbursement Expected to Plummet after CMS Clinical Labor Wage Update, August 16, 2021. Available from: https://www.radiologybusiness.com/topics/economics/radiologist-reimbursement-cms-clinical-labor-wage
6 **Patel AG**, Pizzitola VJ, Johnson CD, Zhang N, Patel MD. Radiologists Make More Errors Interpreting Off-Hours Body CT Studies during Overnight Assignments as Compared with Daytime Assignments. *Radiology* 2020; **297**: E281 [PMID: 33074783 DOI: 10.1148/radiol.2020209018]
7 **Hanna TN**, Lamoureux C, Krupinski EA, Weber S, Johnson JO. Effect of Shift, Schedule, and Volume on Interpretive Accuracy: A Retrospective Analysis of 2.9 Million Radiologic Examinations. *Radiology* 2018; **287**: 205-212 [PMID: 29156150 DOI: 10.1148/radiol.2017170555]
8 **Benjamens S**, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med* 2020; **3**: 118 [PMID: 32984550 DOI: 10.1038/s41746-020-00324-0]
9 **Chen MM**, Golding LP, Nicola GN. Who Will Pay for AI? *Radiol Artif Intell* 2021; **3**: e210030 [PMID: 34142090 DOI: 10.1148/ryai.2021210030]
10 **Bloom J**, Dyrda L. 100+ artificial intelligence companies to know in healthcare. Beckers Hospital Review. Published July 19, 2019. Accessed August 4, 2020. Available from: https://www.beckershospitalreview.com/Lists/100-artificial-intelligence-companies-to-know-in-healthcare-2019.html
11 **Langlotz CP**. Will Artificial Intelligence Replace Radiologists? *Radiol Artif Intell* 2019; **1**: e190058 [PMID: 33937794 DOI: 10.1148/ryai.2019190058]
12 **Siwicki B**. Mass General Brigham and the Future of AI in Radiology. Healthcare IT News, May 10, 2021. Available from: https://www.healthcareitnews.com/news/mass-general-brigham-and-future-ai-radiology
13 **Gimenez F**. Real-Time Interpretation: The next Frontier in Radiology AI - MedCity News. Medcitynews, July 22, 2022. Available from: https://medcitynews.com/2021/07/real-time-interpretation-the-next-frontier-in-radiology-ai/
14 **Do HM**, Spear LG, Nikpanah M, Mirmomen SM, Machado LB, Toscano AP, Turkbey B, Bagheri MH, Gulley JL, Folio LR. Augmented Radiologist Workflow Improves Report Value and Saves Time: A Potential Model for Implementation of Artificial Intelligence. *Acad Radiol* 2020; **27**: 96-105 [PMID: 31818390 DOI: 10.1016/j.acra.2019.09.014]
15 **Hahn LD**, Baeumler K, Hsiao A. Artificial intelligence and machine learning in aortic disease. *Curr Opin Cardiol* 2021; **36**: 695-703 [PMID: 34369401 DOI: 10.1097/HCO.0000000000000903]
16 **Chin CW**, Pawade TA, Newby DE, Dweck MR. Risk Stratification in Patients With Aortic Stenosis Using Novel Imaging Approaches. Circ Cardiovasc Imaging 2015; 8: e003421 [PMID: 26198161 DOI: 10.1161/CIRCIMAGING.115.003421]

17 **Desai V**, Flanders A, Zoga AC. Leveraging Technology to Improve Radiology Workflow. *Semin Musculoskelet Radiol* 2018; **22**: 528-539 [PMID: 30399617 DOI: 10.1055/s-0038-1673385]

18 **Hassanpour S**, Langlotz CP. Information extraction from multi-institutional radiology reports. *Artif Intell Med* 2016; **66**: 29-39 [PMID: 26481140 DOI: 10.1016/j.artmed.2015.09.007]

19 **Bhatia N**, Trivedi H, Safdar N, Heilbrun ME. Artificial Intelligence in Quality Improvement: Reviewing Uses of Artificial Intelligence in Noninterpretative Processes from Clinical Decision Support to Education and Feedback. *J Am Coll Radiol* 2020; **17**: 1382-1387 [PMID: 33153542 DOI: 10.1016/j.jacr.2020.08.002]

20 **Bernardy M**, Ullrich CG, Rawson JV, Allen B Jr, Thrall JH, Keysor KJ, James C, Boyes JA, Saunders WM, Lomers W, Mollura DJ, Pyatt RS Jr, Taxin RN, Mabry MR. Strategies for managing imaging utilization. *J Am Coll Radiol* 2009; **6**: 844-850 [PMID: 19945039 DOI: 10.1016/j.jacr.2009.08.003]

21 **Lehnert BE**, Bree RL. Analysis of appropriateness of outpatient CT and MRI referred from primary care clinics at an academic medical center: how critical is the need for improved decision support? *J Am Coll Radiol* 2010; **7**: 192-197 [PMID: 20193924 DOI: 10.1016/j.jacr.2009.11.010]

22 **Onder O**, Yarasir Y, Azizova A, Durhan G, Onur MR, Ariyurek OM. Errors, discrepancies and underlying bias in radiology with case examples: a pictorial review. *Insights Imaging* 2021; **12**: 51 [PMID: 33877458 DOI: 10.1186/s13244-021-00986-8]

23 **Lakhani P**, Prater AB, Hutson RK, Andriole KP, Dreyer KJ, Morey J, Prevedello LM, Clark TJ, Geis JR, Itri JN, Hawkins CM. Machine Learning in Radiology: Applications Beyond Image Interpretation. *J Am Coll Radiol* 2018; **15**: 350-359 [PMID: 29158061 DOI: 10.1016/j.jacr.2017.09.044]

24 **Chung CY**, Makeeva V, Yan J, Prater AB, Duszak R Jr, Safdar NM, Heilbrun ME. Improving Billing Accuracy Through Enterprise-Wide Standardized Structured Reporting With Cross-Divisional Shared Templates. *J Am Coll Radiol* 2020; **17**: 157-164 [PMID: 31918874 DOI: 10.1016/j.jacr.2019.08.034]

25 **Sanford T**, Harmon SA, Turkbey EB, Kesani D, Tuncer S, Madariaga M, Yang C, Sackett J, Mehralivand S, Yan P, Xu S, Wood BJ, Merino MJ, Pinto PA, Choyke PL, Turkbey B. Deep-Learning-Based Artificial Intelligence for PI-RADS Classification to Assist Multiparametric Prostate MRI Interpretation: A Development Study. *J Magn Reson Imaging* 2020; **52**: 1499-1507 [PMID: 32478955 DOI: 10.1002/jmri.27204]

26 **Winkel DJ**, Wetterauer C, Matthias MO, Lou B, Shi B, Kamen A, Comaniciu D, Seifert HH, Rentsch CA, Boll DT. Autonomous Detection and Classification of PI-RADS Lesions in an MRI Screening Population Incorporating Multicenter-Labeled Deep Learning and Biparametric Imaging: Proof of Concept. Diagnostics (Basel) 2020; 10 [PMID: 33202680 DOI: 10.3390/diagnostics10110951]

27 **Raya-Povedano JL**, Romero-Martín S, Elías-Cabot E, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M. AI-based Strategies to Reduce Workload in Breast Cancer Screening with Mammography and Tomosynthesis: A Retrospective Evaluation. *Radiology* 2021; **300**: 57-65 [PMID: 33944627 DOI: 10.1148/radiol.2021203555]

28 **Schaffter T**, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, Lotter W, Jie Z, Du H, Wang S, Feng J, Feng M, Kim HE, Albiol F, Albiol A, Morrell S, Wojna Z, Ahsen ME, Asif U, Jimeno Yepes A, Yohanandan S, Rabinovici-Cohen S, Yi D, Hoff B, Yu T, Chaibub Neto E, Rubin DL, Lindholm P, Margolies LR, McBride RB, Rothstein JH, Sieh W, Ben-Ari R, Harrer S, Trister A, Friend S, Norman T, Sahiner B, Strand F, Guinney J, Stolovitzky G; and the DM DREAM Consortium, Mackey L, Cahoon J, Shen L, Sohn JH, Trivedi H, Shen Y, Buturovic L, Pereira JC, Cardoso JS, Castro E, Kalleberg KT, Pelka O, Nedjar I, Geras KJ, Nensa F, Goan E, Koitka S, Caballero L, Cox DD, Krishnaswamy P, Pandey G, Friedrich CM, Perrin D, Fookes C, Shi B, Cardoso Negrie G, Kawczynski M, Cho K, Khoo CS, Lo JY, Sorensen AG, Jung H. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Netw Open* 2020; **3**: e200265 [PMID: 32119094 DOI: 10.1001/jamanetworkopen.2020.0265]

29 **Do S**, Song KD, Chung JW. Basics of Deep Learning: A Radiologist's Guide to Understanding Published Radiology Articles on Deep Learning. *Korean J Radiol* 2020; **21**: 33-41 [PMID: 31920027 DOI: 10.3348/kjr.2019.0312]

30 **Futoma J**, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit Health* 2020; **2**: e489-e492 [PMID: 32864600 DOI: 10.1016/S2589-7500(20)30186-2]

31 **Zech JR**, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med* 2018; **15**: e1002683 [PMID: 30399157 DOI: 10.1371/journal.pmed.1002683]

32 **Annarumma M**, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated Triaging of Adult Chest Radiographs with Deep Artificial Neural Networks. *Radiology* 2019; **291**: 196-202 [PMID: 30667333 DOI: 10.1148/radiol.2018180921]

33 **Shen D**, Wu G, Suk HI. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* 2017; **19**: 221-248 [PMID: 28301734 DOI: 10.1146/annurev-bioeng-071516-044442]

34 **Barlow RD**. Larger Volume Data Sets Redefining PACS. Imaging Technology News, June 12, 2008. Available from: http://www.itnonline.com/article/Larger-volume-data-sets-redefining-pacs

35 **Tadavarthi Y**, Vey B, Krupinski E, Prater A, Gichoya J, Safdar N, Trivedi H. The State of Radiology AI: Considerations for Purchase Decisions and Current Market Offerings. *Radiol Artif Intell* 2020; **2**: e200004 [PMID: 33937846 DOI: 10.1148/ryai.2020200004]

36 **Aggarwal R**, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, Ashrafian H, Darzi A. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med* 2021; **4**: 65 [PMID: 33828217 DOI: 10.1038/s41746-021-00438-z]

37 **Chartrand G**, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, Kadoury S, Tang A. Deep Learning: A Primer for Radiologists. *Radiographics* 2017; **37**: 2113-2131 [PMID: 29131760 DOI: 10.1148/rg.2017170077]

38 **McKinney SM**, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GS, Darzi A, Etemadi M, Garcia-Vicente F, Gilbert FJ, Halling-Brown M, Hassabis D, Jansen S, Karthikesalingam A, Kelly CJ, King D, Ledsam JR, Melnick D, Mostofi H, Peng L, Reicher JJ, Romera-Paredes B, Sidebottom R, Suleyman M, Tse D, Young KC, De Fauw J, Shetty S. International evaluation of an AI system for breast cancer screening. *Nature* 2020; **577**: 89-94 [PMID: 31894144 DOI: 10.1038/s41586-019-1799-6]

39 **Wu JT**, Wong KCL, Gur Y, Ansari N, Karargyris A, Sharma A, Morris M, Saboury B, Ahmad H, Boyko O, Syed A,

Jadhav A, Wang H, Pillai A, Kashyap S, Moradi M, Syeda-Mahmood T. Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents. *JAMA Netw Open* 2020; **3**: e2022779 [PMID: 33034642 DOI: 10.1001/jamanetworkopen.2020.22779]

40 **Svoboda E**. Artificial intelligence is improving the detection of lung cancer. *Nature* 2020; **587**: S20-S22 [PMID: 33208974 DOI: 10.1038/d41586-020-03157-9]

41 **Cui S**, Ming S, Lin Y, Chen F, Shen Q, Li H, Chen G, Gong X, Wang H. Development and clinical application of deep learning model for lung nodules screening on CT images. *Sci Rep* 2020; **10**: 13657 [PMID: 32788705 DOI: 10.1038/s41598-020-70629-3]

42 **Ottawa (ON): Canadian Agency for Drugs and Technologies in Health**. Artificial Intelligence for Classification of Lung Nodules: A Review of Clinical Utility, Diagnostic Accuracy, Cost-Effectiveness, and Guidelines [Internet]. 2020-Jan-22 [PMID: 33074628]

43 **Weikert T**, Akinci D'Antonoli T, Bremerich J, Stieltjes B, Sommer G, Sauter AW. Evaluation of an AI-Powered Lung Nodule Algorithm for Detection and 3D Segmentation of Primary Lung Tumors. *Contrast Media Mol Imaging* 2019; **2019**: 1545747 [PMID: 31354393 DOI: 10.1155/2019/1545747]

44 **Chilamkurthy S**, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, Mahajan V, Rao P, Warier P. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* 2018; **392**: 2388-2396 [PMID: 30318264 DOI: 10.1016/S0140-6736(18)31645-3]

45 **Liu Z**, Wang S, Dong D, Wei J, Fang C, Zhou X, Sun K, Li L, Li B, Wang M, Tian J. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* 2019; **9**: 1303-1322 [PMID: 30867832 DOI: 10.7150/thno.30309]

46 **Bi WL**, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH, Tamimi RM, Tempany CM, Swanton C, Hoffmann U, Schwartz LH, Gillies RJ, Huang RY, Aerts HJWL. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin* 2019; **69**: 127-157 [PMID: 30720861 DOI: 10.3322/caac.21552]

47 **Wynants L**, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, Bonten MMJ, Dahly DL, Damen JAA, Debray TPA, de Jong VMT, De Vos M, Dhiman P, Haller MC, Harhay MO, Henckaerts L, Heus P, Kammer M, Kreuzberger N, Lohmann A, Luijken K, Ma J, Martin GP, McLernon DJ, Andaur Navarro CL, Reitsma JB, Sergeant JC, Shi C, Skoetz N, Smits LJM, Snell KIE, Sperrin M, Spijker R, Steyerberg EW, Takada T, Tzoulaki I, van Kuijk SMJ, van Bussel B, van der Horst ICC, van Royen FS, Verbakel JY, Wallisch C, Wilkinson J, Wolff R, Hooft L, Moons KGM, van Smeden M. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020; **369**: m1328 [PMID: 32265220 DOI: 10.1136/bmj.m1328]

48 **Huang L**, Han R, Ai T, Yu P, Kang H, Tao Q, Xia L. Serial Quantitative Chest CT Assessment of COVID-19: A Deep Learning Approach. *Radiol Cardiothorac Imaging* 2020; **2**: e200075 [PMID: 33778562 DOI: 10.1148/ryct.2020200075]

49 **Li Z**, Zhong Z, Li Y, Zhang T, Gao L, Jin D, Sun Y, Ye X, Yu L, Hu Z, Xiao J, Huang L, Tang Y. From community-acquired pneumonia to COVID-19: a deep learning-based method for quantitative analysis of COVID-19 on thick-section CT scans. *Eur Radiol* 2020; **30**: 6828-6837 [PMID: 32683550 DOI: 10.1007/s00330-020-07042-x]

50 **Allen B**, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR Data Science Institute Artificial Intelligence Survey. *J Am Coll Radiol* 2021; **18**: 1153-1159 [PMID: 33891859 DOI: 10.1016/j.jacr.2021.04.002]

Artificial Intelligence in Medical Imaging

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

Artif Intell Med Imaging 2022 April 28; 3(2): 42-54

DOI: 10.35711/aimi.v3.i2.42

ISSN 2644-3260 (online)

SYSTEMATIC REVIEWS

Applications of artificial intelligence in lung ultrasound: Review of deep learning methods for COVID-19 fighting

Laura De Rosa, Serena L'Abbate, Claudia Kusmic, Francesco Faita

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

P-Reviewer: Haurylenka D, Belarus; Zhang W, China

Received: December 19, 2021
Peer-review started: December 19, 2021
First decision: February 10, 2022
Revised: February 22, 2022
Accepted: April 27, 2022
Article in press: April 27, 2022
Published online: April 28, 2022

Laura De Rosa, Serena L'Abbate, Claudia Kusmic, Francesco Faita, Institute of Clinical Physiology, Consiglio Nazionale delle Ricerche, Pisa 56124, Italy

Serena L'Abbate, Institute of Life Sciences, Scuola Superiore Sant'Anna, Pisa 56124, Italy

Corresponding author: Claudia Kusmic, MSc, PhD, Research Scientist, Institute of Clinical Physiology, Consiglio Nazionale delle Ricerche, Via Giuseppe Moruzzi 1, Pisa 56124, Italy. kusmic@ifc.cnr.it

Abstract

BACKGROUND

The pandemic outbreak of the novel coronavirus disease (COVID-19) has highlighted the need to combine rapid, non-invasive and widely accessible techniques with the least risk of patient's cross-infection to achieve a successful early detection and surveillance of the disease. In this regard, the lung ultrasound (LUS) technique has been proved invaluable in both the differential diagnosis and the follow-up of COVID-19 patients, and its potential may be destined to evolve. Recently, indeed, LUS has been empowered through the development of automated image processing techniques.

AIM

To provide a systematic review of the application of artificial intelligence (AI) technology in medical LUS analysis of COVID-19 patients using the preferred reporting items of systematic reviews and meta-analysis (PRISMA) guidelines.

METHODS

A literature search was performed for relevant studies published from March 2020 - outbreak of the pandemic - to 30 September 2021. Seventeen articles were included in the result synthesis of this paper.

RESULTS

As part of the review, we presented the main characteristics related to AI techniques, in particular deep learning (DL), adopted in the selected articles. A survey was carried out on the type of architectures used, availability of the source code, network weights and open access datasets, use of data augmentation, use of the transfer learning strategy, type of input data and training/test datasets, and explainability.

CONCLUSION

Finally, this review highlighted the existing challenges, including the lack of large datasets of reliable COVID-19-based LUS images to test the effectiveness of DL methods and the ethical/regulatory issues associated with the adoption of automated systems in real clinical scenarios.

Key Words: Lung ultrasound; Deep learning; Neural network; COVID-19 pneumonia; Medical imaging

Core Tip: Challenging coronavirus disease 2019 (COVID-19) pandemic through the identification of effective diagnostic and prognostic tools is of outstanding importance to tackle the healthcare system burdening and improve clinical outcomes. Application of deep learning (DL) in medical lung ultrasound may offer the advantage of combining non-invasiveness and wide accessibility of ultrasound imaging techniques with higher diagnostic performance and classification accuracy. This paper overviews the current applications of DL models in medical lung ultrasound imaging in COVID-19 patients, and highlight the existing challenges associated with the effective clinical application of automated systems in the medical imaging field.

Citation: De Rosa L, L'Abbate S, Kusmic C, Faita F. Applications of artificial intelligence in lung ultrasound: Review of deep learning methods for COVID-19 fighting. *Artif Intell Med Imaging* 2022; 3(2): 42-54
URL: https://www.wjgnet.com/2644-3260/full/v3/i2/42.htm
DOI: https://dx.doi.org/10.35711/aimi.v3.i2.42

INTRODUCTION

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a life-threatening infectious virus and its related disease (COVID-19) represents a still ongoing challenge for humans. At time of writing, over 497 million infections have been recorded worldwide including more than 6.1 million attributable deaths[1]. Despite the large number of vaccination programs introduced from the end of 2020 has represented an opportunity to minimise the risk of severe COVID-19 and death, the spread of new genetic viral variants with a higher probability of contagion has raised a renewed strong concern for either not vaccinated and vaccinated people. Thus, since the outbreak of the pandemic, research has continuously looked for a quick and reliable way to diagnose the disease, treat and monitor people affected by coronavirus.

To date, molecular test based on real time quantitative reverse transcription polymerase chain reaction (RT-qPCR) assay by nasopharyngeal swabs along with the serological antibody-detecting and antigen-detecting tests are the current accepted diagnostic tools for the conclusive diagnosis of COVID-19[2]. RT-qPCR may take up to 24 h to provide information and requires multiple tests for definitive results and, in addition, it is not relevant to assess the disease severity. Furthermore, the accuracy of molecular and serological tests remains highly dependent on timing of sample collection relative to infection, improper sampling of respiratory specimens, inadequate preservation of samples and technical errors, particularly contamination during RT-qPCR process and cross-reactivity in the immunoassay[3,4].

To complement conventional *in vitro* analytical techniques of COVID-19, biomedical imaging techniques have demonstrated great potential in clinical diagnostic evaluation by providing rapid patient assessment in the presence of high pre-test probability. Furthermore, imaging techniques are currently important in the follow-up of subjects with COVID-19[5,6]. Among the imaging techniques, chest computed tomography (CT) is considered the primary diagnostic modality and an important indicator for assessing severity and progression of COVID-19 pneumonia[7,8], although it has been reported to have limited specificity[9-11]. Indeed, the CT imaging features can overlap between COVID-19 and other viral pneumonia. Moreover, CT scanning is expensive, not easy to perform in the COVID-19 context, and multiple risks are associated with it, such as radiation exposure and cross-infection risk associated with repeated use of a CT suite[12], along with unavailability of CT in many parts of the world.

In the last few years, lung ultrasound (LUS) technique has become increasingly popular and a good option for real-time point-of-care testing, with several advantages making it a valuable tool in the fight against COVID-19[13], although it has specificity limits comparable to those of chest CT.

Ultrasound (US) is a low-cost, non-radioactive medical imaging method, particularly indicated for evaluation in pregnant women and children, which is portable to the bedside or patient's home and is easy to sterilise. Moreover, the risk of COVID-19 cross-infection can be limited by making use of

disposable ultrasound gel with a portable probe[14]. In addition, some studies indicate that LUS shows excellent performances in speed of execution and accuracy of diagnosis in case of respiratory failure [15]. Furthermore, compared with chest X-ray, LUS demonstrated higher sensitivity in detecting pneumonia[16] and similar specificity in the diagnosis of pneumothorax[15]. On the other hand, the distinctive LUS features (B-lines, consolidations, pleural thickening and rupture) observed in patients with varying severity of COVID pneumonia are similar to the features seen in patients with pneumonia of different aetiologies. Indeed, a recent review[17] on ultrasound findings of LUS in COVID-19 demonstrated that LUS has high sensitivity and reliability in ruling out lung involvement, but at the expense of low specificity. Therefore, especially in the case of low prevalence of the disease, at present LUS cannot be considered a valid gold standard in clinical practice.

Ultrasound image processing techniques have assumed great importance in recent years, with the growing experience that accurate image processing can significantly help in extracting quantitative characteristics to assess and classify the severity of diseases. Accordingly, sophisticated techniques of automated image processing, that include the use of artificial intelligence (AI) methods, have been developed and applied to assist LUS imaging in the detection of COVID-19 and make such assessment more objective and accurate. AI methods - from machine learning (ML) to deep learning (DL), indeed, aim to imitate cognitive functions and stand out in automatically recognizing complex patterns in imaging data, providing quantitative rather than qualitative assessments. The primary purpose of applying AI methods in medical imaging is to improve the visual recognition of certain features in images to produce lower-than-human error rates. Furthermore, an enhancement in LUS performance can reduce the use of more invasive and time-consuming techniques, facilitating both faster diagnosis and recognition of earlier stages of the disease[18]. To allow a quick development of highly performant AI models, a large amount of accessible and validated data to train and test AI models is a critical requirement that can be achieved, for instance, with the development of shared big data archives. Indeed, one of the most common problems associated with using limited training samples is the over-fitting of DL models. To address this issue, two main approaches can be selected: model optimization and transfer learning. These strategies significantly improve the performance of DL models. Likewise, data pre-processing and data augmentation/enhancement can be useful additional strategies[19,20].

The most common applications of DL methods in clinical imaging, and hence in medical ultrasound imaging as well, are object detection, object segmentation, and object classification[21]. The main architectures applied in current analysis are convolutional neural networks (CNNs) and recurrent neural networks (RNNs)[22]. CNNs are architectures able to work with 2D and 3D input images and RNNs recognize the image's sequential characteristics and use patterns to predict the next likely scenario[23].

Since the outbreak of the pandemic, many proposals have been made based on AI methods applied to LUS scans of COVID-19 patients. Here we propose a comprehensive systematic review of the literature on the use of AI technology, DL in particular, to aid in the fight against COVID-19.

MATERIALS AND METHODS

Study selection

A literature search to identify all relevant articles on the use of DL tools applied to LUS imaging in patients affected by COVID-19 virus was conducted.

This systematic review was carried out using the PubMed/Medline electronic database and according to the preferred reporting for systematic reviews and meta-analysis (PRISMA) guidelines[24, 25]. We performed a systematic search covering the period from March 2020 (from the outbreak of the pandemic) to 30 September 2021. The search strategy was restricted to English-language publications.

We performed an advanced research concatenating terms with Boolean operators. In particular, search words and key terms used in the search included ("lung ultrasound" OR "lus") AND ("COVID-19" OR "coronavirus" OR "SARS-CoV2") AND ("artificial intelligence" OR "deep learning" OR "neural networks" OR "CNN").

Eligibility criteria

The inclusion criteria were: Studies that include COVID-19 patients with LUS acquisitions and developed or tested DL-based algorithms on LUS images or on features extracted from the images; No restriction on the ground truth adopted to analyse the presence/absence of COVID-19 and/or the severity of lung disease (*e.g.*, PCR, visual evaluation of video/images and score assignment by expert clinicians); No restriction on the type of DL architecture used in the studies. Studies on paediatric population were excluded. Studies were restricted to peer reviewed articles and conference proceedings. However, the following publication types were excluded: reviews and conference abstracts.

Data extraction and analysis

Two investigators (DRL and FF) screened the articles independently. Disagreement between reviewers was resolved by consensus *via* discussion. The reasons for the exclusion of some trials are described in the Results section. Publications by the same research group or by different groups using the same dataset were included in the analysis. After the selection of the articles, we collected the following characteristics: First author's surname, date of publication, sample size, general characteristics of the study populations, AI techniques used, validation methods and main results obtained. The study selection process is presented in Figure 1.

RESULTS

Search results

Twenty-four articles resulted after querying the database and screened for eligibility (Figure 1). Of the 24 articles, we discarded four references as review papers. After examining the titles and abstracts, we excluded five articles: one manuscript did not include DL methods applied on US imaging, three papers were not based on AI and DL approaches, and one article was focused on the paediatric population. Moreover, two additional papers, retrieved from the checking of references of the eligible articles, were included. Finally, 17 articles[26-42] were selected for full-text screening and included in our analysis (Table 1 and 2). The following part of the section provides a concise overview of the studies' main features.

Dataset and source code availability

Authors of seven[27-30,33,39,40] of the seventeen selected articles (41.2%) extrapolated their datasets from the free access LUS database acquired by point-of-care ultrasound imaging and made available firstly by Born *et al*[30]. Instead, an Italian group firstly introduced the Italian COVID-19 Lung Ultrasound DataBase (ICLUS-DB)[38], which is accessible upon mandatory request to the authors, and that was used in two other studies[32,37]. Noteworthy, Roy *et al*[38] have created a platform through which physicians can access algorithms, upload their data and see the algorithm's evaluation of the data.

Besides dataset open access, access to the code for the neural network is also important to reproduce results and compare performances. Seven articles[26-30,32,38] (41.2%) made the source code implementing the proposed DL architecture available for download from the Git-hub repository.

Single-frame/multi-frames or video based architecture

In the majority of the selected papers, DL architectures work with single frame images as input and only three publications[29,34,41] (17.6%) report DL architectures based on image sequences (*i.e.*, video). However, six studies[28,30,32,37-39] (35.3%), despite adopting a DL architecture designed to perform single-frame classification, also propose additional methods to fulfil video-based classification. In particular, Roy *et al*[38] proposed an aggregation layer system of frame-level scores to produce predictions on LUS videos and Mento *et al*[37] proposed an alternative video-based classification using a threshold-based system on the frame-level scores obtained from DL architecture.

Other authors[32] adopted a Long Short-Term Memory (LSTM) system, which has been used to exploit temporal relationships between multiple frames by taking long time series as input, over performing their results obtained by CNN without LSTM.

Finally, Xue *et al*[42] applied AI models for patient-level assessment of severity using a final module across the entire architecture that works with ML rather than DL systems.

Test strategy of DL models

The proposed DL models have been tested on a database entirely independent from the training database in seven articles[26,35-39,42] (41.2%); five-fold and ten-fold cross-validation techniques were applied in nine[27-34,40] (52.9%) and one[41] (5.9%) studies, respectively. Among the papers that tested DL models on an independent database, the percentage of data used for the testing ranged from 33%[35] to 20%[38] and 10%[26,36] of the overall data. Born *et al*[29], alongside the five-fold cross-validation technique in the training/test phase of the DL model, also used an independent validation dataset made-up of 31 videos (28 convex and 3 linear probes) from six patients. Indeed, Roy *et al*[38], for instance, used 80 videos/10709 frames out of the total 277 videos/58924 frames to test their DL model.

In all studies, the splitting of data between training set and test set was performed either at the patient-level or at the video-level. Thus, all the frames of a single video clip belonged either to the training or to the test set.

Data augmentation

Twelve (70.6%) research groups extended their LUS database by augmentation. The main strategies for data augmentation applied to LUS images were: Horizontal/vertical flipping[26,27,29,30,32,33,36,38-40,

Table 1 General characteristics of the studies included in the analysis (part I)

Ref.	Publication date	Journal	Sample size[1], N° pts/videos/images	Subjects	Main results
Arntfield *et al* [26]	22/02/2021	BMJ Open	243/612/121k	COVID +, COVID -, HPE	Overall Acc = 0.978AUC = 1/0.934/1 for COVID +, COVID -, HPE
Awatshi *et al* [27]	23/03/2021	IEEE Trans Ultrason Ferroelectr Freq Control	-/64/1.1k	COVID +, Healthy, PN	5-fold validation: Acc = 0.829
Barros *et al*[28]	14/08/2021	Sensors	131/185/-	COVID +, PN bacterial, Healthy	Best model (Xception+LSTM): Acc = 0.93 – Se = 0.97
Born *et al*[29]	12/01/2021	Applied Sciences	216/202/3.2k	COVID +, Healthy, PN	External validation: Se = 0.806 – Sp = 0.962
Born *et al*[30]	24/01/2021	ISMB TransMed	-/64/1.1k	COVID +, Healthy, PN	Overall Acc = 0.89Binarization COVID y/n: Se = 0.96 – Sp = 0.79 – F1score = 0.92
Chen *et al*[31]	29/06/2021	IEEE Trans Ultrason Ferroelectr Freq Control	31/45/1.6k	COVID-19 PN	5-fold validation: Acc = 0.87
Dastider *et al* [32]	20/02/2021	Comput Biol Med	29/60/14.3k	COVID-19 PN	Independent data validation: Acc = 0.677 – Se = 0.677 – Sp = 0.768 – F1score = 0.666
Diaz Escobar *et al*[33]	13/08/2021	PLos One	216/185/3.3k	COVID +, PN bacterial, Healthy	Best model (InceptionV3): Acc = 0.891 – AUC = 0.971
Erfanian Ebadi *et al*[34]	04/08/2021	Inform Med Unlocked	300/1.5k/288k	COVID +, PN	5-fold validation: Acc = 0.90 – PP=0.95
Hu *et al*[35]	20/03/2021	BioMed Eng OnLine	108/-/5.7k	COVID +	COVID detection: Acc = 0.944 – PP = 0.823 – Se = 0.763 – Sp=0.964
La Salvia *et al* [36]	03/08/2021	Comput Biol Med	450/5.4k/> 60k	Hospitalised COVID-19	External validation (ResNet50): Acc = 0.979 – PP=0.978 – F1score = 0.977 – AUC = 0.998
Mento *et al*[37]	27/05/2021	J Acoust Soc Am	82/1.5k/315k	COVID-19 confirmed	% Agreement DL and LUS = 96%
Roy *et al*[38]	14/05/2020	IEEE Trans	35/277/58.9k	COVID-19 confirmed, COVID-19 suspected, Healthy	Segmentation: Acc = 0.96 – DICE = 0.75
Sadik *et al*[39]	09/07/2021	Health Inf Sci Syst	-/123/41.5k	COVID +, PN, Healthy	COVID y/n (VGG19+SpecMen): PP = 0.81 – F1score = 0.89
Muhammad *et al*[40]	25/02/2021	Information Fusion	121 videos + 40 frames	COVID +, PN bacterial, Healthy	Overall: Acc = 0.918 – PP = 0.925
Tsai *et al*[41]	08/03/2021	Phys Med	70/623/99.2k	Healthy, Pleural effusion pts	Pleural effusion detection:Acc = 0.924
Xue *et al*[42]	20/01/2021	Med Image Anal	313/-/6.9k	COVID-19 confirmed	4-level and binary disease severity:Acc = 0.75 and Acc = 0.85

[1]k: Indicates × 10^3.

pts: Patients; HPE: Hydrostatic pulmonary edema; PN: Pneumonia; Acc: Accuracy; Se: Sensitivity; Sp: Specificity; AUC: Area under the curve; PP: Precision; DL: Deep learning; LUS: Lung ultrasound.

42], bidirectional arbitrary rotation[26,27,29,30,32,33,35,38-40,42], horizontal and vertical shift[30,32,38,39,42]; filtering, colour transformation, adding salt and pepper noise, Gaussian noise[36,38,42], normalisation of grey levels' intensity[38]. Although proposed by all the authors, only seven papers[26,29,30,32,33,38,40] provided details on the amplitude of image rotation. In particular, Dastider *et al*[32] applied rotations in the range of 0 ± 360 degrees, while other authors have limited image rotations to 10 degrees [26,29,30,33], ± 15 degrees[38] and ± 20 degrees[40], respectively. The remaining five papers[28,31,34,37,41] (29.4%) did not perform data augmentation.

Explainability

Among the selected articles, tools for interpreting the network output were provided in twelve studies (70.6%), whereas in the remaining five (29.4%) the DL algorithms' outcomes were proposed as black box systems. The majority of papers[26-29,32,35,36,38,40] reported the Gradient-weighted Class Activation

Table 2 General characteristics of the studies included in the analysis (part II)

Ref.	DL architecture	Input of DL models	Available dataset	Available code	Pre-trained/TL	Test independent	Data Augmentation	Explainability
Arntfield *et al* [26]	CNN	SF	No	Yes (on github)	Yes	Yes	Yes	Yes
Awatshi *et al* [27]	CNN	SF	No	Yes (on github)	Yes	No (five-fold)	Yes	Yes
Barros *et al*[28]	CNN+LSTM	SF	Yes	Yes (on github)	Yes	No(five-fold)	No	Yes
Born *et al*[29]	3D CNN	MF	Yes	Yes (on github)	Yes	No(five-fold)	Yes	Yes
Born *et al*[30]	CNN	SF	Yes	Yes (on github)	Yes	No(five-fold)	Yes	No
Chen *et al*[31]	MLFCNN	SF	No	Yes (on github)	No	No(five-fold)	No	No
Dastider *et al* [32]	CNN+LSTM	SF	No	Yes (on github)	Yes	No(five-fold)	Yes	Yes
Diaz Escobar *et al*[33]	CNN	SF	No	No	Yes	No(five-fold)	Yes	No
Erfanian Ebadi *et al*[34]	3D CNN	MF	No	Yes (on github)	Yes	No(five-fold)	No	Yes
Hu *et al*[35]	CNN + MCRF	SF	No	No	Yes	Yes	Yes	Yes
La Salvia *et al* [36]	CNN	SF	No	No	Yes	Yes	Yes	Yes
Mento *et al*[37]	CNN+ STN	SF	No	No	No	-	No	No
Roy *et al*[38]	CNN+ STN	SF	Yes (on request)	Yes (on github)	No	Yes	Yes	Yes
Sadik *et al*[39]	CNN	SF	No	No	Yes	Yes	Yes	Yes
Muhammad *et al*[40]	CNN	SF	Yes	No	No	No(five-fold)	Yes	Yes
Tsai *et al*[41]	CNN+ STN	MF	No	No	Yes	No(ten-fold)	No	No
Xue *et al*[42]	CNN	SF	No	No	No	Yes	Yes	Yes

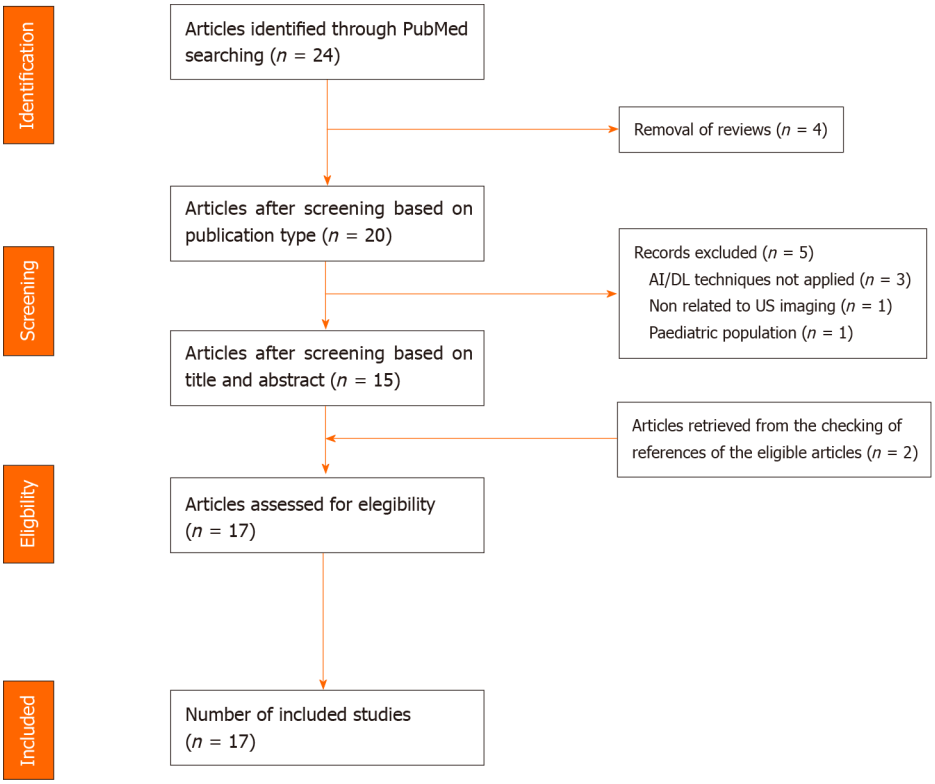
CNN: Convolutional neural network; LSTM: Long short-term memory; MCRF: Multimodal channel and receptive field; MLFCNN: Multi-layer fully connected neural network; STN: Spatial transformer network; SF: Single-frame; MF: Multi-frame; DL: Deep learning; TL: Transfer learning.

Mapping (Grad-CAM) as the preferred explainability tool. Grad-CAM uses gradients to create a location map to highlight the region of interest of the images[43]. Instead, Sadik *et al*[39] used a colormap jet to visualise a heat map overlay to US images; Erfanian Ebadi *et al*[34] adopted an activation map system to detect and segment features in LUS scans. Furthermore, one study[42] showed LUS images with overlaid colormaps to indicate the segmentation zone of ultrasound according to the different severity. Roy *et al*[38], differently, provided an ultrasound colormap overlay on the LUS frame/video and used four colours to distinguish the different classes of disease severity recognized by DL architecture.

Clinical use

Most of the selected papers applied the AI system to diagnose COVID-19 and/or discriminate between COVID-19 and other lung diseases (such as bacterial pneumonia)[26-30,33,34,39,40]. The first approach using DL architecture for automatic differential diagnosis of COVID-19 from LUS data was POCOVID-Net[30].

However, a fair number of studies have focused on assessing the severity of COVID-19[31,32,35-38,42]. In particular, a disease severity score is assigned to the single image according to some characteristics visible in the image pattern. Most of the articles used four severity classes by assigning a score to the single frame from 0 to 3[31,32,35-38], as defined by Soldati *et al*[44]. Xue *et al*[42] proposed a classification in five classes of pneumonia severity (score from 0 to 4) along with a binary severe/non-severe classification. Furthermore, these authors used the DL technology exclusively to implement a segmentation phase based on a VGG network, while the classification phase still employed a more traditional, features-based machine learning approach. Finally, La Salvia *et al*[36] proposed a classi-



DOI: 10.12998/wjcc.v3.i2.42

Figure 1 Flow diagram of systematic identification, screening, eligibility and inclusion of publications that applied deep learning methods to lung ultrasound imaging in coronavirus disease 2019 patients. AI: Artificial intelligence; DL: Deep learning; US: Ultrasound.

fication based on three severity classes and a modified version considering a seven-classes scenario.

Furthermore, Arntfield *et al*[26] showed that their network was able to recognize pathological pattern in LUS images with higher sensitivity than sonographers; whilst an InceptionV3 network proposed by Diaz-Escobar *et al*[33] was able to discriminate COVID-19 pneumonia from healthy lung and other bacterial pneumonia with an accuracy of 89.1% and an area under the ROC curve of 97.1%.

Curiously, one of the eligible papers[41] did not include confirmed cases of COVID-19 patients. The authors' aim was to design an algorithm capable of identifying the presence of pleural effusion. However, we have included this work in our systematic review, because small pleural effusions are rarely reported in COVID-19 patients. Therefore, the detection of pneumonia with pleural effusion can help rule out the hypothesis of COVID-19 disease.

Transfer learning and DL architecture

From our analysis, it emerged that most of the studies have proposed convolutional neural networks (CNNs) as DL models to generate screening systems for COVID-19. In particular, all publications with the exception of one[31] used the CNN network. Conversely, Chen *et al*[31] developed a multi-layer fully connected neural network for scoring LUS images in assessing the severity of COVID-19 pneumonia.

Among the DL systems included in this review, most of them were generated starting from DL architectures already proposed for other tasks[26-30,32-36,39,42], suitably modified and trained for new tasks. Furthermore, many works compared the results of their architectures with those obtained using existing and well-known architectures[27-30,32,33,35,38-40]. In particular, the following DL architectures were adapted to fulfil the requirements of LUS analysis to assist in COVID-19 detection and/or assessment of the severity of the lung disease, or just to compare their performances: VGG-19 [28,33,39] and VGG-50[28-30,33]; Xception[26,28,39]; ResNet 50[27,33,36,40]; NasNetMobile[27,29,39]; DenseNet[32,39].

More in detail, Awasthi *et al*[27] proposed Mini-COVIDNet, a modified MobileNet model belonging to the CNN's networks family and originally developed for detecting objects in mobile applications[45]. Barros *et al*[28], along with their proposed DL model, also investigated the impact of using different pre-trained CNN architectures in extracting spatial features that were successively classified by a LSTM model. Finally, Born *et al*[29] derived their DL video-based models from a model that was pre-trained on lung CT scans[46].

All aforementioned architectures are pre-trained on ImageNet[47].

Sample size

Partly due to the recent outbreak of the pandemic and to the difficulty of having standardised high quality archives of US images, only few of the selected studies relied on a large dataset in terms of enrolled patients. Six papers (35.3%) reported a sample size greater than 200 subjects (namely, 243, 216, 216, 300, 450 and 313 in references[26,29,33,34,36,42] respectively).

However, despite the relatively low number of subjects, the total number of LUS videos reaches up to 5400 in one study[36], with an average equal to 1589 videos[26,29,33,34,36]. Among the studies carried out on a low sample size, Dastider *et al*[32] included 29 patients and 60 videos, whilst 35 patients/45 videos and 35 patients/277 videos were analysed in references Chen *et al*[31] and Roy *et al*[38], respectively. However, it should be noted that Roy *et al*[38] published their work at the beginning of the COVID-19 pandemic, when the total number of COVID-19 patients was still relatively limited. In the paper by Xue *et al*[42], the number of frames/video was not reported.

DISCUSSION

The paper reviews the different DL techniques able to work with LUS images in assisting the diagnosis and/or prognosis of the COVID-19 disease published since the outbreak of the pandemic. In the selected documents, the use of DL systems aimed to achieve an accuracy comparable to or better than clinical standards to provide a faster diagnosis and/or follow-up in COVID-19 patients.

Most of the papers present pre-trained DL architectures[26-30,32-36,39,42] that were modified and adapted to new data. This approach is also known as transfer learning (TL) technique - *i.e.*, a training strategy for new DL models with reduced datasets. The network is pre-trained on a very large dataset, such as ImageNet, with millions of images intentionally created to facilitate the training of DL models, focusing on image classification and object location/detection tasks[48]. Indeed, deeper models are difficult to train and provide inconsistent performances when trained on a limited amount of data[49]. Therefore, most of the studies based on DL systems to classify COVID-19 images appropriately use the TL strategy as large datasets of US images from COVID-19 patients are not yet easily available, partly because the coronavirus disease is a relatively recent concern.

Furthermore, most of the proposed systems shared the same design, *i.e.*, CNN's architectures. CNNs have several applications in medical imaging – among others, image segmentation and object detection [50]. However, CNNs are particularly suited for image classification problems[51] and, consequently, represent an optimal solution for the classification of the disease severity from US images.

To date, one of the main challenges faced by DL architectures applied to LUS images of COVID-19 patients are the limited datasets in the available databases. This problem could benefit from creating open access databases that collect large amounts of data from multiple centres. In some of the selected studies, a first attempt to overcome this issue is evident, with particular emphasis on the work by Born *et al*[30], the authors who first collected a free access dataset of lung images from healthy controls and patients affected by COVID-19 or other pneumonia.

The development of public and multicentre platforms would guarantee the collection of a continuously growing amount of data, large and highly heterogeneous, suited for the training and testing of new DL applications in medical imaging, both in the COVID-19 and LUS field. Furthermore, this would allow an easier comparison of performances among DL models proposed in different studies. However, alternative approaches are often used in the testing phase that do not require the use of independent data sets to evaluate the performance of the model in the event of a limited number of images available. Among these, the k-fold cross-validation is a statistical method used to evaluate the ability of ML models to generalise to previously unseen data. Despite being widely used in ML models, the k-fold cross validation approach is less reliable than tests performed using an external dataset; the latter is always preferable to test model's ability to adapt properly to new, previously unseen data.

Data augmentation techniques are an alternative strategy to overcome the issue of the limited amounts of data, largely adopted in practice. These techniques generate different versions of a real dataset artificially to both increase its size and the power of model's generalisation. Despite the great advantage in increasing data to feed DL architectures, data augmentation techniques should be used with awareness, as some geometric transformations could be unrealistic when applied to LUS images (*e.g.*, angles of rotations greater than 30°). In the field of DL applied to medical imaging, the use of architectures designed to work with 3D images is another interesting challenge. Indeed, a DL system that operates with 3D data input usually requires a larger amount of data for training, as a 3D network contains a parameters' number that is orders of magnitude greater than a 2D network. This could significantly increase the risk of overfitting, especially in the case of limited dataset availability. In addition, the training on large amounts of data requires high computational costs associated with memory and performance requirements of the tools used. LUS images are usually recorded in the form of videoclips (2D + time) and can be assimilated to 3D data. Exploitation of dynamic information naturally embedded in image sequences has proven very important in the analysis of lung echoes. In

particular, changes induced by COVID-19 viral pneumonia are better detectable in LUS through the analysis of multi-frames acquisition due to its ability in capturing dynamic features, *e.g.*, pleural sliding movements and generation of B-line artefacts[44].

Regardless of the data format (*i.e.*, 3D, 2D or 2D+time images), the labelling of ground truth data is required in supervised DL applications and should be provided by skilled medical professionals. However, it is a time-consuming activity, in particular in the 2D approach that is characterised by a high number of samples.

Indeed, some authors demonstrated that the performance in pleural effusion classification on LUS images obtained with the video-based approach was comparable to that obtained with frame-based analysis, despite a significant reduction in labelling effort[41]. Furthermore, Kinetics-I3D network was able to classify LUS video sequences with great accuracy and efficiency[34]. On the other hand, the video-based approach has also revealed a reduced accuracy in patients classification with respect to the single frame analysis; however, this could be explained by the relatively reduced number of available LUS clips[29].

Extending the use of DL architectures beyond multi-frame analysis with respect to single 2D images is highly desirable. In particular, these methods could be effectively used to assign a patient-level disease severity score. In fact, this information plays a key role in the selection of treatment, monitoring of disease progression and management of medical resources (*e.g.*, mechanical ventilator needed).

Code availability is another very critical issue in applications of AI in medical imaging. Indeed, the lack of ability to reproduce the training of the proposed DL models or to test these models on new US images is a rather widespread problem. Often, authors do not provide access to either the source code used to train NNs or the final weight of the trained network. On the other hand, the availability of this information would greatly facilitate the diffusion of new AI systems in the clinical setting.

DL systems are often presented as black boxes - *i.e.*, they produce a result without providing a clear understanding in "human terms" of how it was obtained. The black-box nature of the algorithms has restricted their clinical use until now. Consistently, the explainability - *i.e.*, making clear and understandable the features that influence the decisions of a DL model - is a critical point to guarantee a safe, ethical, and reliable use of AI. Especially in medical imaging applications, explainability is very important as it gives the opportunity to highlight regions of the image containing the visual features that are critical for the diagnosis. Gradient-weighted Class Activation Mapping (Grad-CAM) is a promising technique for producing "visual explanations" of decisions taken from a large class of CNN-based models, making their internal behaviour more understandable, thus partially overcoming the black-box problem. The basic idea is to produce a rough localization map that highlights the key regions in the image that have a major effect on customization of network parameters, thus maximally contributing to the prediction of outcomes[43].

These maps visualised areas using a blue-to-red scale, with the highest/lowest contribution to the class prediction operated by the model. The clinical use of DL systems is a crucial issue. One of the major current limitations of LUS imaging in COVID patients is the specificity. Focusing the design of DL systems to overcome this limit could really represent a benefit in the clinical setting.

Along this line, some of the included studies tested the agreement between physicians' ability to classify COVID-19 patients and that proposed by neural networks. Furthermore, this finding suggests that the automated system can capture some features (biomarkers) in US images that are not clearly visible to the human eye.

Finally, another important issue to mention is the use of the quantitative evaluation indicators and the analysis of the benchmarking techniques adopted to evaluate the effectiveness of the proposed methods. Unfortunately, the tools examined in the selected manuscripts had very heterogeneous targets (Table 1, Main results column), ranging from diagnostic to prognostic purposes or assessment of disease severity. This dispersion of intent and the few articles published in the literature at present make any comparison or analysis very difficult.

CONCLUSION

The studies analysed in this article have shown that DL systems applied to LUS images for the diagnosis/prognosis of COVID-19 disease have the potential to provide significant support to the medical community. However, there are a number of challenges to overcome before AI systems can be regularly employed in the clinical setting. On the one hand, the critical issues related to the availability of high-quality databases with large sample size of lung images/videos of COVID-19 patients and free access to datasets must be addressed. On the other hand, existing concerns about the methodological transparency (*e.g.*, explainability and reproducibility) of DL systems and the regulatory/ethical and cultural issues that the clinical use of AI methods raise must be resolved. Finally, a closer collaboration between the communities of informatics/engineers and medical professionals is desirable to facilitate the outcome of adequate guidelines for the use of DL in US pulmonary imaging and, more generally, in medical imaging.

ARTICLE HIGHLIGHTS

Research background

The current coronavirus disease 2019 (COVID-19) pandemic crisis has highlighted the need for biomedical imaging techniques in rapid clinical diagnostic evaluation of patients. Furthermore, imaging techniques are currently important in the follow-up of subjects with COVID-19. The lung ultrasound technique has become increasingly popular and is considered a good option for real-time point-of-care testing, although it has specificity limits comparable to those of chest computed tomography.

Research motivation

The application of artificial intelligence, and of deep learning in particular, in medical pulmonary ultrasound can offer an improvement in diagnostic performance and classification accuracy to a non-invasive and low-cost technique, thus implementing its diagnostic and prognostic importance to COVID-10 pandemic.

Research objectives

This review presents the state of the art of the use of artificial intelligence and deep learning techniques applied to lung ultrasound in COVID-19 patients.

Research methods

We performed a literature search, according to preferred reporting items of systematic reviews and meta-analysis guidelines, for relevant studies published from March 2020 - to 30 September 2021 on the use of deep learning tools applied to lung ultrasound imaging in COVID-19 patients. Only English-language publications were selected.

Research results

We surveyed the type of architectures used, availability of the source code, network weights and open access datasets, use of data augmentation, use of the transfer learning strategy, type of input data and training/test datasets, and explainability.

Research conclusions

Application of deep learning systems to lung ultrasound images for the diagnosis/prognosis of COVID-19 disease has the potential to provide significant support to the medical community. However, there are critical issues related to the availability of high-quality databases with large sample size and free access to datasets.

Research perspectives

Close collaboration between the communities of computer scientists/engineers and medical professionals could facilitate the outcome of adequate guidelines for the use of deep learning in ultrasound lung imaging.

FOOTNOTES

Author contributions: Kusmic C and Faita F designed the research study; Faita F and De Rosa L collected and analysed the references mentioned in the review; De Rosa L wrote the initial draft; Kusmic C, Faita F and L'Abbate S revised and edited the manuscript; all authors have read and approve the final manuscript.

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

PRISMA 2009 Checklist statement: The authors have read the PRISMA 2009 Checklist, and the manuscript was prepared and revised according to the PRISMA 2009 Checklist.

Country/Territory of origin: Italy

ORCID number: Laura De Rosa 0000-0002-6030-8483; Serena L'Abbate 0000-0002-4462-0645; Claudia Kusmic 0000-0002-8064-8311; Francesco Faita 0000-0002-6201-1843.

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

REFERENCES

1 **World Health Organization**. (2022, April 13). Coronavirus (COVID-19) Dashboard. Available from: https://covid19.who.int
2 **Li Z**, Yi Y, Luo X, Xiong N, Liu Y, Li S, Sun R, Wang Y, Hu B, Chen W, Zhang Y, Wang J, Huang B, Lin Y, Yang J, Cai W, Wang X, Cheng J, Chen Z, Sun K, Pan W, Zhan Z, Chen L, Ye F. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J Med Virol* 2020; **92**: 1518-1524 [PMID: 32104917 DOI: 10.1002/jmv.25727]
3 **Yang Y**, Yang M, Yuan J, Wang F, Wang Z, Li J, Zhang M, Xing L, Wei J, Peng L, Wong G, Zheng H, Wu W, Shen C, Liao M, Feng K, Yang Q, Zhao J, Liu L, Liu Y. Laboratory Diagnosis and Monitoring the Viral Shedding of SARS-CoV-2 Infection. *Innovation (N Y)* 2020; **1**: 100061 [PMID: 33169119 DOI: 10.1016/j.xinn.2020.100061]
4 **Mardian Y**, Kosasih H, Karyana M, Neal A, Lau CY. Review of Current COVID-19 Diagnostics and Opportunities for Further Development. *Front Med (Lausanne)* 2021; **8**: 615099 [PMID: 34026773 DOI: 10.3389/fmed.2021.615099]
5 **Hoffmann T**, Bulla P, Jödicke L, Klein C, Bott SM, Keller R, Malek N, Fröhlich E, Göpel S, Blumenstock G, Fusco S. Can follow up lung ultrasound in Coronavirus Disease-19 patients indicate clinical outcome? *PLoS One* 2021; **16**: e0256359 [PMID: 34432835 DOI: 10.1371/journal.pone.0256359]
6 **Martini K**, Larici AR, Revel MP, Ghaye B, Sverzellati N, Parkar AP, Snoeckx A, Screaton N, Biederer J, Prosch H, Silva M, Brady A, Gleeson F, Frauenfelder T; European Society of Thoracic Imaging (ESTI), the European Society of Radiology (ESR). COVID-19 pneumonia imaging follow-up: when and how? *Eur Radiol* 2021 [PMID: 34713328 DOI: 10.1007/s00330-021-08317-7]
7 **Bernheim A**, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, Diao K, Lin B, Zhu X, Li K, Li S, Shan H, Jacobi A, Chung M. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. *Radiology* 2020; **295**: 200463 [PMID: 32077789 DOI: 10.1148/radiol.2020200463]
8 **Lyu P**, Liu X, Zhang R, Shi L, Gao J. The Performance of Chest CT in Evaluating the Clinical Severity of COVID-19 Pneumonia: Identifying Critical Cases Based on CT Characteristics. *Invest Radiol* 2020; **55**: 412-421 [PMID: 32304402 DOI: 10.1097/RLI.0000000000000689]
9 **Raptis CA**, Hammer MM, Short RG, Shah A, Bhalla S, Bierhals AJ, Filev PD, Hope MD, Jeudy J, Kligerman SJ, Henry TS. Chest CT and Coronavirus Disease (COVID-19): A Critical Review of the Literature to Date. *AJR Am J Roentgenol* 2020; **215**: 839-842 [PMID: 32298149 DOI: 10.2214/AJR.20.23202]
10 **Inui S**, Gonoi W, Kurokawa R, Nakai Y, Watanabe Y, Sakurai K, Ishida M, Fujikawa A, Abe O. The role of chest imaging in the diagnosis, management, and monitoring of coronavirus disease 2019 (COVID-19). *Insights Imaging* 2021; **12**: 155 [PMID: 34727257 DOI: 10.1186/s13244-021-01096-1]
11 **Xu B**, Xing Y, Peng J, Zheng Z, Tang W, Sun Y, Xu C, Peng F. Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy. *Eur Radiol* 2020; **30**: 5720-5727 [PMID: 32415585 DOI: 10.1007/s00330-020-06934-2]
12 **Self WH**, Courtney DM, McNaughton CD, Wunderink RG, Kline JA. High discordance of chest x-ray and computed tomography for detection of pulmonary opacities in ED patients: implications for diagnosing pneumonia. *Am J Emerg Med* 2013; **31**: 401-405 [PMID: 23083885 DOI: 10.1016/j.ajem.2012.08.041]
13 **Smith MJ**, Hayward SA, Innes SM, Miller ASC. Point-of-care lung ultrasound in patients with COVID-19 - a narrative review. *Anaesthesia* 2020; **75**: 1096-1104 [PMID: 32275766 DOI: 10.1111/anae.15082]
14 **Akl EA**, Blažić I, Yaacoub S, Frija G, Chou R, Appiah JA, Fatehi M, Flor N, Hitti E, Jafri H, Jin ZY, Kauczor HU, Kawooya M, Kazerooni EA, Ko JP, Mahfouz R, Muglia V, Nyabanda R, Sanchez M, Shete PB, Ulla M, Zheng C, van Deventer E, Perez MDR. Use of Chest Imaging in the Diagnosis and Management of COVID-19: A WHO Rapid Advice Guide. *Radiology* 2021; **298**: E63-E69 [PMID: 32729811 DOI: 10.1148/radiol.2020203173]
15 **Walden A**, Smallwood N, Dachsel M, Miller A, Stephens J, Griksaitis M. Thoracic ultrasound: it's not all about the pleura. *BMJ Open Respir Res* 2018; **5**: e000354 [PMID: 30305907 DOI: 10.1136/bmjresp-2018-000354]
16 **Amatya Y**, Rupp J, Russell FM, Saunders J, Bales B, House DR. Diagnostic use of lung ultrasound compared to chest radiograph for suspected pneumonia in a resource-limited setting. *Int J Emerg Med* 2018; **11**: 8 [PMID: 29527652 DOI: 10.1186/s12245-018-0170-2]
17 **Gil-Rodríguez J**, Pérez de Rojas J, Aranda-Laserna P, Benavente-Fernández A, Martos-Ruiz M, Peregrina-Rivas JA, Guirao-Arrabal E. Ultrasound findings of lung ultrasonography in COVID-19: A systematic review. *Eur J Radiol* 2022; **148**: 110156 [PMID: 35078136 DOI: 10.1016/j.ejrad.2022.110156]
18 **Langlotz CP**, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, Flanders AE, Lungren MP, Mendelson DS, Rudie JD, Wang G, Kandarpa K. A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology* 2019; **291**: 781-791 [PMID: 30990384 DOI: 10.1148/radiol.2019190613]
19 **Abu Anas EM**, Seitel A, Rasoulian A, St John P, Pichora D, Darras K, Wilson D, Lessoway VA, Hacihaliloglu I, Mousavi P, Rohling R, Abolmaesumi P. Bone enhancement in ultrasound using local spectrum variations for guiding percutaneous scaphoid fracture fixation procedures. *Int J Comput Assist Radiol Surg* 2015; **10**: 959-969 [PMID: 25847667 DOI: 10.1007/s11548-015-1181-6]
20 **Krizhevsky A**, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017; **60**: 84-90 [DOI: 10.1145/3065386]

21 **Currie GM**. Intelligent Imaging: Artificial Intelligence Augmented Nuclear Medicine. *J Nucl Med Technol* 2019; **47**: 217-222 [PMID: 31401616 DOI: 10.2967/jnmt.119.232462]

22 **Liu S**, Wang Y, Yang X, Lei B, Liu L, Li SX, Ni D, Wang T. Deep learning in medical ultrasound analysis: A review. Engineering 2019; 5: 261-275 [DOI: 10.1016/j.eng.2018.11.020]

23 **LeCun Y**, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436-444 [PMID: 26017442 DOI: 10.1038/nature14539]

24 **Moher D**, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339**: b2535 [PMID: 19622551 DOI: 10.1136/bmj.b2535]

25 **Liberati A**, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700 [PMID: 19622552 DOI: 10.1136/bmj.b2700]

26 **Arntfield R**, VanBerlo B, Alaifan T, Phelps N, White M, Chaudhary R, Ho J, Wu D. Development of a convolutional neural network to differentiate among the etiology of similar appearing pathological B lines on lung ultrasound: a deep learning study. *BMJ Open* 2021; **11**: e045120 [PMID: 33674378 DOI: 10.1136/bmjopen-2020-045120]

27 **Awasthi N**, Dayal A, Cenkeramaddi LR, Yalavarthy PK. Mini-COVIDNet: Efficient Lightweight Deep Neural Network for Ultrasound Based Point-of-Care Detection of COVID-19. *IEEE Trans Ultrason Ferroelectr Freq Control* 2021; **68**: 2023-2037 [PMID: 33755565 DOI: 10.1109/TUFFC.2021.3068190]

28 **Barros B**, Lacerda P, Albuquerque C, Conci A. Pulmonary COVID-19: Learning Spatiotemporal Features Combining CNN and LSTM Networks for Lung Ultrasound Video Classification. *Sensors (Basel)* 2021; **21** [PMID: 34450928 DOI: 10.3390/s21165486]

29 **Born J**, Wiedemann N, Cossio M, Buhre C, Brändle G, Leidermann K, Aujayeb A, Moor M, Rieck B, et al Accelerating detection of lung pathologies with explainable ultrasound image analysis. Appl Sci 2021; 11: 672 [DOI: 10.3390/app11020672]

30 **Born J**, Brändle G, Cossio M, Disdier M, Goulet J, Roulin J, Wiedemann N. POCOVID-Net: Automatic detection of COVID-19 from a new lung ultrasound dataset (POCUS). ISMB TransMed COSI 2020 2021. Preprint

31 **Chen J**, He C, Yin J, Li J, Duan X, Cao Y, Sun L, Hu M, Li W, Li Q. Quantitative Analysis and Automated Lung Ultrasound Scoring for Evaluating COVID-19 Pneumonia With Neural Networks. *IEEE Trans Ultrason Ferroelectr Freq Control* 2021; **68**: 2507-2515 [PMID: 33798078 DOI: 10.1109/TUFFC.2021.3070696]

32 **Dastider AG**, Sadik F, Fattah SA. An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound. *Comput Biol Med* 2021; **132**: 104296 [PMID: 33684688 DOI: 10.1016/j.compbiomed.2021.104296]

33 **Diaz-Escobar J**, Ordóñez-Guillén NE, Villarreal-Reyes S, Galaviz-Mosqueda A, Kober V, Rivera-Rodriguez R, Lozano Rizk JE. Deep-learning based detection of COVID-19 using lung ultrasound imagery. *PLoS One* 2021; **16**: e0255886 [PMID: 34388187 DOI: 10.1371/journal.pone.0255886]

34 **Erfanian Ebadi S**, Krishnaswamy D, Bolouri SES, Zonoobi D, Greiner R, Meuser-Herr N, Jaremko JL, Kapur J, Noga M, Punithakumar K. Automated detection of pneumonia in lung ultrasound using deep video classification for COVID-19. *Inform Med Unlocked* 2021; **25**: 100687 [PMID: 34368420 DOI: 10.1016/j.imu.2021.100687]

35 **Hu Z**, Liu Z, Dong Y, Liu J, Huang B, Liu A, Huang J, Pu X, Shi X, Yu J, Xiao Y, Zhang H, Zhou J. Evaluation of lung involvement in COVID-19 pneumonia based on ultrasound images. *Biomed Eng Online* 2021; **20**: 27 [PMID: 33743707 DOI: 10.1186/s12938-021-00863-x]

36 **La Salvia M**, Secco G, Torti E, Florimbi G, Guido L, Lago P, Salinaro F, Perlini S, Leporati F. Deep learning and lung ultrasound for Covid-19 pneumonia detection and severity classification. *Comput Biol Med* 2021; **136**: 104742 [PMID: 34388462 DOI: 10.1016/j.compbiomed.2021.104742]

37 **Mento F**, Perrone T, Fiengo A, Smargiassi A, Inchingolo R, Soldati G, Demi L. Deep learning applied to lung ultrasound videos for scoring COVID-19 patients: A multicenter study. *J Acoust Soc Am* 2021; **149**: 3626 [PMID: 34241100 DOI: 10.1121/10.0004855]

38 **Roy S**, Menapace W, Oei S, Luijten B, Fini E, Saltori C, Huijben I, Chennakeshava N, Mento F, Sentelli A, Peschiera E, Trevisan R, Maschietto G, Torri E, Inchingolo R, Smargiassi A, Soldati G, Rota P, Passerini A, van Sloun RJG, Ricci E, Demi L. Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound. *IEEE Trans Med Imaging* 2020; **39**: 2676-2687 [PMID: 32406829 DOI: 10.1109/TMI.2020.2994459]

39 **Sadik F**, Dastider AG, Fattah SA. SpecMEn-DL: spectral mask enhancement with deep learning models to predict COVID-19 from lung ultrasound videos. *Health Inf Sci Syst* 2021; **9**: 28 [PMID: 34257953 DOI: 10.1007/s13755-021-00154-8]

40 **Muhammad G**, Shamim Hossain M. COVID-19 and Non-COVID-19 Classification using Multi-layers Fusion From Lung Ultrasound Images. *Inf Fusion* 2021; **72**: 80-88 [PMID: 33649704 DOI: 10.1016/j.inffus.2021.02.013]

41 **Tsai CH**, van der Burgt J, Vukovic D, Kaur N, Demi L, Canty D, Wang A, Royse A, Royse C, Haji K, Dowling J, Chetty G, Fontanarosa D. Automatic deep learning-based pleural effusion classification in lung ultrasound images for respiratory pathology diagnosis. *Phys Med* 2021; **83**: 38-45 [PMID: 33706149 DOI: 10.1016/j.ejmp.2021.02.023]

42 **Xue W**, Cao C, Liu J, Duan Y, Cao H, Wang J, Tao X, Chen Z, Wu M, Zhang J, Sun H, Jin Y, Yang X, Huang R, Xiang F, Song Y, You M, Zhang W, Jiang L, Zhang Z, Kong S, Tian Y, Zhang L, Ni D, Xie M. Modality alignment contrastive learning for severity assessment of COVID-19 from lung ultrasound and clinical information. *Med Image Anal* 2021; **69**: 101975 [PMID: 33550007 DOI: 10.1016/j.media.2021.101975]

43 **Selvaraju RR**, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22-29; Venice, Italy. IEEE, 2017: 618–626 [DOI: 10.1109/ICCV.2017.74]

44 **Soldati G**, Smargiassi A, Inchingolo R, Buonsenso D, Perrone T, Briganti DF, Perlini S, Torri E, Mariani A, Mossolani EE, Tursi F, Mento F, Demi L. Proposal for International Standardization of the Use of Lung Ultrasound for Patients With COVID-19: A Simple, Quantitative, Reproducible Method. *J Ultrasound Med* 2020; **39**: 1413-1419 [PMID: 32227492 DOI: 10.1002/jum.15285]

45 **Howard AG**, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam M. Mobilenets: Efficient convolutional neural networks for mobile vision applications; 2017. Preprint. Cited 17 April 2017

46 **Zhou Z**, Sodha V, Rahman Siddiquee MM, Feng R, Tajbakhsh N, Gotway MB, Liang J. Models genesis: Generic autodidactic models for 3D medical image analysis. In Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap PT, Khan A editors. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019. Proceedings of 22nd International Conference MICCAI; 2019 Oct 13-17; Shenzhen, China. Lecture Notes in Computer Science. Springer, Cham, Switzerland, 2019; 11767: 384-393 [DOI: 10.1007/978-3-030-32251-9_42]

47 **Krizhevsky A**, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ. Advances in neural information processing systems. New York: Curran Associates Inc, 2012; 25: 1097–1105

48 **Russakovsky O**, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, et al ImageNet large scale visual recognition challenge. Int J Comput Vis 2015; 115: 211–252 [DOI: 10.1007/s11263-015-0816-y]

49 **Horry MJ**, Chakraborty S, Paul M, Ulhaq A, Pradhan B, Saha M, Shukla N. Covid-19 detection through transfer learning using multimodal imaging data. IEEE Access 2020; 8: 149808–149824 [DOI: 10.1109/ACCESS.2020.3016780]

50 **Moran M**, Faria M, Giraldi G, Bastos L, Oliveira L, Conci A. Classification of Approximal Caries in Bitewing Radiographs Using Convolutional Neural Networks. *Sensors (Basel)* 2021; **21** [PMID: 34372429 DOI: 10.3390/s21155192]

51 **LeCun Y**, Haffner P, Bottou L, Bengio Y. Object recognition with gradient-based learning. In: Shape, Contour and Grouping in Computer Vision. Lecture Notes in Computer Science, vol 1681. Springer: Berlin/Heidelberg, Germany, 1999 [DOI: 10.1007/3-540-46805-6_19]

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

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

MINIREVIEWS

Artificial Intelligence in Medical Imaging

Bimonthly Volume 3 Number 3 June 28, 2022

Contents

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

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 19th 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 standardization 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 investig-

ations, 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].

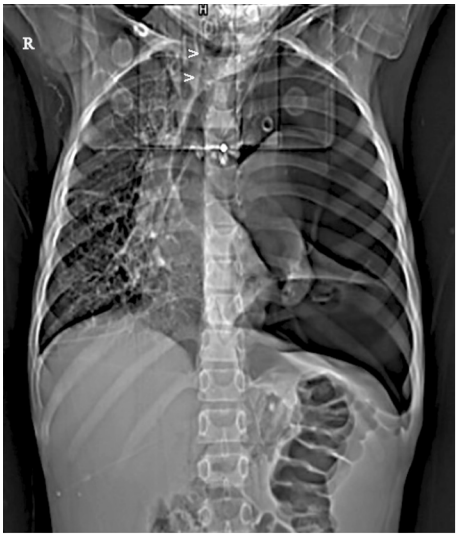
BIASES IN RADIOLOGY

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

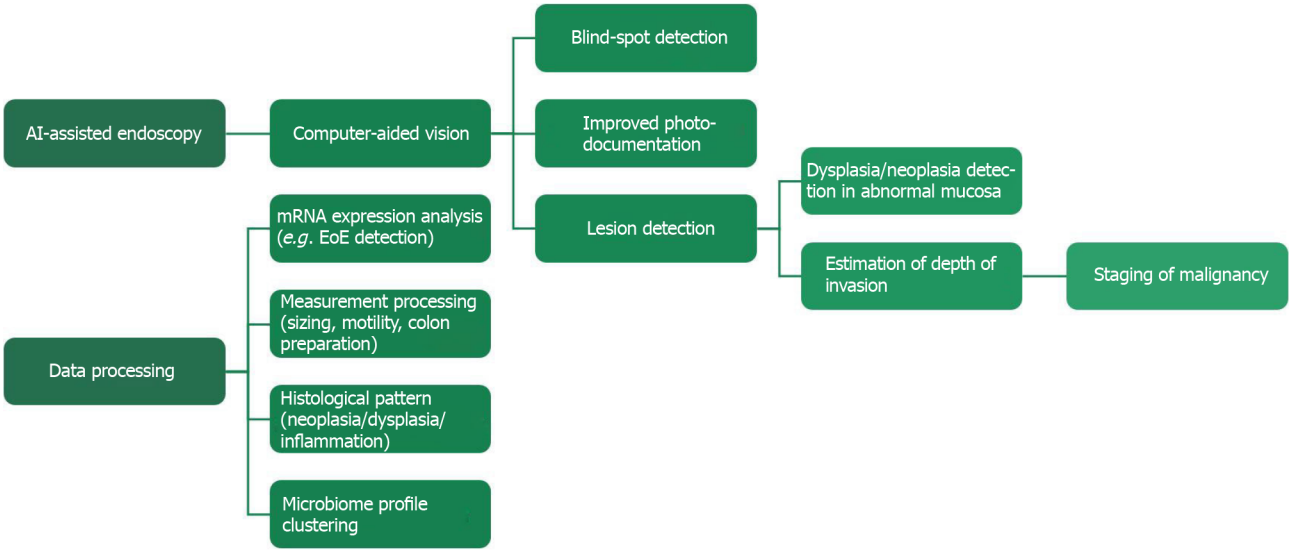
A literature search to identify all relevant articles on the use of AI in endoscopic screening and surveillance of gastric and esophageal diseases was conducted. The search was conducted utilizing PubMed, Medline, and Reference Citation Analysis (RCA) electronic database. We performed a systematic search 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

ISSN 2644-3260 (online)

Artificial Intelligence in *Medical Imaging*

Artif Intell Med Imaging 2022 December 28; 3(4): 87-96

Published by Baishideng Publishing Group Inc

AIMI

Artificial Intelligence in Medical Imaging

Contents

Bimonthly Volume 3 Number 4 December 28, 2022

MINIREVIEWS

Artificial Intelligence in Medical Imaging

Bimonthly Volume 3 Number 4 December 28, 2022

Contents

ABOUT COVER

Editorial Board Member of *Artificial Intelligence in Medical Imaging*, Suchart Kothan, PhD, Associate Professor, Center of Radiation Research and Medical Imaging, Department of Radiologic Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand. suchart.kothan@cmu.ac.th

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

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
December 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

Artificial Intelligence in Medical Imaging

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

Artif Intell Med Imaging 2022 December 28; 3(4): 87-96

DOI: 10.35711/aimi.v3.i4.87

ISSN 2644-3260 (online)

MINIREVIEWS

Radiomics: Status quo and future challenges

Zhi-Yun Jiang, Li-Shuang Qi, Jia-Tong Li, Nan Cui, Wei Li, Wei Liu, Ke-Zheng Wang

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): 0
Grade C (Good): C, C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Anzola F LK, Colombia; Litvin A, Russia

Received: September 20, 2022
Peer-review started: September 21, 2022
First decision: November 30, 2022
Revised: December 8, 2022
Accepted: December 21, 2022
Article in press: December 21, 2022
Published online: December 28, 2022

Zhi-Yun Jiang, Jia-Tong Li, Nan Cui, Wei Li, Wei Liu, Ke-Zheng Wang, Department of Positron Emission Tomography-Computed Tomography/Magnetic Resonance Imaging, Harbin Medical University Cancer Hospital, Harbin 150081, Heilongjiang Province, China

Li-Shuang Qi, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, Heilongjiang Province, China

Wei Li, Department of Interventional Vascular Surgery, The 4th Affiliated Hospital of Harbin Medical University, Harbin 150001, Heilongjiang Province, China

Corresponding author: Ke-Zheng Wang, MD, PhD, Chief Doctor, Professor, Department of Positron Emission Tomography-Computed Tomography/Magnetic Resonance Imaging, Harbin Medical University Cancer Hospital, No. 150 Haping Road, Harbin 150081, Heilongjiang Province, China. wangkezheng9954001@163.com

Abstract

Noninvasive imaging (computed tomography, magnetic resonance imaging, endoscopic ultrasonography, and positron emission tomography) as an important part of the clinical workflow in the clinic, but it still provides limited information for diagnosis, treatment effect evaluation and prognosis prediction. In addition, judgment and diagnoses made by experts are usually based on multiple years of experience and subjective impression which lead to variable results in the same case. With accumulation of medical imaging data, radiomics emerges as a relatively new approach for analysis. Via artificial intelligence techniques, high-throughput quantitative data which is invisible to the naked eyes extracted from original images can be used in the process of patients' management. Several studies have evaluated radiomics combined with clinical factors, pathological, or genetic information would assist in the diagnosis, particularly in the prediction of biological characteristics, risk of recurrence, and survival with encouraging results. In various clinical settings, there are limitations and challenges needing to be overcome before transformation. Therefore, we summarize the concepts and method of radiomics including image acquisition, region of interest segmentation, feature extraction and model development. We also set forth the current applications of radiomics in clinical routine. At last, the limitations and related deficiencies of radiomics are pointed out to direct the future opportunities and development.

Key Words: Radiomics; Methodologies; Quantification; Clinical applications; Limitations

Core Tip: Radiomics is widespread applied in clinical researches through extracting high-dimensional quantitative imaging features as a relatively emerging and mature technique based on medical imaging. The basic principles and methodologies of radiomics were reviewed to make it easy to understand from the relatively fixed processes. The representative clinical utilizations were declared to show the benefits of radiomics in diagnosis, tumor biological features and prognosis. Radiomics has revealed potential of clinical applications, while there are still many limitations to resolve in the further researches.

Citation: Jiang ZY, Qi LS, Li JT, Cui N, Li W, Liu W, Wang KZ. Radiomics: Status quo and future challenges. *Artif Intell Med Imaging* 2022; 3(4): 87-96
URL: https://www.wjgnet.com/2644-3260/full/v3/i4/87.htm
DOI: https://dx.doi.org/10.35711/aimi.v3.i4.87

INTRODUCTION

Radiomics was first proposed by Lambin *et al*[1] in 2012, which converts medical images into high-throughput quantitative features. Radiomic features can capture tissue and lesion properties noninvasively, such as shape and heterogeneity, and radiomics acts as a new approach to extract the information underlying the medical images that fail to be appreciated by naked eyes[2]. In the meantime, radiomics also possesses several advantages over molecular assays, such as being non-tissue-destructive, rapid analysis, easily serialized, fairly inexpensive, and being fully compatible with the existing clinical workflows[3]. In 2014, Aerts *et al*[4] demonstrated the role of radiomics in disease prognostication, promoting the development of radiomic-based signatures. Subsequently, the Pyradiomics framework based on the image biomarker standardization initiative (IBSI) criteria published in 2017 strongly supported the standardized application of radiomics[5].

Radiomics has evolved tremendously in the last decade, with the objective of precision medicine. However, the interpretability of radiomic-based signatures and the correlation with biology and pathology need to be further discussed. Additional multi-center data and prospective validation are also required for verification, in order to improve the confidence of applications[6]. There are still several substantial barriers to realize the objective of transforming artificial intelligence (AI) into the real clinical practice.

In the present study, the basic principles and methodologies of radiomics were reviewed and an outline of the representative clinical utilization was provided to highlight the benefits of radiomics in diagnosis, staging, tumor biological features, and prognosis. Additionally, it is essential to explore the deficiencies of radiomics to achieve a balanced interpretation between AI and clinical practice.

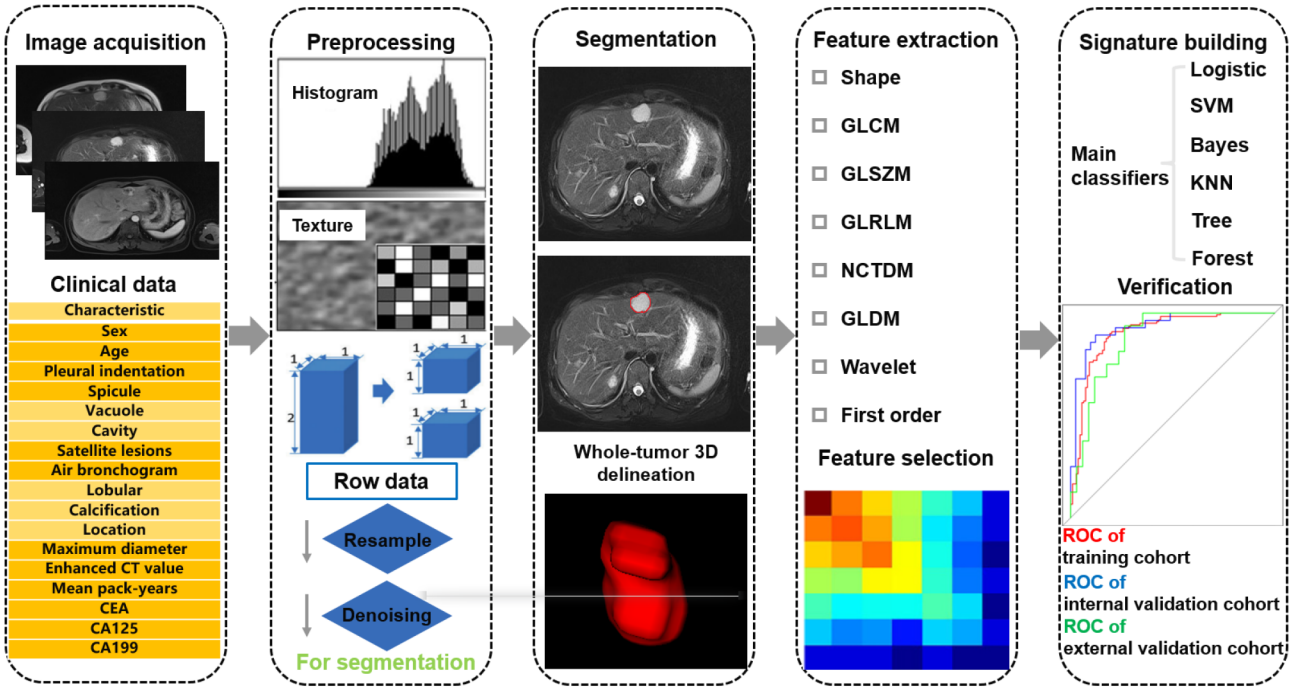
CONCEPT AND METHODOLOGIES

"Radiomics," a term that describes the "omics" approach for the analysis of imaging data, has emerged as a novel tool for diagnosis and prognosis[2]. Using advanced computational tools, high-throughput quantitative imaging features beyond inspections of naked human eyes are extracted and the desensitized medical images are transformed into multiple textural features for quantitative assessment[7-9]. With semantic features, radiomics enables clinicians to make more objective and accurate clinical decisions in diagnosis and prognosis[10,11]. The workflow of radiomics analysis, consisting of several steps, is illustrated in Figure 1.

Image acquisition

Image acquisition is approved by the ethics committee and informed consent form is signed by participants or their close relatives. The right to know patients is protected by relevant regulations. As the research of radiomics concentrated on human participants, it complies with the basic principles of 1964, Helsinki Manifesto and its later revisions. Sensitive information is erased from medical imaging data exported from imaging databases, including but not limited to organization name, organization address, physician's name, patient's name, patient's birthday, *etc*. Besides, personal data are kept confidential, such as ID number, home address, contact information, medical insurance information, *etc*. Acquisition, transmission, and use of data should meet relevant legal requirements.

In addition, medical imaging data, which are consistent with standard imaging protocols, are the foundation of radiomics[12,13]. It can be single- or multi-center, and retrospective or prospective. Although there are various types of imaging examinations, including computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound, *etc*.[11,14-16] for different research purposes, the dominant examination methods or sequences are more recommended.



DOI: 10.35711/aimi.v3.i4.87

Figure 1 The flow diagram of radiomics. CEA: Carcinoembryonic antigen; CA125: Carbohydrate antigen 125; GLCM: Gray-level co-occurrence matrix; GLSZM: Gray-level size zone matrix; GLRLM: Gray level run length matrix; GLDM: Gray-level difference method; SVM: Support vector machine; KNN: K Nearest Neighbor; ROC: Receiver operating characteristic; NCTDM: Neighbourhood gray-tone difference matrix.

Hence, more eligible cases are included to find out common features, which may contribute to the stability of models[17]. There is no general standard for the medical imaging data from different examination methods using different imaging methods, acquisition methods, imaging parameters, and imaging quality that may affect the subsequent analysis. Therefore, how to normalize the data and conform to the imaging standard is the focus of radiomics studies at present.

After data collection, the data need to be checked and confirmed, in order to correct or eliminate unqualified data. The specific inspection content includes the validity of the file format, the integrity of the sequence, and the correctness of the image content, in order to exclude unrecognizable images, sequence deletion, and wrong image layers. More detailed image quality specifications can also be formed according to specific research requirements. In the process of image quality control, it is necessary to sort out the imaging problems encountered, so that the data can be traced back when the inclusion and exclusion criteria are defined.

Preprocessing

Because of different scanning parameters, reconstruction procedures (slice thickness, voxel size, and reconstruction algorithm), and inconsistent imaging acquisition of multi-brand manufactories, it has a significant influence on distribution of features[18,19]. In order to decrease this discrepancy, preprocessing of the collected imaging data is essential. At present, the most common methods include resampling, gray-level discretization, and intensity normalization. Image resampling involves generation of equal-size voxels by applying the linear interpolation algorithm to improve image quality and to eliminate bias introduced by non-uniform imaging resolution[20]. Gray-level discretization refers to the bundling of pixels based on their density, either by relative discretization (fixed number) or absolute discretization (fixed size)[21]. Image intensity normalization is used to correct inter-subject intensity variation by transforming all images from original greyscale into a standard greyscale. Furthermore, image enhancement approaches, such as image flipping, image rotation, image distortion, image transformation, and image scaling, can enrich data diversity, improve model generalization ability, and reduce the risk of model overfitting.

In addition to the above-mentioned methods, not only for images, we also need to preprocess clinical data. Deidentification of data is beneficial to protect personal information and query data among multiple departments. Hospital number is advised to be the unique identification, realizing the mapping of images. In order to effectively eliminate the deficiency of data inconsistency and bias in multi-center studies, it is necessary to conduct data consistency processing, which is advantageous to realize cross-center data modeling and verification. The methods of data consistency processing include: (1) Standardization of data collection: Data are collected according to the unified data acquisition standard in each center; (2) Consistency processing based on extracted features: The method of Z-score

can be used to standardize data; and (3) Consistency processing based on image domain: According to the annotated information, the size of region of interest (ROI) is kept consistent.

Segmentation

Segmentation of ROI can be divided into manual and semiautomatic/automatic segmentation, two-dimensional (2D) and three-dimensional (3D) segmentation, and intratumoral and peritumoral segmentation[22-26]. This process is relatively tedious and requires open-source or dedicated software to support[12]. The process at least needs one labeling physician and one senior physician. The knowledge of relevant anatomy and imaging should be well known by labeling physicians and they must be familiar with the sketching software. In addition, for manual segmentation, intra-class correlation coefficient and concordance correlation coefficient can be advantageous to reduce the discrepancy of subjective judgement and the intra- and inter-reader variability[17,27]. Due to the rapid development of computer science, semiautomatic/automatic segmentation has been frequently applied. Automatic segmentation aims to draw ROIs automatically[28], while semiautomatic segmentation still requires partially manual intervention to mark the center of the lesion before automatic segmentation [29]. They both decrease instability to a certain extent, however, they are less applied because of technical restriction. At present, automatic segmentation can be summarized into three categories[30]: (1) Algorithms based on intensity thresholds and regions; (2) algorithms based on statistical approaches and deformable models; and (3) algorithms incorporating empirical knowledge into the segmentation process.

Feature extraction

Features are extracted from ROIs using different software with the similar code, which consist of first-order, second-order, and higher-order features. First-order features describe the geometric attributes and the distribution of voxel intensities of the ROIs, including mean, median, maximum, and minimum values, as well as the skewness, kurtosis, and entropy. Second-order features represent the relationships between adjacent voxels to measure features[31]. Second-order textural features describe the gray-scale alterations and are extracted by different algorithms. Higher-order features are extracted *via* wavelet, Laplacian, and Gaussian filters from multiple dimensions[32]. With the combination of multiple omics, semantic features, which are based on the experience and knowledge of radiologists, pathological features, genetic features, *etc*., all promote the transformation of radiomics into clinical practice. In recent years, depiction of deep learning (DL)-based features, which are supplementary high-dimensional features, by observers has been reported as a challenge[33]. Although DL-based features reveal certain advantages in terms of estimating prognosis of malignancies, it is enslaved to be widely used by data size and technological development.

Feature selection

According to the fourth step (feature extraction), the great number of extracted features is achieved, and how to select the most relevant features is the key to establish a robust radiomics model. This process simplifies the mathematical problem by decreasing the number of parameters and also reduces the risk of overfitting. Specific methods include univariate, the least absolute shrinkage and selection operator (LASSO), RELIEF algorithm, redundancy maximum relevance (MRMR), *etc*[34].

Modeling and verification

The ultimate objective of radiomics is to establish an effective model for classification and prediction. The data should be clustered into training and validation datasets. Different classifiers, including logistics, support vector machine, Bayes, k-Nearest Neighbor algorithm, Tree and Forest, are used to set up models and to select the most effective model by seed circling for clinical transformation[35]. Meanwhile, the predictive performance of the final model should be verified on a separate cohort, and an external validation cohort is highly appropriate to confirm its generalization. Owing to the lack of data sharing, obtaining the results of external validation of the model is a challenge at this stage.

CLINICAL APPLICATION OF RADIOMICS

Diagnosis and staging

In previous studies, radiomics has shown a great potential in the diagnosis and staging of different diseases. Although the diagnosis of some lesions is easy according to imaging manifestations, radiomics can improve physicians' diagnostic confidence and patients' examination strategies. In a plain CT study, 168 patients with hepatocellular carcinoma (HCC) and 117 patients with hepatic hemangioma were analyzed. Textural features were extracted from plain CT images and 13 features were selected from 1223 candidate features to constitute the radiomics signature, in order to establish a logistic regression model to classify benign and malignant liver tumors. The final model achieved an average area under the curve (AUC) of 0.87. In spite of the lack of innovation, it helps patients who cannot successfully

undergo contrast-enhanced CT (CECT) because of iodine contrast agent allergy for a relatively accurate diagnosis[36].

In another study, Ding *et al*[37] explored the capacity of the combined model for differentiating HCC from focal nodular hyperplasia (FNH) in non-cirrhotic livers using Gd-DTPA contrast-enhanced MRI. For this purpose, 8 radiomics features were selected for the radiomics model, and 4 clinical factors (age, gender, hepatitis B surface antigen (HbsAg), and enhancement pattern) were chosen for the clinical model. The combined model was established using the factors from the previous models. The classification accuracy of the combined model that differentiated HCC from FNH in both the training and validation datasets was 0.956 and 0.941, respectively. The model could support clinicians to make more reliable clinical decisions.

Serous cystadenomas (SCN) are considered as mostly benign cystic neoplasm in the pancreas. Mucinous cystic neoplasm (MCN) is an easily misdiagnosed lesion of SCN, which is associated with the risk of malignant transformation[38]. Therefore, Xie *et al*[39] confirmed the value of CT-based radiomics analysis in preoperatively discriminating pancreatic MSN and SCN. A total of 103 MCN and 113 SCN patients who underwent surgery were retrospectively enrolled. The Rad-score model was proved to be robust and reliable (average AUC, 0.784; sensitivity, 0.847; specificity, 0.745; positive-predictive value (PPV), 0.767; negative-predictive value, 0.849; accuracy, 0.793), which could serve as a novel tool for guiding clinical decision-making.

In another multi-center study, researchers took advantages of radiomics to develop a nomogram for preoperatively predicting grade 1 and grade 2/3 tumors in patients with pancreatic neuroendocrine tumors (PNETs). Totally, 138 patients from two institutions with pathologically confirmed PNETs were included in that retrospective study. The nomogram integrating an independent risk factor of tumor margin and fusion radiomic signature showed a strong discrimination with an AUC of 0.974 (95% confidence interval (CI): 0.950–0.998) in the training cohort and 0.902 (95% CI: 0.798–1.000) in the validation cohort, with a satisfactory calibration. Decision curve analysis (DCA) verified the clinical applicability of the predictive nomogram[40].

Evaluation of tumor biological behaviors

Concurrent advancements in imaging and genomic biomarkers have facilitated identification of noninvasive imaging surrogates of molecular phenotypes. Villanueva *et al*[41] investigated the genomic features of HCC and peritumoral tissues that were associated with patients' outcomes, and they explored the relationship between imaging traits and genomic signatures. Patients who underwent preoperative CT or MRI and transcriptome profiling were assessed using 11 qualitative and 4 quantitative (size, enhancement ratio, wash-out ratio, tumor-to-liver contrast ratio) imaging traits. Several imaging traits, including infiltrative pattern and macrovascular invasion were found to be associated with gene signatures of aggressive HCC phenotype, such as proliferative signatures and CK19 signature.

Microvascular invasion (MVI) is one of the strongest predictors of hepatic transplantation or hepatectomy for HCC, which is one of the independent factors for early recurrence and poor prognosis [42]. MVI could be diagnosed postoperatively and it was defined as the presence of tumor within microscopic vessels of the portal vein, hepatic artery, and lymphatic vessels[43]. Conventional imaging methods cannot reveal MVI because of the poor resolution before operation. Therefore, it is important to develop a non-invasive tool to detect MVI for clinical decision-making. Zhu *et al*[44] proposed a nomogram for the prediction of MVI that included a radiomic score and alpha fetoprotein, tumor type, peritumoral enhancement, arterial rim, and internal arteries. This nomogram was superior to a clinical and radiologic model with an AUC of 0.858 versus 0.729. In another research, Renzulli *et al*[45] demonstrated that non-smooth tumor margins and peritumoral enhancement, combined with the radiogenomic features were independent predictors for MVI with a PPV of 0.95. In a large-scale study, Xu *et al*[46] collected CT scan images from 495 patients and developed a combined model which consisted of semantic features (aspartate aminotransferase, alpha fetoprotein (AFP), non-smooth tumor margin, extrahepatic growth, ill-defined pseudocapsule, and peritumoral arterial enhancement) and radiomic features to predict histological MVI, with an AUC of 0.909 and 0.889 in the training cohort and the test cohort, respectively.

Gao *et al*[47] assessed the preoperative prediction of TP53 status based on multiparametric MRI (mp-MRI) radiomic features extracted from 3D images. In total, 57 patients with pancreatic cancer who underwent preoperative MRI were included. The 3D ADC-ap-DWI-T2WI model with 11 selected features yielded the best performance for differentiating TP53 status, with an accuracy of 0.91 and an AUC of 0.96. The model revealed a good calibration, and the DCA proved the clinical value of the model. The radiomics model derived from mp-MRI provided a non-invasive, quantitative method to predict mutational status of TP53 in patients with pancreatic cancer that might contribute to the precision treatment.

Prognosis

Current guidelines recommend surgical resection as the first-line therapy for patients with HCC[48]. However, postoperative recurrence rate remains high and there is no reliable prediction tool. In a multi-center study, the potential of radiomics coupled with machine learning algorithms was assessed to improve the predictive accuracy for HCC recurrence. Using the machine learning framework, they

identified a three-feature signature that demonstrated a favorable prediction of HCC recurrence across all datasets, with C-index of 0.633-0.699. AFP, albumin-bilirubin, hepatic cirrhosis, tumor margin, and radiomic signature were selected for developing a preoperative model; the postoperative model incorporated satellite nodules into the above-mentioned predictors. The two models showed a superior prognostic performance, with C-index of 0.733-0.801 and integrated Brier score of 0.147-0.165, compared with rival models without radiomics, and are widely used in staging systems. Combined with clinical data, a three-feature fusion signature generated by aggregated ML-based framework could accurately predict individual recurrence risk, enabling appropriate management and surveillance of HCC[49]. In another study, CECT with measurement of Gabor and Wavelet radiomics features in patients with a single HCC tumor treated by hepatectomy revealed that several features were associated with both overall survival (OS) and disease-free survival (P values < 0.05)[50]. Similar results were reported by a separate study that risk scores developed from radiomics nomograms obtained from CECT textural data overmatched traditional clinical staging systems in both the training and validation cohorts for both tumor recurrence and OS[51].

Patients with pancreatic cancer have a poor prognosis, therefore, it is necessary to identify tumor characteristics associated with prognosis. Toyama *et al*[52] enrolled 161 patients with pancreatic cancer who underwent fluorodeoxyglucose (FDG)-PET/CT before treatment. The area of the primary tumor was semi-automatically contoured with a threshold of 40% of the maximum standardized uptake value, and 42 PET-based features were extracted. Among the PET parameters, 10 features showed statistical significance for predicting OS. Multivariate Cox regression analysis revealed gray-level zone length matrix (GLZLM)-gray-level non-uniformity (GLNU) as the only PET parameter showing statistical significance. In the random forest model, GLZLM-GLNU was the most relevant factor for predicting 1-year survival, followed by total lesion glycolysis. Radiomics with machine learning using FDG-PET in patients with pancreatic cancer provided valuable prognostic information.

DISCUSSION

There is no doubt that radiomics as a newly emerged quantitative technique is burgeoning in disease management. Nevertheless, the majority of the research of radiomics encountered common problems, and whether the radiomic-based signatures can be used in clinical practice needs to be discussed.

Reproducibility is one of the primary challenges that radiomic techniques must overcome for clinical application. At present, imaging protocols are not standardized worldwide, and hence, variability in image acquisition and reconstruction parameters is inevitable in clinical practice. A recent study demonstrated that the quantitative values of radiomic features varied according to imaging protocols [53]. In addition, although IBSI seeks standardization for radiomic extraction, the differences in techniques or platforms adopted in different centers may lead to differences in feature values[5], propagating to the radiomic signatures. Most radiomic signatures have a sharp drop in performance from training cohort to validation cohort. Researchers have adopted data normalization methods to correct for multicenter effects, such as ComBat harmonization[54]. However, whether the radiomic-based signature developed by normalized radiomic features is appropriate for clinical practice has not yet been studied. It is urgent to develop a reproducible radiomic signature that could overcome inherent multicenter effects, which is the basis for clinical individualized application.

Data sharing for independent validation is a challenge for radiomic signatures. To date, studies have mainly developed and validated the radiomic signatures using imaging data derived from their own center or multiple centers according to the same imaging protocols[55]. However, whether the signatures would be effective in completely independent centers needs further validation. Although images are more readily available than tissue molecular assays, the current open radiomic datasets are not enough for the independent validation. To eliminate this deficiency, data sharing among institutes and hospitals around the country or even around the world is important for radiomics, although it presents complex logistical problems. The Cancer Imaging Archive provides a good example of data sharing with a large portion of clinical data[56], and it is still growing with contribution from different institutes and hospitals. A previous study indicated that signatures should be validated using an open dataset that could become the standard to demonstrate their effectiveness[9].

Biological interpretability of radiomic signatures would accelerate their clinical application. Clinical experts mainly assume the radiomic model as a black box that can provide promising prediction results for clinical outcomes, which may make radiomics as a less accepted approach. The problem is further aggravated in the context of deconvolutional neural or DL networks, which even lack the observable model that solely concentrates on maximizing performance. A great number of these so-called "black-box" approaches may be perfectly viable in the diagnostic setting; however, when it comes to radiomic signatures for optimizing treatment, the question of interpretability becomes more paramount because a biomarker-driven treatment decision needs an explanation rooted in pathophysiology[57]. The emergence of radio-genomics provides a bridge for linking the radiomics to the underlying biological progression. The biological interpretability may provide biological evidence for the predictive ability of the radiomic signatures.

Clinical operability is the key in the clinical adoption of prognostic and predictive radiomic tools. To date, radiomic-based studies have mainly concentrated on developing robust signatures, and their application details in clinical practice are lack. Therefore, translating the computer language into a simple software or system may be an effective method to promote clinical application of radiomics.

CONCLUSION

In conclusion, the current researches have achieved encouraging results of radiomics and revealed potential of clinical applications, while poor standardization and generalization of radiomics limit the further translation of this method into clinical routine. How to make reproducibility of data, multi-center data sharing, biological interpretability of radiomic signatures and clinical operability come true, will become the crucial issue for development of radiomics. Only then will radiomics be more comparable and increase reliability to get clinician's approval. In foreseeable future, the development of radiomics will occupy a significant position in personalization and precision medicine. At present, it is more important to make clinical participants be conscious of benefits and limitations of radiomics in order to obtain reasonable decision towards clinical practice.

FOOTNOTES

Author contributions: Jiang ZY and Qi LS contributed equally to this work; Jiang ZY and Qi LS wrote the manuscript; Wang KZ designed the summary of the article and made constructive comments; Li JT, Cui N and Li W searched the relevant literatures and corrected the content; Liu W modified the format of the article; All authors have read and approve the final manuscript.

Conflict-of-interest statement: All the authors have no potential conflicts of interest to disclose.

Country/Territory of origin: China

ORCID number: Ke-Zheng Wang 0000-0001-8776-5812.

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

REFERENCES

1 **Lambin P**, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; **48**: 441-446 [PMID: 22257792 DOI: 10.1016/j.ejca.2011.11.036]

2 **Lambin P**, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**: 749-762 [PMID: 28975929 DOI: 10.1038/nrclinonc.2017.141]

3 **Verma V**, Simone CB 2nd, Krishnan S, Lin SH, Yang J, Hahn SM. The Rise of Radiomics and Implications for Oncologic Management. *J Natl Cancer Inst* 2017; **109** [PMID: 28423406 DOI: 10.1093/jnci/djx055]

4 **Aerts HJ**, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; **5**: 4006 [PMID: 24892406 DOI: 10.1038/ncomms5006]

5 **Zwanenburg A**, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R, Bogowicz M, Boldrini L, Buvat I, Cook GJR, Davatzikos C, Depeursinge A, Desseroit MC, Dinapoli N, Dinh CV, Echegaray S, El Naqa I, Fedorov AY, Gatta R, Gillies RJ, Goh V, Götz M, Guckenberger M, Ha SM, Hatt M, Isensee F, Lambin P, Leger S, Leijenaar RTH, Lenkowicz J, Lippert F, Losnegård A, Maier-Hein KH, Morin O, Müller H, Napel S, Nioche C, Orlhac F, Pati S, Pfaehler EAG, Rahmim A, Rao AUK, Scherer J, Siddique MM, Sijtsema NM, Socarras Fernandez J, Spezi E, Steenbakkers RJHM, Tanadini-Lang S, Thorwarth D, Troost EGC, Upadhaya T, Valentini V, van Dijk LV, van Griethuysen J, van Velden FHP, Whybra P, Richter C, Löck S. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* 2020; **295**: 328-

338 [PMID: 32154773 DOI: 10.1148/radiol.2020191145]
6 **Hu W**, Yang H, Xu H, Mao Y. Radiomics based on artificial intelligence in liver diseases: where we are? *Gastroenterol Rep (Oxf)* 2020; **8**: 90-97 [PMID: 32280468 DOI: 10.1093/gastro/goaa011]
7 **Gillies RJ**, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016; **278**: 563-577 [PMID: 26579733 DOI: 10.1148/radiol.2015151169]
8 **Limkin EJ**, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, Schernberg A, Paragios N, Deutsch E, Ferté C. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* 2017; **28**: 1191-1206 [PMID: 28168275 DOI: 10.1093/annonc/mdx034]
9 **Liu Z**, Wang S, Dong D, Wei J, Fang C, Zhou X, Sun K, Li L, Li B, Wang M, Tian J. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* 2019; **9**: 1303-1322 [PMID: 30867832 DOI: 10.7150/thno.30309]
10 **Paul R**, Schabath M, Balagurunathan Y, Liu Y, Li Q, Gillies R, Hall LO, Goldgof DB. Explaining Deep Features Using Radiologist-Defined Semantic Features and Traditional Quantitative Features. *Tomography* 2019; **5**: 192-200 [PMID: 30854457 DOI: 10.18383/j.tom.2018.00034]
11 **Fu S**, Wei J, Zhang J, Dong D, Song J, Li Y, Duan C, Zhang S, Li X, Gu D, Chen X, Hao X, He X, Yan J, Liu Z, Tian J, Lu L. Selection Between Liver Resection Versus Transarterial Chemoembolization in Hepatocellular Carcinoma: A Multicenter Study. *Clin Transl Gastroenterol* 2019; **10**: e00070 [PMID: 31373932 DOI: 10.14309/ctg.0000000000000070]
12 **Park S**, Chu LC, Hruban RH, Vogelstein B, Kinzler KW, Yuille AL, Fouladi DF, Shayesteh S, Ghandili S, Wolfgang CL, Burkhart R, He J, Fishman EK, Kawamoto S. Differentiating autoimmune pancreatitis from pancreatic ductal adenocarcinoma with CT radiomics features. *Diagn Interv Imaging* 2020; **101**: 555-564 [PMID: 32278586 DOI: 10.1016/j.diii.2020.03.002]
13 **Dalal V**, Carmicheal J, Dhaliwal A, Jain M, Kaur S, Batra SK. Radiomics in stratification of pancreatic cystic lesions: Machine learning in action. *Cancer Lett* 2020; **469**: 228-237 [PMID: 31629933 DOI: 10.1016/j.canlet.2019.10.023]
14 **Zhou T**, Xie CL, Chen Y, Deng Y, Wu JL, Liang R, Yang GD, Zhang XM. Magnetic Resonance Imaging-Based Radiomics Models to Predict Early Extrapancreatic Necrosis in Acute Pancreatitis. *Pancreas* 2021; **50**: 1368-1375 [PMID: 35041335 DOI: 10.1097/MPA.0000000000001935]
15 **Gao X**, Wang X. Performance of deep learning for differentiating pancreatic diseases on contrast-enhanced magnetic resonance imaging: A preliminary study. *Diagn Interv Imaging* 2020; **101**: 91-100 [PMID: 31375430 DOI: 10.1016/j.diii.2019.07.002]
16 **Wei M**, Gu B, Song S, Zhang B, Wang W, Xu J, Yu X, Shi S. A Novel Validated Recurrence Stratification System Based on (18)F-FDG PET/CT Radiomics to Guide Surveillance After Resection of Pancreatic Cancer. *Front Oncol* 2021; **11**: 650266 [PMID: 34055620 DOI: 10.3389/fonc.2021.650266]
17 **Li Y**, Reyhan M, Zhang Y, Wang X, Zhou J, Yue NJ, Nie K. The impact of phantom design and material-dependence on repeatability and reproducibility of CT-based radiomics features. *Med Phys* 2022; **49**: 1648-1659 [PMID: 35103332 DOI: 10.1002/mp.15491]
18 **Duron L**, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik JC, Thomassin-Naggara I, Fournier L, Lecler A. Gray-level discretization impacts reproducible MRI radiomics texture features. *PLoS One* 2019; **14**: e0213459 [PMID: 30845221 DOI: 10.1371/journal.pone.0213459]
19 **Moradmand H**, Aghamiri SMR, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. *J Appl Clin Med Phys* 2020; **21**: 179-190 [PMID: 31880401 DOI: 10.1002/acm2.12795]
20 **Larue RTHM**, van Timmeren JE, de Jong EEC, Feliciani G, Leijenaar RTH, Schreurs WMJ, Sosef MN, Raat FHPJ, van der Zande FHR, Das M, van Elmpt W, Lambin P. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol* 2017; **56**: 1544-1553 [PMID: 28885084 DOI: 10.1080/0284186X.2017.1351624]
21 **Zhuge Y**, Udupa JK, Liu J, Saha PK. Image background inhomogeneity correction in MRI *via* intensity standardization. *Comput Med Imaging Graph* 2009; **33**: 7-16 [PMID: 19004616 DOI: 10.1016/j.compmedimag.2008.09.004]
22 **Peng J**, Zhang J, Zhang Q, Xu Y, Zhou J, Liu L. A radiomics nomogram for preoperative prediction of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma. *Diagn Interv Radiol* 2018; **24**: 121-127 [PMID: 29770763 DOI: 10.5152/dir.2018.17467]
23 **Huang X**, Long L, Wei J, Li Y, Xia Y, Zuo P, Chai X. Radiomics for diagnosis of dual-phenotype hepatocellular carcinoma using Gd-EOB-DTPA-enhanced MRI and patient prognosis. *J Cancer Res Clin Oncol* 2019; **145**: 2995-3003 [PMID: 31664520 DOI: 10.1007/s00432-019-03062-3]
24 **Ciaravino V**, Cardobi N, DE Robertis R, Capelli P, Melisi D, Simionato F, Marchegiani G, Salvia R, D'Onofrio M. CT Texture Analysis of Ductal Adenocarcinoma Downstaged After Chemotherapy. *Anticancer Res* 2018; **38**: 4889-4895 [PMID: 30061265 DOI: 10.21873/anticanres.12803]
25 **D'Onofrio M**, Ciaravino V, Cardobi N, De Robertis R, Cingarlini S, Landoni L, Capelli P, Bassi C, Scarpa A. CT Enhancement and 3D Texture Analysis of Pancreatic Neuroendocrine Neoplasms. *Sci Rep* 2019; **9**: 2176 [PMID: 30778137 DOI: 10.1038/s41598-018-38459-6]
26 **Braman NM**, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, Plecha D, Madabhushi A. Erratum to: Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res* 2017; **19**: 80 [PMID: 28693537 DOI: 10.1186/s13058-017-0862-1]
27 **Rios Velazquez E**, Aerts HJ, Gu Y, Goldgof DB, De Ruysscher D, Dekker A, Korn R, Gillies RJ, Lambin P. A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol* 2012; **105**: 167-173 [PMID: 23157978 DOI: 10.1016/j.radonc.2012.09.023]
28 **Häme Y**, Pollari M. Semi-automatic liver tumor segmentation with hidden Markov measure field model and non-parametric distribution estimation. *Med Image Anal* 2012; **16**: 140-149 [PMID: 21742543 DOI: 10.1016/j.media.2011.06.006]

29 **Permuth JB**, Choi J, Balarunathan Y, Kim J, Chen DT, Chen L, Orcutt S, Doepker MP, Gage K, Zhang G, Latifi K, Hoffe S, Jiang K, Coppola D, Centeno BA, Magliocco A, Li Q, Trevino J, Merchant N, Gillies R, Malafa M; Florida Pancreas Collaborative. Combining radiomic features with a miRNA classifier may improve prediction of malignant pathology for pancreatic intraductal papillary mucinous neoplasms. *Oncotarget* 2016; **7**: 85785-85797 [PMID: 27589689 DOI: 10.18632/oncotarget.11768]

30 **Chen W**, Liu B, Peng S, Sun J, Qiao X. Computer-Aided Grading of Gliomas Combining Automatic Segmentation and Radiomics. *Int J Biomed Imaging* 2018; **2018**: 2512037 [PMID: 29853828 DOI: 10.1155/2018/2512037]

31 **Li J**, Lu J, Liang P, Li A, Hu Y, Shen Y, Hu D, Li Z. Differentiation of atypical pancreatic neuroendocrine tumors from pancreatic ductal adenocarcinomas: Using whole-tumor CT texture analysis as quantitative biomarkers. *Cancer Med* 2018; **7**: 4924-4931 [PMID: 30151864 DOI: 10.1002/cam4.1746]

32 **Nougaret S**, Tardieu M, Vargas HA, Reinhold C, Vande Perre S, Bonanno N, Sala E, Thomassin-Naggara I. Ovarian cancer: An update on imaging in the era of radiomics. *Diagn Interv Imaging* 2019; **100**: 647-655 [PMID: 30555018 DOI: 10.1016/j.diii.2018.11.007]

33 **Thawani R**, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, Madabhushi A. Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer* 2018; **115**: 34-41 [PMID: 29290259 DOI: 10.1016/j.lungcan.2017.10.015]

34 **Matzner-Lober E**, Suehs CM, Dohan A, Molinari N. Thoughts on entering correlated imaging variables into a multivariable model: Application to radiomics and texture analysis. *Diagn Interv Imaging* 2018; **99**: 269-270 [PMID: 29751945 DOI: 10.1016/j.diii.2018.04.011]

35 **Rizzo S**, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, Bellomi M. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* 2018; **2**: 36 [PMID: 30426318 DOI: 10.1186/s41747-018-0068-z]

36 **Yin J**, Qiu JJ, Qian W, Ji L, Yang D, Jiang JW, Wang JR, Lan L. A radiomics signature to identify malignant and benign liver tumors on plain CT images. *J Xray Sci Technol* 2020; **28**: 683-694 [PMID: 32568166 DOI: 10.3233/XST-200675]

37 **Ding Z**, Lin K, Fu J, Huang Q, Fang G, Tang Y, You W, Lin Z, Pan X, Zeng Y. An MR-based radiomics model for differentiation between hepatocellular carcinoma and focal nodular hyperplasia in non-cirrhotic liver. *World J Surg Oncol* 2021; **19**: 181 [PMID: 34154624 DOI: 10.1186/s12957-021-02266-7]

38 **Fernández-del Castillo C**. Mucinous cystic neoplasms. *J Gastrointest Surg* 2008; **12**: 411-413 [PMID: 17955316 DOI: 10.1007/s11605-007-0347-0]

39 **Xie T**, Wang X, Zhang Z, Zhou Z. CT-Based Radiomics Analysis for Preoperative Diagnosis of Pancreatic Mucinous Cystic Neoplasm and Atypical Serous Cystadenomas. *Front Oncol* 2021; **11**: 621520 [PMID: 34178619 DOI: 10.3389/fonc.2021.621520]

40 **Gu D**, Hu Y, Ding H, Wei J, Chen K, Liu H, Zeng M, Tian J. CT radiomics may predict the grade of pancreatic neuroendocrine tumors: a multicenter study. *Eur Radiol* 2019; **29**: 6880-6890 [PMID: 31227882 DOI: 10.1007/s00330-019-06176-x]

41 **Villanueva A**, Hoshida Y, Battiston C, Tovar V, Sia D, Alsinet C, Cornella H, Liberzon A, Kobayashi M, Kumada H, Thung SN, Bruix J, Newell P, April C, Fan JB, Roayaie S, Mazzaferro V, Schwartz ME, Llovet JM. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology* 2011; **140**: 1501-12.e2 [PMID: 21320499 DOI: 10.1053/j.gastro.2011.02.006]

42 **Mazzaferro V**, Llovet JM, Miceli R, Bhoori S, Schiavo M, Mariani L, Camerini T, Roayaie S, Schwartz ME, Grazi GL, Adam R, Neuhaus P, Salizzoni M, Bruix J, Forner A, De Carlis L, Cillo U, Burroughs AK, Troisi R, Rossi M, Gerunda GE, Lerut J, Belghiti J, Boin I, Gugenheim J, Rochling F, Van Hoek B, Majno P; Metroticket Investigator Study Group. Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis. *Lancet Oncol* 2009; **10**: 35-43 [PMID: 19058754 DOI: 10.1016/S1470-2045(08)70284-5]

43 **Lim KC**, Chow PK, Allen JC, Chia GS, Lim M, Cheow PC, Chung AY, Ooi LL, Tan SB. Microvascular invasion is a better predictor of tumor recurrence and overall survival following surgical resection for hepatocellular carcinoma compared to the Milan criteria. *Ann Surg* 2011; **254**: 108-113 [PMID: 21527845 DOI: 10.1097/SLA.0b013e31821ad884]

44 **Zhu YJ**, Feng B, Wang S, Wang LM, Wu JF, Ma XH, Zhao XM. Model-based three-dimensional texture analysis of contrast-enhanced magnetic resonance imaging as a potential tool for preoperative prediction of microvascular invasion in hepatocellular carcinoma. *Oncol Lett* 2019; **18**: 720-732 [PMID: 31289547 DOI: 10.3892/ol.2019.10378]

45 **Renzulli M**, Brocchi S, Cucchetti A, Mazzotti F, Mosconi C, Sportoletti C, Brandi G, Pinna AD, Golfieri R. Can Current Preoperative Imaging Be Used to Detect Microvascular Invasion of Hepatocellular Carcinoma? *Radiology* 2016; **279**: 432-442 [PMID: 26653683 DOI: 10.1148/radiol.2015150998]

46 **Xu X**, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, Yang G, Yan X, Zhang YD, Liu XS. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol* 2019; **70**: 1133-1144 [PMID: 30876945 DOI: 10.1016/j.jhep.2019.02.023]

47 **Gao J**, Chen X, Li X, Miao F, Fang W, Li B, Qian X, Lin X. Differentiating TP53 Mutation Status in Pancreatic Ductal Adenocarcinoma Using Multiparametric MRI-Derived Radiomics. *Front Oncol* 2021; **11**: 632130 [PMID: 34079753 DOI: 10.3389/fonc.2021.632130]

48 **Vogel A**, Cervantes A, Chau I, Daniele B, Llovet JM, Meyer T, Nault JC, Neumann U, Ricke J, Sangro B, Schirmacher P, Verslype C, Zech CJ, Arnold D, Martinelli E. Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2019; **30**: 871-873 [PMID: 30715202 DOI: 10.1093/annonc/mdy510]

49 **Ji GW**, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, Li XC, Wang XH. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. *EBioMedicine* 2019; **50**: 156-165 [PMID: 31735556 DOI: 10.1016/j.ebiom.2019.10.057]

50 **Chen S**, Zhu Y, Liu Z, Liang C. Texture analysis of baseline multiphasic hepatic computed tomography images for the prognosis of single hepatocellular carcinoma after hepatectomy: A retrospective pilot study. *Eur J Radiol* 2017; **90**: 198-204 [PMID: 28583634 DOI: 10.1016/j.ejrad.2017.02.035]

51 **Zheng BH**, Liu LZ, Zhang ZZ, Shi JY, Dong LQ, Tian LY, Ding ZB, Ji Y, Rao SX, Zhou J, Fan J, Wang XY, Gao Q. Radiomics score: a potential prognostic imaging feature for postoperative survival of solitary HCC patients. *BMC Cancer*

2018; **18**: 1148 [PMID: 30463529 DOI: 10.1186/s12885-018-5024-z]

52 **Toyama Y**, Hotta M, Motoi F, Takanami K, Minamimoto R, Takase K. Prognostic value of FDG-PET radiomics with machine learning in pancreatic cancer. *Sci Rep* 2020; **10**: 17024 [PMID: 33046736 DOI: 10.1038/s41598-020-73237-3]

53 **Midya A**, Chakraborty J, Gönen M, Do RKG, Simpson AL. Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. *J Med Imaging (Bellingham)* 2018; **5**: 011020 [PMID: 29487877 DOI: 10.1117/1.JMI.5.1.011020]

54 **Orlhac F**, Frouin F, Nioche C, Ayache N, Buvat I. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. *Radiology* 2019; **291**: 53-59 [PMID: 30694160 DOI: 10.1148/radiol.2019182023]

55 **Wu G**, Woodruff HC, Shen J, Refaee T, Sanduleanu S, Ibrahim A, Leijenaar RTH, Wang R, Xiong J, Bian J, Wu J, Lambin P. Diagnosis of Invasive Lung Adenocarcinoma Based on Chest CT Radiomic Features of Part-Solid Pulmonary Nodules: A Multicenter Study. *Radiology* 2020; **297**: 451-458 [PMID: 32840472 DOI: 10.1148/radiol.2020192431]

56 **Lu L**, Sun SH, Yang H, E L, Guo P, Schwartz LH, Zhao B. Radiomics Prediction of EGFR Status in Lung Cancer-Our Experience in Using Multiple Feature Extractors and The Cancer Imaging Archive Data. *Tomography* 2020; **6**: 223-230 [PMID: 32548300 DOI: 10.18383/j.tom.2020.00017]

57 **Bera K**, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol* 2022; **19**: 132-146 [PMID: 34663898 DOI: 10.1038/s41571-021-00560-7]

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